EXPANDING PERSPECTIVES ON OPEN SCIENCE: COMMUNITIES, CULTURES AND DIVERSITY IN CONCEPTS AND PRACTICES

This page intentionally left blank

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices

Proceedings of the 21st International Conference on Electronic Publishing

Edited by

Leslie Chan

Centre for Critical Development Studies, University of Toronto Scarborough, Canada

and

Fernando Loizides

School of Mathematics and Computer Science, Faculty of Science and Engineering, University of Wolverhampton, UK

IOS Press

Amsterdam • Berlin • Washington, DC

ISBN 978-1-61499-768-9 (print)
ISBN 978-1-61499-769-6 (online)
Library of Congress Control Number: 2017942293

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

For book sales in the USA and Canada:
IOS Press, Inc.
6751 Tepper Drive
Clifton, VA 20124
USA
Tel.: +1 703 830 6300
Fax: +1 703 830 2300
sales@iospress.com

The cover shows Ruins of bishop's house – Early Christian basilica – Kourion archaeological site, Limassol. Cover photo by Alfonso Lorenzetto, Cyprus Tourism Organisation.

PRINTED IN THE NETHERLANDS

Preface

The term “electronic publishing” sounds a bit quaint today. But 21 years ago when the ELPUB conference series first started, the term promised all manner of potential that the Web and network technologies could bring to scholarly communication, scientific research and technical innovation. Indeed, over the last two decades we have seen tremendous developments across all these domains, and at the same time our social, economical and political lives have been completely transformed.

Open Science represents one such transformation, and not surprisingly, the elements that make Open Science possible, including open access, open data, open software, and other domains of open have been regular topics presented and debated at previous ELPUB conferences.

However, development and diffusion of open research practices are highly uneven across disciplines and across regions. And despite the common claims that Open Science improves transparency and accountability throughout the research life cycle while democratizing the knowledge production process, empirical research and conceptual validation of these ideas has been limited. In addition, there is a growing tendency to conceptualize Open Science as a set of conditions waiting to be met, without regard for regional differences, including cultural and historical contexts of knowledge production.

The theme of the conference this year, *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*, is intended to generate discussion and debate on the potential and limitations of openness. We thus invited, researchers and practitioners from diverse backgrounds to share their results and ideas at what we trust will be a highly interactive forum.

We also asked potential presenters to consider exploring alternative models of interaction and co-creation between scholars and citizen scientists, and the role of dissemination and publishing within these interactions. To stimulate submissions, we included these questions in the open call: Who determines the agenda and direction of emerging discourses around Open Science? How does Open Science challenge the current positions and power of players and agents in varying institutional contexts? Are we seeing a converging global view of Open Science, or are there disciplinary, regional, and other differences that are important to consider? What are the gaps between existing Open Science policies, regulatory frameworks, and implementation requirements and how should they be addressed? How do Open Science agendas relate to the Open Innovation agendas of governments, funders and institutions? What is the impact of these agendas on research funding and dissemination practices?

By assessing these interlinked questions, the aim is to improve our understanding of current challenges and opportunities in the ecosystem of open science, and how to move forward collaboratively in developing an inclusive system that works for a much broader range of participants.

All submissions were subjected to peer review, performed by members of the Program Committee. In all, a total of 27 research and practitioner papers and 7 posters are being presented at this year’s conference, along with 4 workshops on the first day of the conference. The papers represent a broad range of topics related to Open Science, from provision of common infrastructure, innovative tools, new publishing models, sustainability models, and policy provisions. We also have a broad range of conceptual papers exploring the boundaries and diversities of open research practices in varying institutional and cultural contexts. Perhaps for the first time in ELPUB history, we have speakers coming from countries that span the globe. This was due in part to the fact that several of the presenters are members of the Open and Collaborative Science in Development Network (OCSDNet), and they are coming from countries including Senegal, Jamaica, Brazil, Peru, Colombia, and Argentina. At the conference, we hope to have a productive discussion about the challenges and opportunities facing researchers and citizens in the global South.

In keeping with the theme of the conference, this year we have three diverse keynote speakers with diverse expertise speaking on diverse topics, but all related to the implications of how networked technologies are changing the way we produce, consume, and circulate knowledge. Rachel Harding, an early career researcher in genomics at the University of Toronto, will speak on "Open science and accelerating discovery in rare and neglected diseases." Dr. Hebe Vessuri, CIGA (Centro de Investigación en Geografía Ambiental) UNAM in Mexico, will speak on "Tapping knowledge globally: open access and mobile objects in an asymmetric world." And Mimis Sophocleous, Academic Director of the Historical Archives and Research Centre of Limassol, will be asking "What happens to poetry and prose when they go in digital form online instead of reaching their readers in a book form."

This year's conference takes place in Cyprus, the third largest island in the Mediterranean, after Sicily and Sardinia. At the crossroads of Europe, Asia, and Africa, its geographic location, rich, unique history and cultural diversity make it particularly well suited for hosting ELPUB 2017. We would like to thank the Library Director, Marios Zervas, of Cyprus University of Technology, and his Staff for their involvement regarding the sponsorships from publishers, the articles that will be presented on behalf of the CUT, and the promotional actions taken to advertise the ELPUB 2017 conference in the local community. We also thank the Publishers who readily and positively responded to the CUT Library's request for sponsorship and support. Last but not least, we thank Easy Conferences for their communication and collaboration with the CUT Library regarding the organization of the conference.

The staff at Easy Conference have been tremendously helpful and supportive. They were extremely prompt and attentive with our many requests. Not only did they assist with logistics and social event planning, they also provided valuable input regarding programing. We could not have managed this conference without their dedicated support.

We would like to express our sincere thanks to members of the ELPUB Executive Committee who, together with the Programme Committee, helped us to bring together a diverse and exciting programme. A special thanks to Saman Goudarzi, an undergraduate student at the University of Toronto at Scarborough, for her editorial assistance with the manuscripts. And thanks to Anne Marie de Rover and Paul Weij at the IOS Press for their support in the production of this proceedings.

We wish everyone a productive an inspiring conference. We would like to extend an invitation to all of you to the 22st edition of the conference, which will be held in Toronto, Canada. We hope to see an even more diverse group of presenters, topics, and attendees at this conference, and look forward to welcoming you to Toronto!

Leslie Chan and Fernando Loizides

June 6th, 2017

Contents

This page intentionally left blank

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-1

Open Science and Accelerating Discovery in Rare and Neglected Diseases

Rachel J. HARDING[1]
Structural Genomics Consortium, University of Toronto

Abstract. New medicines for many diseases, in particular neurodegenerative disorders, are not forthcoming, despite patient demands and billions of dollars spent on biomedical research globally. Traditional publishing methods in biomedical sciences are generally slow and disseminate manuscripts, sometimes without the inclusion of primary data, to a privileged audience affiliated to institutions which can afford publication subscription costs. To overcome this barrier to progressive scientific endeavors, many researchers are championing the use of preprints, transparent subject-relevant data repositories, open access journals and open lab notebooks in an effort to more effectively and efficiently communicate their research to a wider audience. In this talk I shall discuss these options and the decisions I have made as an early career researcher, to share my research output on Huntington's disease in real-time through an open lab notebook. Included will be a discussion of the motivations, methods and assessment of open online publishing, including an evaluation of my own open notebook endeavors.[2]

Keywords. Open Science, Open Data, Open Access, Open Notebook, Repository, Preprint, Huntington's Disease.

New medicines for many diseases, in particular neurodegenerative disorders, are not forthcoming, despite patient demands and billions of dollars spent on biomedical research in laboratories throughout the world. Following the complete sequencing of the human genome, many researchers hoped that they could use this information as a manual to human biology and disease, expecting that an era of rapid drug discovery would follow. In fact, studies examining productivity of pharmaceutical and biotech companies, show a steady decline since the genome's publication [1]. The reasons why the research community is struggling to develop such therapies to meet patient demand are complex and multifaceted. The way in which research output is disseminated to the wider scientific community has been identified as a key problem area by many biomedical researchers in hindering the development of novel therapies in a timely manner.

Traditional publishing methods in the biomedical sciences are fraught with numerous problems. A well-documented issue is publication bias [2], which leads to the preferential publication of "positive" data and complete research stories, not reflecting the true breadth of academic output. Recent efforts to resolve this issue have seen the rise in short communication-style manuscript journals, such as PLoS Currents, and funding agency publishing platforms, such as Wellcome Open Research, both of which encourage the rapid publication of data, protocols and findings important to

[1] Corresponding author: Rachel Harding, Structural Genomic Consortium, University of Toronto. Email: rachel.harding@utoronto.ca
[2] This paper is the text of the keynote delivered at the ELPUB2017 conference.

various scientific fields, irrespective of the scale of the study. Nonetheless, the esteem in which extensive and comprehensive research publications are held, persists in the mainstream.

The conventional publishing process itself is generally slow with often long timeframes between manuscript submission, review and acceptance [3]. This is unsatisfactory for researchers in fast-paced, innovative and competitive fields where developments can take place rapidly and authors are nervous of being scooped by competing laboratories. The long time-frame from bench to publication can slow translation of key breakthroughs and discoveries into the clinic, meaning that major field advancements may not be felt by patients within their lifetimes or the time frame of their particular condition.

Manuscripts are often accepted by journals, reporting only polished results with no commitment for the researchers to share or deposit raw data or code for the readers of their paper. Many subject areas report a reproducibility crisis and it is common for researchers to not be able to reproduce data and outcomes reported in the published literature [4]. Clear, full and honest reporting of data, methods and analysis would create a more transparent model in which mistakes could be highlighted earlier, validation of results by independent groups would be more achievable and a more cooperative ethos would exist between groups researching similar areas.

In a bid to chase increasingly shrinking research grants and positions, researchers aim to publish in journals held in high esteem by their peers, funding agencies and departmental tenure panels, which, whilst appearing “glamourous”, often have substantial fees for publication, subscription and double-dip pricing structures for open access [5]. This firstly places an onus on the research groups publishing in such journals to find additional funding by which to cover these publication costs and meet their funding stipulations, which increasingly demand short embargo times or open access publication. Secondly, where publication is not open access at the first point of publication, this hinders researchers at less well funded institutes around the world, which cannot afford large numbers of high cost subscription packages, as well as the general public, from being able to access up-to-date publications.

Beyond rigorous changes to the traditional publishing system, there are alternative communication strategies which may alleviate some of these issues to provide fast, inclusive, inexpensive, transparent and open dissemination of research and data.

Preprints have been a viable self-archiving arrangement in some disciplines for more than 20 years, working synergistically with traditional publication systems and other forms of scientific communication. In particular, the physics research community has created a discipline standard by which manuscripts are routinely published on arXiv, a preprint server which has now amassed more than a million manuscripts [6]. Released on specialist platforms prior to journal submission, preprints allow fast communication of research findings in an open access manner. ArXiv has many examples of highly cited manuscripts which have not been published in the traditional system, showing that in some cases, preprinting can be sufficient means of communicating findings as well as subsequent critique and citation by peers [7]. As articles are all presented in a similar format, readers are perhaps less biased in their assessment of the preprint manuscripts with relation to journal based factors such as impact factor, and instead can focus their attentions to the academic merit of the content within. Whilst allowing fast communication of research findings to peers, preprint servers still primarily publish traditional manuscript formats and deposition of associated data is not obligatory. None the less, preprint servers are increasingly popular in a wide-range of subject areas and are rapidly growing in number as are the number of articles which are self-archived by their authors prior to, or post-printed following, formal publication [8].

Many researchers are in fact depositing both raw and analyzed data sets, associated with manuscripts or independently accumulated, a trend which has grown during a time when the reliability and credibility of scientific findings has been drawn into question [9]. Many researchers now publish the data generated within their investigations in a bid to improve transparency as well as allowing secondary analysis by interested parties for posterity. Repositories exist for both specialist as well as broad interests and generally allow the archiving of almost any digital output from researchers. This can prove particularly lucrative for early career researchers who can digitally archive early works such as project reports, literature reviews, conference posters and lab group presentations, allowing them to generate an online presence of their scientific output from an early stage, as well as preserving their works in the process. Both preprint servers and data repositories typically permit fast sharing of research outputs through social media platforms, promoting works among peers for review and assessment with online commentary of secondary analyses and opinions of the works. Repositories also represent a critical resource for data scientists and those working with big data for data mining and meta-analyses.

However, many experimental findings still elude publication of any kind for various reasons. These include failing to complete the research project or story within the time-frame of funding, having "negative" or contradictory data to the field dogma as well as struggling to resolve difficult methodological issues. The incentive to invest time publishing these types of projects within a traditional framework is low given the poor returns in the value added to a researcher's profile and possible negative impact on a researcher's reputation. None-the-less, these outputs include important findings which can be useful to other researchers in the field. Other work evades publication due to the filing of patents of the findings. However, a study of patents relating to genetic data showed that patents do not encourage innovation in their specific areas [10] so perhaps should not be considered to have a positive impact on academic advancement.

I currently work as a postdoctoral fellow at the Structural Genomics Consortium (SGC), a not-for-profit public-private partnership with a focus on accelerating science in understudied areas of human biology and disease. With funding from the CHDI Foundation, the SGC has a number of research projects focused on Huntington's disease (HD), a devastating inherited neurodegenerative disease with limited therapies no available cures. Both organizations have agreed not to file for patents on anything that results from this collaboration, as well as committing to make all their data and biochemical materials resulting generated during this relationship, freely available to the broader research community.

The aim of my particular project is to understand how the underlying genetic mutation of HD, the hallmark of the disease, gives rise to the disease phenotype seen in patients using structural biology methods. Very little evidence of similar research is available in the literature despite anecdotal evidence that similar projects have been pursued in both academic and industrial labs. In an effort to accelerate research and innovate within this field, I am writing up all of my findings in close to real-time in an open notebook using the data repository Zenodo in combination with my blog, labscribbles.com. I hope that by sharing all data generated for this project freely, honestly, effectively and efficiently within the field, to generate an online international team of scientists who can critique my research, offer suggestions for future experiments as well as collaborate on certain aspects of the project to try and reach our common research goals as quickly as possible. To date this project has resulted in numerous open collaborations being established both within the HD field and beyond, resulting in faster progression towards research milestones. I am now working to establish a portal by which I can freely share reagents generated in the course of this project in addition to the data to the scientific community.

The SGC is now considering adopting the open notebook concept for other rare disease projects. These ventures would likely have common goals; to establish an openly shared tool-box of reagents and data as a starting platform for fellow and future researchers to use to continue research in specific disease areas. Creating a range of high quality research tools should also accelerate research in fields where few materials or starting reagents are commercially available. In under-studied fields where researchers lack the risk of being scooped, an open notebook would be an excellent resource to the fields future research base, with minimal career risk to the scientist generating the data and materials in the first instance.

In the last year, the SGC launched its target enabling package (TEP) scheme. This initiative is built upon the recognition that genetic data is a good starting point for understanding certain diseases, but is insufficient alone to propel a translational project in drug discovery or even deeper understanding of the drug target or disease. As such, TEPs generated by the SGC will provide a critical mass of reagents and knowledge on a given protein target, the aim of which is to allow rapid biochemical and chemical exploration as well as characterization of proteins with genetic linkage to key disease areas. The primary goal is to accelerate drug discovery for these new targets with a fast, open access approach. All data and reagents generated as part of a TEP are shared without restriction with interested parties, even prior to formal publication.

A crucial point for the successful implementation of innovative or novel communication strategies, is that they provide positive impact to the fields adopting them. This requires comprehensive assessment of the “success” of the different methodologies compared to traditional publication methods. For different fields and different strategies, success will likely be defined differently and I believe it unlikely that all disciplines will implement all approaches well. The TEP initiative and my own open notebook project are subject to internal assessment at the SGC for their effectiveness as well as providing insight as to what future initiatives might be developed. In particular, for labscribbles, I am keen to take risks and try different approaches to find the most efficient and effective way of running this project. Whilst striving to maintain key standards with respect to time from bench to online update, the depth of detail describing each notebook installment and so forth, I hope to continue to develop and evolve the project into a success, as well as promoting this approach to others.

In embarking on my open access endeavors, I have been fortunate to have the full support of the SGC, in whose labs I am based, as well as the CHDI Foundation, who have generously funded this work. In particular, I have appreciated the mentorship of Aled Edwards, Cheryl Arrowsmith and Leticia Toledo-Sherman.

References

[1] Bernstein Research Report, The Long View - R&D productivity, 2010
[2] Nissen S.B. et al., Research: Publication bias and the canonization of false facts, *eLife*, **5** (2016), e21451

[3] Powell K., Does it take too long to publish research? *Nature*, **530(7589)** (2016), 148-51
[4] Baker M., 1,500 scientists lift the lid on reproducibility, *Nature* **533(7604)** (2016), 452-4
[5] Pinfield S. et al., The "Total Cost of Publication" in a Hybrid Open-Access Environment: Institutional Approaches to Funding Journal Article-Processing Charges in Combination With Subscriptions, *ASIS&T* **67(7)** (2016), 1751-1766
[6] Vence T., One Million Preprints and Counting: A conversation with arXiv founder Paul Ginsparg, *The Scientists* (2014)
[7] Perelman G., The entropy formula for the Ricci flow and its geometric applications, *ArXiv* (2002)
[8] http://asapbio.org/preprint-info/biology-preprints-over-time
[9] How science goes wrong, *The Economist* (2013)
[10] Sampat B. et al, How do patents affect follow-on innovation? Evidence from the human genome, *NBER* (2015) 21666

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-6

Openness in Scholarship: A Return to Core Values?

Cameron NEYLON[1]
Centre for Culture and Technology, Curtin University, Perth, Western Australia

Abstract. The debate over the meaning, and value, of open movements has intensified. The fear of co-option of various efforts from Open Access to Open Data is driving a reassessment and re-definition of what is intended by "open". In this article I apply group level models from cultural studies and economics to argue that the tension between exclusionary group formation and identity and aspirations towards inclusion and openness are a natural part of knowledge-making. Situating the traditional Western Scientific Knowledge System as a culture-made group, I argue that the institutional forms that support the group act as economic underwriters for the process by which groups creating exclusive knowledge invest in the process of making it more accessible, less exclusive, and more public-good-like, in exchange for receiving excludable goods that sustain the group. A necessary consequence of this is that our institutions will be conservative in their assessment of what knowledge-goods are worth of consideration and who is allowed within those institutional systems. Nonetheless the inclusion of new perspectives and increasing diversity underpins the production of general knowledge. I suggest that instead of positioning openness as new, and in opposition to traditional closed systems, it may be more productive to adopt a narrative in which efforts to increase inclusion are seen as a very old, core value of the academy, albeit one that is a constant work in progress.

Keywords. openness, open access, cultural science, culture, club economics, collective action, knowledge, epistemology

1. The Many Strands of "Open"

"Open" is a contested and increasingly it seems polarized term. It is also highly contextual. A number of different efforts have been made to disentangle the various discourses that underpin the advocacy programs that operate under the banner of open, but there is, as yet, little consistency between them. Fecher and Friesike's "Five Schools of Thought" [1] sit uneasily beside Pomerantz and Peek's "50 Shades of Open" [2], and while they both refer to the Open Knowledge Definition, various Open Access declarations and the debate between Free and Open Source software there is no clarity of definition.

Arguably all of these roots and their more recent interrogations are strongly rooted in Anglo-American conceptions of scholarship and political economy. "Open" in scholarship borrows heavily from the movements for Free and Open Source Software (F/OSS) while sitting alongside the movements advocating Open Government and Open Data. All of these are rooted in Western and Anglo-American discourses, not

[1] Corresponding author, Centre for Culture and Technology, Curtin University, Perth, Western Australia; E-mail: cn@cameronneylon.net.

infrequently with a techno-utopian and neo-liberal slant. Coleman notes how the distancing of F/OSS discourses from "[…]movements predicated on some political intentionality, direction, or reflexivity or a desire to transform wider social conditions" nonetheless serves those political programs [3].

These discourses connect "open" in scholarship to networked communications systems and usually the web. The connection to F/OSS as the supposed historical root of openness often makes this explicit. This in turn connects "open" to broader discussions of collaboration that are also seen as being supported by networked communications infrastructures. Opportunities to be gained through engagement, both in open sharing of resources and in collaboration are assumed to provide equitable gains. Openness in these discourses is presumed to be uniformly positive for all who engage with it. The presumption of equitable opportunities for the traditionally disenfranchised and disempowered is a driving motivation for many engaged in Open movements.

At the same time Nathaniel Tkacz [4] that "openness" is almost always situated as an oppositional movement, one that opposes "traditional" and "closed" processes whether they be in government, reporting, property, or scholarly communications. He draws a thread from Popper's The Open Society via the neo-liberal discourses inspired by Hayek to the rhetorics of F/OSS and their successors.

> "Openness is conceived as a new mode of being, applicable to many areas of life and gathering significant momentum – 'changing the game' as it were. Once again, this 'spirit of open' is closely articulated with collaboration and participation" - Tkacz (2012)

In a move that is challenging for many who see themselves as advocating "the opens" Tkcaz traces these discourses, and particularly openness as "freedom" to the political agendas of libertarian politicians like Douglas Carswell (the British Conservative MP, better known today for first defecting to the UK Independence Party, and then leaving after it successfully campaigned for the UK to leave the European Union) and the US Tea Party movement. He argues that the freedoms being pursued are largely negative in the sense discussed by Holbrook [5]. Openness is generally the effort to be free *from* the restrictions of the status quo.

They are negative in two ways. First they are absolutist in nature, but secondly they frequently make little sense except in the context of the fight against an existing status quo. Open only exists as a contrast to closed and, as Tkcaz traces in many examples, and other critics have noted, the implementation of open leads to it becoming – or being co-opted by – the status quo. The old open becomes the new closed that a new generation will battle against.

Constructed this way, openness can never win. The old "open" is the new "closed". We see this cycle in criticisms of "open-washing", of the power of those groups who control the definitions of open in software, and in the development of open government and open scholarly communications agendas. From offices of open government, to the Open Source Institute and the Public Library of Science, once an advocate of open has achieved stability and a measure of power they become a target, not just for reactionary forces but for their erstwhile allies.

Tkcaz argues that this means that any open agenda always has enclosure as its endpoint, that the underpinning rhetoric, being negative inevitably sows the seeds of its own demise. In his words:

> "If we wish to understand the divergent political realities of things described as open, and to make visible their distributions of agency and organising forces, we cannot 'go native', as a young, anthropologically-minded Bruno Latour once wrote, meaning that we cannot adopt the language used in the practices we wish to study. To describe the political organisation of all things open requires leaving the rhetoric of open behind". - Tkacz (2012)

In this paper I want to argue that while Tkcaz's challenge needs to be taken seriously, that it is not fatal. The key to this lies in understanding how meso-scale political organization, and the inevitable inclusion and exclusion that arises from group formation, interacts with individual (micro-scale) and macro-scale political economics. To do this I will draw on strands of economics, political economy, and cultural studies to seek to show how the oppositional stance and boundary work necessary to define groups can nonetheless be harnessed to aspirations for inclusion and interoperability.

In particular, I want to examine the political and epistemological challenges raised by the inclusion of knowledge-workers from traditionally "peripheral" positions with respect to power centres of traditional Western scholarship. Understanding how a wider range of knowledge-making groups can interact productively and equitably ultimately requires an understanding of how these groups are sustained and how their differing cultures affect their interactions. My aim is to sketch a route towards how three differing framings might be aligned to develop a philosophical underpinning for open agendas. In doing this my focus is on scholarship, but the argument can be developed for much broader application.

2. Cultural Science as a Model

Central to my argument is the need for an enhanced focus of scholarship on the formation, culture, and sustainability of groups. Many arguments founder on the way they move directly from individual micro-economic concerns to a global macro-level argument. The need for "meso-level" analysis in a range of different disciplines has emerged over the last decade. Here I draw on the model of "Cultural Science" developed by Hartley and Potts(6).

Cultural Science seeks to be an evolutionary model of groups and culture. The unit of analysis is a group or community that shares culture. Hartley and Potts name this culture-defined group a "deme" borrowing from both biological (an interbreeding community) and political (the "demos") terminology. The key to the model is that demes do not merely "share" culture, they are *made by culture*. Culture makes the group and the group enacts and articulates the culture.

Culture is not, in this formulation, the aggregate product of the individual actions or behaviours of members of the group but the thing which draws in members of the group through creating common narrative and meaning. Demes can be seen as a parallel concept to Fleck's "Knowledge Collectives" [7] and Ravetz's [8] or Kuhn's "communities" [9].The primary difference lies in the underlying concept of how demes are formed and sustained.

Any given person may be a member of multiple demes, and demes can be embedded within other demes. As an evolutionary model it poses serious challenges of complexity in analysis, although arguably no more than the emerging complexity of selection operating at many different levels in biological systems. The key question for survival of a deme is how effectively it competes with other demes in the environment it finds itself in.

In the book "Cultural Science" [6] Hartley and Potts emphasize conflict between demes. More recently this has been developed to acknowledge that conflict need not be violent or existential (although it frequently is). We argue that it is through *productive* conflict that knowledge (or more generally capacities to act) are created. Demes may build internal capacities that allows them to act *on* other demes, that is to do violence, but they may alternately build capacities that enable them to interact productively with other demes to create new capacities. Without seeking to provide a strict definition at this stage, we can consider shared capacities that span more than one deme to be shared knowledge.

With the Cultural Science model in hand we can make some assertions about demes that do this successfully. They will have aspects of culture that promote productive interactions – *productive conflict* – across demic boundaries. These demes will invoke narratives and norms, and enact and articulate those norms, where they come into contact with differing view points. Such a set of norms might be expected to include acceptable modes of disagreement, agreed approaches to seeking resolution, a commitment to considering – indeed seeking out – alternative perspectives, and approaches for agreeing to disagree where resolution cannot be achieved.

3. An Epistemological Framing of Western Science Knowledge Systems from Cultural Science

If we were to look for an example of such a deme we would likely rapidly arrive at the Western Scientific Knowledge System (WSKS) as an example of a culture that has achieved both continuity in time and dominance over many other systems. We might note the set of cultural elements sketched out above align quite closely to Merton's Four Norms [10] and to other (claimed) normative aspects of Western scientific culture. It could be further noted that the WSKS has a form of fractal organization in which discipline and subject and topic boundaries create opportunities for conflict at many different scales.

Finally, and crucially, we might note that the cultural elements that define the WSKS describe narrative and cultural *aspirations* not necessarily *practice*. Obviously if there is "too much" of a gap between the claims a deme makes about its practice and actual practice then the internal consistencies will build up and lead to failure. However it is also the case that a perfect alignment is not necessary.

This idea that aspiration towards enacting norms and demic narrative can be of value, even when those aspirations cannot be completely achieved, is also developed by Collins and Evans in *Why Democracies Need Science* [11]. They make a different kind of argument for the value of WSKS in democracies and this has tensions with my argument that will discussed below. What we can adopt directly is the flow of their argument that by recognizing that there is value in the group level aspirations we can reconcile the tools and knowledge developed by both "Wave One" and "Wave Two" Science and Technology Studies (STS).

The so-called Wave One of STS uncritically accepted the value of Western Science and sought to examine how this value was created. Merton in particular worked on showing how individual human frailties could be ameliorated by shared norms and strong institutions that supported the creation of scientific knowledge. The overall group dynamic was assumed to be positive and ultimately objective. Wave Two STS critiqued this position noting that group dynamics was clearly related to power, that expertise and stakeholders from outside the academy were often discounted, and that the social context could determine both the process and outcomes of knowledge creation.

To reduce it to slogan form Wave Two showed that groups and institutions could never approach the objectivity and perfection assigned to them by Wave One. In parallel development of philosophy and epistemology that consistently showed that claims of WSKS to generating "truth" could not be demonstrated to be provable. The strong version of these two strands of scholarship led some to the other extreme. Because knowledge and the WSKS institutions supposed to be safeguarding it could not be shown to be provably reliable it follows that we must reject all authority.

Cultural Science, in common with the "Wave Three" proposed by Collins and Evans [11], offers a middle route. First we observe that a recognizable culture and community of WSK creation has persisted over (at least) several centuries. This evolved community has continuity and therefore its supporting culture has continuity. Through analysis of historical and contemporary narratives we can identify some elements of this culture that appear to persist: a valuing of observation, critique of claims, and interestingly an aspiration to civility in resolving disputes. Robert Boyle [12] writes in the 17th century responding to a critic with whom he has had no previous correspondence:

> "[I will answer Linus' objections] partly, because the Learned Author, whoever he be (for 'tis the Title-Page of his Book that first acquainted me with the name of Franciscus Linus) having forborne provoking Language in his Objections, allowes me in answering them to comply with my Inclinations & Custom of exercising Civility, even where I most dissent in point of Judgement." - Boyle (1662)

Many of the social points Boyle makes about practice in his works, including issues of reproducibility and effective communication are in fact much more comprehensible than his actual observations and theories. These are situated in a language and theoretical framework that is largely incomprehensible to us today. Arguably this shows that while the emerging culture of 17th century Natural Philosophy is recognizably the same as that of modern science, the actual knowledge is lost to us as the Thought Collectives, to use Fleck's language [7], have changed too radically.

The details of this idea that there is a recognizable scientific culture that persists over time, and provides sustainability and continuity to a community of practitioners require much work and are beyond the scope of this paper. If the idea is provisionally accepted then we must immediately ask the crucial question, what is it that makes this culture sustainable? Clearly this will be a mix of historical contingency, social context, and power relationships. But the central claim is that elements of the culture have contributed to that sustainability.

I want to suggest that one element that has contributed is a form of *openness*. It manifests historically in different ways but the valuing of observation, and of critique, the importance of effective communication and more recently efforts towards inclusion both in access to the outputs of research and influence over its conduct, can all be read as a valuing the testing of claims by exposing them across the boundaries of the community. We can use the rich literature on the nature of research communities, and their disciplinary splits and divisions, from Fleck [7], through Kuhn and Ravetz [8], but also to Latour [13] and Wave Two STS and indeed on to the work of Collins and Evans [11] on expertise in Wave Three, to understand how the culture of WSKS creates a myriad of hierarchical boundaries across which claims can be tested, while also driving interoperability across those boundaries by articulating shared values.

The Cultural Science framing suggests that Western Scientific Culture is doing two different things. Firstly at a high level, it creates interoperability though shared values. Secondly it drives the creation of new disciplinary groups at all scale levels creating boundaries across which knowledge claims can be tested. We can suggest that this culture, and at least some of the groups it has created, has thrived over time because it is well situated to creating productive conflict where groups interact. From the process of peer review, a managed form of a conflict in which one research group's claims are tested by another, through to the insights that arise when whole disciplines clash as they come into contact, what emerges, as Ravetz noted is more abstracted, more general, and more widely used than what was initially created within the group.

My suggestion is that it is the various forms of openness that act to maximize the productivity of those conflicts. This is not to say that these values are perfectly enacted. As Wave Two STS tells us, scholars are embedded in social contexts and power structures laced with bias, assumptions and exclusions. Indeed, the tension between the necessary boundary work that defines the group, and the productivity of interactions that arise from relaxing those boundaries, is the key to understanding what is being created, what value it has, and to who.

4. The Economic and Political Sustainability of Knowledge Clubs

While I have sketched out an argument for explaining the sustainability of Western Scientific Culture as a whole, to examine the question of how institutions and groups operate we need to examine the sustainability of the overlapping and hierarchical groups that make up the larger deme. We use the term "Knowledge Clubs" [14] to refer these groups that have a commitment to generating knowledge with value beyond their boundaries, which is underpinned by these elements of openness.

The use of "clubs" is deliberate and has two motivations. Firstly, it emphasizes the tension between the definition of boundaries and the need to operate across them. Secondly it draws on the strand of economic theory that examines how groups can sustain the production of collective goods. The narrative for Knowledge Clubs within the WSKS is that knowledge is being created for the good of all. But such goods, Public Goods in economic terms, cannot support the sustainability of the club itself. This implies that the culture-made group is also capable of generating value, or utility, *for the group itself*.

Buchannan's [15] work on the economic sustainability of clubs is central here. Buchannan identifies a class of goods that are neither public or nor private, but are important in sustaining groups. In modern terminology these are goods that are non-rivalrous (they can be shared out without diminishing them) but are excludable (it is easy to prevent non-group members from benefiting from them).

Where a group generates private goods (such as money) that are passed to individuals then engagement is easy to explain. If a group only generates public goods then a classic collective action problem ensues. Such a group can only be sustained if it is non-rational from an economic perspective. While this is by no means impossible – it can be argued that Wikimedia solves the collective action problem for public good creation of a free encyclopedia by relying on donations from non-(economically) rational actors – evidence suggests this can only operate at the extremely large scales where a sufficiently large number of such actors can be found.

Clubs in Buchannan's terms are sustained by this intermediate class of goods, which are termed club goods. I have previously argued that we can see knowledge as such a club good. Knowledge is created by and within groups. It is non-rivalrous, in Jefferson's memorable language "…he who lights his taper at mine, receives light without darkening me", but on its creation it is exclusive and excludable. Firstly, because it is only available to the group, but later the choices of how, and where to communicate it, what language to use, restrictions to access all create forms of exclusion.

We intuitively understand that knowledge held exclusively by a group, whether the scholars who originated it, or the community that subscribes for access to a specific – closed – journal, will not create as much value as it might. This is also consistent with the epistemological model sketched out above, where it is the process of exchange and translation amongst groups, which makes knowledge both more general and more valuable. We therefore have systems, including our systems of scholarly communication, in place that support the process of making knowledge more like a public good, removing various forms of exclusion piece by piece.

This process of investment in making club-good knowledge more public-like, a process of "public-making", however raises the same collective action problem. Why would a Knowledge Club voluntarily give up a good, indeed invest in reducing the exclusivity that allows them to maintain control? Part of the answer is that we are actually quite selective about the modes of control we give up. Traditionally communication through a journal or a book is directed at and accessible (for many different meanings of the word) to a very select, and identifiably demic, group. Part of the answer is one of culture – and as we shall return to, values – that guide our practice as scholars.

Neither of these answers however will suffice for our economic framing. An economic framing suggests that the club is involved in an exchange where it gains something in return giving up exclusivity. That something must be a club or private good and there are in fact a range of these that can be identified. Some of these are quite abstract goods; recognition, prestige, and membership within disciplinary knowledge clubs. Some are much more concrete; jobs, professional advancement, and funding both for further research and personally.

5. An Economic Framing: Institutions as the Underwriter of the Public-Making Exchange

An important aspect of this exchange process to note is that the immediate benefits of the exchange are the more abstract and nebulous ones, recognition and attention. The more concrete, and more widely exchangeable goods take longer. These are individual benefits such as positions and salaries, and for demic groups recognition as a discipline and strand of scholarship that should be a visible part of a research institution. The coupling between public-making and these longer term benefits is something that we believe in. It is a part of our culture. But from an economic perspective there is a distinct risk that the investment in public-making may not in fact pay off.

In financial terms these kinds of risks can be managed if there is an underwriter available. In the research community this underwriting is managed by institutions acting as a – partial – guarantor that the knowledge club's investment in public-making will be convertible in an understood and predictable way into these concrete club and private goods. Institutions, both in the sense of research performing organizations such as universities, but also in the broader sense used by Ostrom [16] of "...the prescriptions that humans use to organize all forms of repetitive and structured interactions", provide the assurances that support the risks of investing in public-making for the knowledge club.

There is, therefore tension at the heart of our institutions. Their purpose is (in part) to promote public-making, but they do this through acting as a guarantor in a transaction which provides excludable goods. The university itself is an exclusive club and needs to be to support the realization of benefits that arise from prestige and authority. To be predictable and therefore effective as guarantor institutions must necessarily be conservative in both the forms of public-making they support and recognize and in the rewards they award as a result of those activities. But to realize the full benefits of public-making they may need to be adaptable and even radical in a rapidly changing world.

Ostrom [17] showed that the way to understand institutions that resolve collective action problems is to see them as developing through a process of evolution. And that coordination at large scale required the development of hierarchical layers of organization. In turn the development of these layers provides stability and resilience to the system as a whole. All of this emphasizes that our institutions (in the sense of research organizations) should be expected to be resistant to change – should in fact be designed to be stable.

This analysis has implications that spread far beyond scholarly communications. In its role as a guarantor for the provision of club goods, which have as a core characteristic exclusivity, the institution is continually policing boundaries. This means working to protect the identity of the existing clubs, including their historical lack of diversity, it means policing the boundary of what counts as "scholarly" in terms of both work and outputs, and it means a focus on protecting existing and historical markers of prestige and authority.

As scholars we also reinforce this backwards looking boundary work whenever we rely on our research organizations to act as the guarantor of benefits that we exchange for public-making. Our continuing engagement with "traditional" modes of public-making and scholarly communication are both driven by our acceptance of the social contract we have with our institutions and act to reinforce that system.

As is often the case with economic arguments, this one appears to arrive at a profoundly depressing conclusion. Not only must we expect, indeed rely on, our institutions to be conservative, but this appears to open up a gaping hole between the harsh economics and the value of an open culture that the epistemological argument implies. The Cultural Science framing implies diversity is key to generalizing knowledge, whereas the economic argument seems inevitably to point to institutions that will slow the increase in diversity, both of activities and participants.

Arguably framing the opportunities presented by developing technologies as "new" forms of scholarly communication that are "different", aligning ourselves with the oppositional discourse that Tkacz [4] describes, is counterproductive. This offers a potential solution, that is to situate and to design these "new" practices as simply a more effective expression of old values. Successes in innovation in scholarly communication and open practice are often associated with small changes, with far superior but more radical opportunities often failing. Can we avoid the problems of conservatism, or at least speed up the uptake of new tools and practices, and also the oppositional discourse of openness by describing openness as an *old* value?

6. Framing Openness as a Core Value of the Academy

The economic analysis above paints a very harsh and transactional picture, but the reality is of course more complex. The institutions that are taking the role of guarantor spread beyond our research organizations to those broader "institutions" that are part of our research culture. Indeed, we can tie the sustainability of Western Science culture in part to its role in sustaining the cultural institutions that underwrite this exchange of knowledge. That is, our reliance on this exchange as scholars is underpinned by our self-identification as scholars, our identification with the demic group. It is deeply tied to the *values* that we hold. In these final sections I will argue that it is through a framing of openness as a value core to Western Science culture that we can both work for change within our institutions as well as enhance the diversity of our communities and therefore the value of the knowledge we create.

Shapin and Schaeffer [18] in their dissection of the historical conflict between Robert Boyle and Thomas Hobbes and the founding of the UK's Royal Society describe Boyle as deploying three technologies. The three technologies; the material technology of the experiment, the literary technology of printing and dissemination, and a social technology – the scientific culture and institutions in our terms – that defined the interactions of scholars. These various technologies underpinned claims made by Boyle and other natural philosophers to openness and similar claims are made to this day. Openness to criticism and critique, openness to contributors from any place or walk of life, and openness through the accessibility of printing and disseminating accurate descriptions of the experiments.

Shapin and Schaeffer's important contribution is to critically examine these claims and to show that in practice Boyle and others involved in defining and creating the culture and institutions of science that continues to this day fell a substantial distance short of their aspirations. Boyle sharply circumscribed what he would accept as legitimate criticism, claims and evidence from those of more noble birth were to be preferred over that from commoners, and access to the halls and demonstrations of the Royal Society were certainly not open to all. Indeed, it is only in the past 25 years that a ban on women (at least those who are not Fellows) entering the headquarters of the Royal Society was lifted.

Here we see again exactly the tension that has played out through this discussion. A claim of openness, and a narrative that this openness sits at the core of the value system, that is not quite realized in practice. The building of institutions that seek to enhance openness – the Royal Society holding formalized meetings, open to members, in the place of private demonstrations – that are nonetheless exclusive. Membership of the club, whether the Royal Society or other National Academies, has always been a marker of prestige and authority, even as the actual criteria for membership have changed radically over the years. Yet what is passed down to us today, is less that exclusive gentleman's club and more the core values that it sought to express.

Move forward 200 years from the 17th to the mid-19th century and a debate was raging in the United Kingdom about who could contribute to the conduct of science.

Lightman [19] reveals what might appear to our 21st century eyes as a startlingly modern debate on the interest "that not alone scientific readers, but those of every class, [...] to approach the source from whence this species of knowledge is derived". Lightman describes the growth of popular science journals to meet this demand. It is perhaps a sign of the strength of the tension we are discussing that the most visible survivor of this growth is the journal *Nature* which has been so entirely co-opted by our scholarly culture as an institutional signal of internal club prestige, that it can stand symbolically for the entire system of journal hierarchies.

In an illustration that progress is clearly not linear Lightman also discusses the positioning of Darwin – whose beard today stands as a (not particularly inclusive) symbol of a professional scientist – as a demonstration that amateurs can contribute to science. Lightman quotes Grant Allen, a 19th century popularizer of science describing Darwin as "merely an amateur, a lover of truth, who was impelled by curiosity". The professionalization of the academy through the 20th century alongside the celebration of Darwin as a key figure in the history of science seems to have necessitated an assumption of his place as a "real scientist". If we are to aspire to be part of the club that included Darwin then we must necessarily place him in that club. Arguably this was a backwards step in a trajectory of gradually implementing greater openness. Lightman notes that the "...appropriation of [...] Darwin as [an] iconic figure [...] served to undermine the participatory ideal of the 19th-century popularizers and reflected the increasing power of professionalization". That is, the evolution of the professionalized institutions, that stabilize and allow the scaling of the culture of Western Science created exclusion, even in the way that we create and describe iconic figures.

It would be straightforward to follow the gradual opening up of aspects of our institutions and culture through the 20th and 21st century. Examples could be given from increasing access to tertiary education, the public funding of research, through open access, the shift from "public understanding of science" through "public engagement" to "responsible research", to issues of data availability and citizen science. However my point is to establish the deep roots of this agenda. Despite, or even in some cases because of, the limitations in putting it into practice, the idea that critical contributions to scholarship will come from outside has persisted. Indeed a case for the inverse can be made, that the culture of Western Science has persisted precisely because a commitment to openness, to public-making, is one of its core values.

7. An Aspiration to Openness as a Conservative Position

I began by noting that openness refers to many different things, and that as many others have noted, that the narrative associated with this variety is frequently one of new-ness, of technological possibilities, and of opposition to a status quo. As Tkacz notes this can lead to a cyclic inevitability as openness eats itself and becomes the new status quo, the new establishment.

I want to flip this on its head. In Boyle's writings we see the concern for completeness of description, for reproducibility and for a commitment to observations, wherever they come from as the final arbiters. In the 19th, and again in the 20th and 21st centuries we see movements arise in which contributions are sought from anyone. In Merton's norms [10] of communalism and universalism, Popper's conception of falsifiability [20], and Kuhn's idea that scientific revolutions are precipitated by the build up of external information [9], even in Latour's model for the gradual expansion of the collective [13] we see repeated attempts to articulate the importance of openness to claims and ideas from the outside as a core part of the social activity of science.

Clearly this value is quietly ignored at least as frequently as it is found in practice, but the aspiration is a common thread. Indeed the institutionalization of imperfection may be critical in solving the economic problem of sustaining knowledge-making clubs that choose to invest in public-making. The argument made here has only provided the barest sketch of how Knowledge Clubs interacting may be engaged in both economic exchanges and productive general-knowledge producing conflict. If the most significant insights come from across boundaries then the boundaries themselves are also of value. A deeper analysis may provide a route to identifying the ways in which this tension can be managed both to create value in the economic sense and to maximize the public-good nature of generalized knowledge.

It is therefore the aspiration to openness, and its adoption as element of the identity and core values of the researcher, its centrality to our culture, that provokes us to attempt to move across boundaries and to create knowledge. That “full openness” or “total inclusion” can never be achieved is the consequence of an imperfect world. The aspiration to seek it still has value. In this sense the argument aligns with the claims for Elective Modernism made by Collins and Evans [11]. However my conclusion is diametrically opposed.

Collins and Evans state that the scientific community must be protected so that its value system, its culture in our terms, can operate without disturbance. I argue here that disturbance is fundamental to its function, that the process of generalizing knowledge requires that new efforts are constantly made to break down barriers and reduce exclusion. Nonetheless the institutions that underwrite the exchanges fundamental to public-making do need protection. Understanding how they can change at an optimal pace remains a challenge.

Part of the answer may lie in the problem. It may be that an argument can be made that this tension is fundamental, that progress towards greater openness is a return to core values, that such progress must underpin any claim of real progress arising from Western Science. In that sense situating openness as a profoundly conservative position may be a viable political move. In the end the answer is not that openness is any one thing, it is that it is many different expressions of one underlying process. That it proceeds through cycles of change, institutionalization and reaction is then unsurprising. And if that is correct then we can start to pull the threads together that will allow us not merely to respond the institutions and culture that we have as they evolve around us, but to design them.

If this is true then we are perhaps living in a time of unprecedented opportunity for science and for scholarship. There are profound challenges to adapting our institutions to interact productively with differing knowledge systems, but we are perhaps for the first time well placed to do so. By understanding how tension between boundary work – and its exclusionary tendencies – and the value of diverse perspectives we may be able to improve, by design, on our institutions. If we can develop a narrative thread within our culture that this is merely the extension of an ongoing process that has served the academy well, then we arguably make this gradual and high imperfect progress a highly conservative position. This may offer us the best opportunity to accelerate the progress we are making on access, inclusion, and diversity and build a more generally valuable, and accessible knowledge system that truly includes the insights and perspective of those beyond the walls of the academy.

References

[1] Fecher B, Friesike S. Open Science: One Term, Five Schools of Thought. In: Bartling S, Friesike S, editors. Opening Science [Internet]. Heidelburg: Springer; 2014 [cited 2016 Jan 19]. p. 17–47. Available from: http://papers.ssrn.com/abstract=2272036
[2] Pomerantz J, Peek R. Fifty shades of open. First Monday [Internet]. 2016 Apr 12 [cited 2017 Jan 16];21(5). Available from: http://firstmonday.org/ojs/index.php/fm/article/view/6360
[3] Coleman G. The Political Agnosticism of Free and Open Source Software and the Inadvertent Politics of Contrast. Anthropological Quarterly. 2004 Sep 3;77(3):507–19.
[4] Tkacz N. From open source to open government: A critique of open politics | ephemera. ephemera: theory & politics in organization. 2012;12(4):386–405.
[5] Holbrook JB. We Scholars: How Libraries Could Help Us with Scholarly Publishing, if Only We'd Let Them. In: Bonn M, Furlough M, editors. Getting the Word Out: Academic Libraries as Scholarly Publishers. ACRL; 2015. p. 43–54.
[6] Hartley J, Potts J. Cultural Science. London ; New York: Bloomsbury 3PL; 2014. 266 p.
[7] Fleck L. Genesis and Development of a Scientific Fact. New edition edition. Trenn TJ, Merton RK, editors. Chicago u.a: University of Chicago Press; 1981. 232 p.
[8] Ravetz J. Scientific Knowledge and Its Social Problems. Reprint edition. Oxford: Oxford University Press; 1971. 449 p.
[9] Kuhn TS. The Structure of Scientific Revolutions. 2nd Revised edition edition. Chicago, Ill: University of Chicago Press; 1970. 222 p.
[10] Merton R. Science and technology in a democratic order. Journal of Legal and Political Sociology. 1942;1:115–26.
[11] Collins H, Evans R. Why Democracies Need Science. Cambridge, UK ; Malden, MA: Polity Press; 2017. 200 p.
[12] Boyle R. A defence of the doctrine touching the spring and weight of the air propos'd by Mr. R. Boyle in his new physico-mechanical experiments, against the objections of Franciscus Linus ; wherewith the objector's funicular hypothesis is also examin'd, by the author of those experiments. Oxford: Thomas Robinson; 1662.
[13] Latour B. Politics of Nature: How to Bring the Sciences into Democracy. Cambridge, Mass: Harvard University Press; 2004. 320 p.
[14] Potts J, Hartley J, Montgomery L, Neylon C, Rennie E. A Journal is a Club: A New Economic Model for Scholarly Publishing [Internet]. Rochester, NY: Social Science Research Network; 2016 Apr [cited 2017 Apr 8]. Report No.: ID 2763975. Available from: https://papers.ssrn.com/abstract=2763975
[15] Buchanan JM. An Economic Theory of Clubs. Economica. 1965;32(125):1–14.
[16] Ostrom E. Understanding Institutional Diversity. Princeton: Princeton University Press; 2005. 376 p.
[17] Ostrom E. Governing the Commons: The Evolution of Institutions for Collective Action. unknown edition. Cambridge ; New York: Cambridge University Press; 1991. 298 p.
[18] Shapin S, Schaffer S. Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life. With a New introduction by the authors edition. Princeton, N.J: Princeton University Press; 1985. 448 p.
[19] Lightman B. Popularizers, participation and the transformations of nineteenth-century publishing: From the 1860s to the 1880s. Notes Rec. 2016 Dec 20;70(4):343–59.
[20] Popper KR. The logic of scientific discovery. London: Hutchinson and Co; 1959.

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-18

Framing a Situated and Inclusive Open Science: Emerging Lessons from the Open and Collaborative Science in Development Network

Rebecca HILLYER[1,a], Alejandro POSADA[b], Denisse ALBORNOZ[b], Leslie CHAN[b] and Angela OKUNE[c]
[a] *University of Stellenbosch*
[b] *University of Toronto Scarborough*
[c] *University of California, Irvine*

Keywords. Inclusive open science, situated science, collaboration, sustainable development, contextualised openness, knowledge diversity

Abstract. What is open science and under what conditions could it contribute towards addressing persistent development challenges? How could we re-imagine and enrich open science so that it is inclusive of local realities and a diversity of knowledge traditions? These are some of the questions that the Open and Collaborative Science in Development Network (OCSDNet) is attempting to answer. In this paper, we provide the rationale and principles underlying OCSDnet, the conceptual and methodological frameworks guiding the research, and preliminary findings from the network's twelve globally diverse research projects. Instead of a "one-size-fits-all" approach to open science, our findings suggest that it is important to take into account the local dynamics and power structures that affect the ways in which individuals tend to collaborate (or not) within particular contexts. Despite the on-going resistance of powerful actors towards new forms of creating and sharing diverse knowledge, concluding evidence from the twelve research teams suggests that open science does indeed have an important role to play in facilitating inclusive collaboration and transformatory possibilities for development.

1. Introduction

The idea of 'open science' has gained momentum over the past few years, emerging alongside other 'open' initiatives - including open access, open government, open source, open data and others [1]. A common conception of open science is the opening of the entire research cycle - from designing the question and methods, to collecting and analysing data, through to the communication and dissemination of findings [2] [3]. In principle, these concepts collectively strive for an environment that facilitates opportunities to participate in knowledge production and circulation for people who have been historically excluded. As such, the growing momentum around open science provides a key opportunity to reflect on and reimagine the ways in which we

[1] Corresponding author Rebecca Hillyer, email: becky.hillyer@gmail.com

understand and conduct science, and how knowledge-making could be made fairer and more inclusive of diverse ways of knowing.

To date, however, the majority of action and discussion on open science has been dominated by Western actors and institutions, with a tendency to focus on the tools, infrastructure and cost models of producing knowledge 'openly' [4] [5] [6], with less focus on the underlying power structures that tend to determine who is or is not able to participate in knowledge-production processes, and for what aims [7] [8] [9] [10] [11].

The Open and Collaborative Science in Development Network (OCSDNet) is an international research network, launched in 2015, to address the fundamental question of whether and how open science has the potential to contribute to the achievement of development goals and opportunities [12]. Funded by IDRC in Canada and DFID in the UK, with coordination support from Kenya's iHub[2] and the University of Toronto, OCSDNet is composed of twelve international research teams[3] throughout Latin America, Africa, the Middle East and Asia. The teams are from highly diverse disciplinary backgrounds, including law, education, climate change, the maker movement, intellectual property rights, biodiversity, health and environmental conservation. Over the course of two years, and using an array of diverse research methods within distinctly different contexts, each team explored the challenges and opportunities for an open and collaborative science, and the potential of open science to facilitate fair and sustainable development.

OCSDNet recognises that throughout recent history, processes of knowledge production and dissemination have been shaped and solidified by a privileged and exclusive set of actors, ultimately influencing the way in which the world understands 'valid' and 'legitimate' scientific knowledge and research. This limited representation of knowledge leads to an incomplete understanding of the world and of the issues affecting local populations [10] [13]. Unchallenged, this system will continue to exacerbate knowledge and research inequalities, with serious consequences for sustainable and equitable development [14].

As the projects in the network will have reached their completion by June 2017, this paper provides a preliminary analysis of some of the key lessons that have shaped the ways in which OCSDNet members have come to re-imagine the potential of open science to transform processes of knowledge production and contribute to sustainable development. The paper will begin with a discussion of the network's background, including the methodologies that have guided research conducted between 2015 and 2017. This will be followed by an overview of the ways in which individual projects have contributed towards co-constructing a new and more nuanced understanding of open science.

Some projects have contributed towards refining open science at the 'grassroots' level of sustainable development through the implementation of small-scale citizen science projects at the community level. At the same time, others have contributed towards the reimagining of the field through a case-study analysis of existing, longer-term open science initiatives, including the sustainability challenges and social tensions that tend to arise as openness 'scales up' within or between institutions and their networks. Finally, other research teams have sought to apply network-defined open science principles within their unique contexts to develop new tools and frameworks

[2] See: https://ihub.co.ke/

[3] See Annex 1 in this paper for a list of project names and keywords, or visit www.ocsdnet.org for full project descriptions

for understanding the potential of open science to contribute towards complex development and societal challenges.

Despite the diversity of projects within the network, many overlapping findings emerge, which demonstrate the importance of re-imagining open science in the context of complex development issues. Through the application of a contextualized or 'situated' approach towards defining and practicing open science, this paper concludes with the need to focus on making science more inclusive of a diverse set of actors and their epistemic traditions.

2. Network Background & Methodologies

The conceptual framework that informed the initial research questions for the network was based on the Institutional and Development Analysis (IDA) framework developed by Elinor Ostrom and colleagues over several decades of work on natural resource commons and their governance. Ostrom's work challenged the conventional wisdom around the need for government regulation of public resources (such as forests, fisheries, etc.) in order to attain sustainability and benefit sharing [15] [16].

In more recent years, this framework has also been applied to knowledge as a "commons," which cross-cuts national and disciplinary boundaries [17] [18]. Taking into account the unique attributes of knowledge and information that are distinct from natural resources, Frischmann, Madison and Strandburg [19] modified the IAD framework into a "Knowledge Commons framework" to aid other researchers with empirical research on different forms of commons. The framework provides a number of guiding research questions around the nature of the community in question, the kind of the resources in use, the existing institutional arrangements, and the interactions that take place within the community. Within OCSDNet, these questions were used and adapted to structure our data collection activities with the sub-projects, by including them in monthly and annual report templates, semi-structured interview questions and general group discussions throughout the network's duration.

While using this framework as a guideline for collecting data from research teams, observations around team and network working dynamics were also drawn from exchanges within a closed Google Group established for network communication, as well as offline network dialogues, social media discussions (e.g. Facebook groups and Twitter) and formal academic communications - including publications and conference presentations. Project teams were encouraged to share events, resources and best practices as part of the field and network-building exercises. It should be noted here that the OCSDNet Research Coordination team (consisting of five members positioned variably in five countries around the world) also participated in similar processes of reflection and discussion, around their own perpetuation of power dynamics within the network.

Along with these more traditional data collection activities, OCSDNet explored the potential of participatory, consensus-building exercises through the design of an "OCSDNet Manifesto" - a document that has attempted to consolidate the shared understanding of what Open and Collaborative Science offers to scientific research and development.[4] These discussions and the seven consequent "open science principles" that were developed, have had a substantial effect on the way in which many projects

[4] see Albornoz et. al, [21] for ELPUB for more information about the manifesto creation process

have assessed their own findings and ways of working.

The various mixed methods described above, guided by an iterative process of reflection towards our original conceptual framework, has generated a large volume of qualitative data and media artifacts. Much of this data has been analysed iteratively, over the course of the network's duration, but the final analysis continues to take place through qualitative-data coding processes to uncover themes and ideas that allow for greater comparison between diverse and complex projects.

The next section will discuss some of the key, preliminary findings that have emerged from the twelve individual research projects, with the intention of presenting a framing of open science that extends beyond a discussion of the 'tools' and 'cost models' associated with working openly. Instead, all cases look at the innovative ways in which OS principles can be applied to complex development questions and scenarios, with a focus on the socio-cultural contexts that have the potential to enable or curtail the potential for open science as an effective tool for achieving sustainable development objectives.

3. Emerging Lessons from OCSDNet Projects

An advantage afforded by the diversity of project membership and contexts is we are afforded the opportunity to interrogate the manifestation of open science practices at varying scales, from the grassroots, to the institutional, regional and national levels. With this in mind, we have grouped the projects into three thematic categories for analysis:

1) Practicing OS at the 'Grassroots;' *(4 projects)*
2) Analysing existing OS projects in the context of development *(2 projects)*; and
3) Exploring the potential of Open and Collaborative Science through new Tools and Frameworks *(6 projects)*

The complex discussions that OCSDNet members have had around defining 'development,' are beyond the scope of this paper. However, it should be noted that network members have broadly agreed on a notion of development that encapsulates Appadurai's "Right to Research," [20] which acknowledges that all humans have the capacity to aspire towards imagining their own knowledge and futures. Appadurai's work echoes Amartya Sen's Human Capabilities Approach, which posits human development as the process of enlarging a person's "functionings and capabilities to function, the range of things that a person could do and be in her life,'' as expressed in terms of one's agency to exercise ''choices'' [22]. The purpose of development is thus to improve human lives by expanding the range of things that a person can be and do, such as to be well nourished and be healthy, to be knowledgeable by taking part in knowledge making, and to actively participate in community life. In this regard, the Latin American concept of *buen vivir*[5] ("the good living") has also informed the network's conceptual framework, as has the ancient African concept of *Ubuntu* - a philosophy that celebrates the strength of humans working and living in community with one another [23]. Taken together, these concepts comprise a framework of development that positions human beings as agents, working towards common goals,

[5] For a description of *buen vivir*, see [26]

and using the tools and forms of knowledge that are most relevant to their unique socio-cultural contexts[6].

3.1. Practicing Open Science at the 'Grassroots'

"Grassroots" development, well-known since the 1990's, refers to development research and activities that tend to focus on the community or micro-level context. The intention of this approach is to facilitate and pursue a 'bottom-up' approach to development, in which ordinary people are directly involved within activities meant to improve their lives [24]. This approach arose largely due to the growing opposition against 'top-down,' macro-development strategies that tended to dominate development discourse and practice during the 1980's. In particular, these macro-level approaches tended to dismiss local contexts and prevailing power structures, and hence failed to procure anticipated outcomes [25].

We borrow the notion of 'grassroots' development here due to its similarities to the localized and small-scale citizen science-initiatives that are present in four OCSDNet projects. These projects allow for a deeper understanding around the possibilities of initiating, managing and assessing small-scale open science initiatives that demand minimal funding, and which can be initiated, planned and completed in a relatively short time frame. They likewise permit a unique, 'insider' perspective regarding the day-to-day negotiations and complexities associated with the practice of open science, as well as a chance to compare dilemmas and opportunities across contexts. Most importantly, they provide the opportunity to assess whether a small-scale open science project-approach can have positive implications for sustainable community development.

Table 1. Practicing Open Science at the 'Grassroots'

Project Name	*Key Words*
Water Quality and Social Transformation in rural Kyrgyzstan	Kyrgyzstan, rural communities, citizen science, environmental conservation, water quality, participatory action research, open science motivation, teachers and students
Community-driven environmental conservation in Costa Rica and Colombia	Costa Rica, Colombia, participatory action research, citizen science, Model Forests, human capabilities, adaptive capacity, sustainable development, biodiversity
Water quality and community velopment in Lebanon	Lebanon, Citizen science, participatory research, community-based environmental management, water quality, empowering conservation, bottom-up policy making
Open Science Hardware for Development in Southeast Asia	SEast Asia, open science hardware (OSH), transnational networks, little science, citizen science, do it yourself (DIY), Indonesia, Thailand, Nepal, tools, participation, tinkering, Right to Science

In reference to *Table 1*, all four grassroots open science projects position the concept of 'citizen science,' as central to their methodologies and conceptual

[6] As our conceptual framework draws heavily from a diversity of thinkers from different fields, we have put together an annotated bibliography and reading list to allow interested readers to go deeper into the literature that we consulted:
https://docs.google.com/document/d/10g0U2_aNsOWCSNulfsw3Ea0TEhbx18JoCL8I7a8QLZ8/edit

framework. In general, ‘citizen science’ is a broad term that has come to convey an array of meanings and actions, depending on the context. Perhaps the most common conception of a ‘citizen scientist,’ is an individual who voluntarily spends time contributing towards the crowd-sourcing of data (often using online tools and infrastructure) as part of a larger research investigation with predefined questions and objectives. For instance, Silvertown [27] refers to a citizen scientist as “a volunteer who collects and/or processes data as part of a scientific enquiry;” while Cohn [28] defines them as “volunteers who participate as field assistants in scientific studies.” While these forms of citizen science may indeed have important outcomes for knowledge production and development, there tends to be less focus on the individual as a local expert, or co-researcher, who is able to have input in the design of the research process, questions and data analysis.

Three of the four teams listed above have positioned *transformation* and/or *empowerment* as key objectives within their projects. On the one hand, while citizens are involved, in various ways, within processes of data collection, they also have the opportunity to participate in the identification of key local challenges, and perhaps to provide input on how and where data is collected, as well as any consequent actions that should be taken once information is collected and assessed.

In Kyrgyzstan, the OCSDNet research team worked with rurally-located school teachers and students to design an experiment to test local water quality, after the communities acknowledged that water pollution is a significant issue within the area. This was by no means simply an act of “designing and rolling out” an experiment, but instead involved complex discussions with teachers, students and research organisations that focused on *who* should be able to participate in scientific knowledge production and for what purposes. Throughout the duration of the project, teachers and students began to re-define their ideas of who a “scientist” is, and what scientific research could entail. Similar findings were encountered by the research team in Lebanon, who recruited a group of local volunteers (all of whom happened to be women), to conduct water-quality testing in fifty rural villages. In the end, not only were citizen scientists feeling more informed about water issues in their respective areas, but felt empowered, through their acquired knowledge, to begin making demands on government to pay attention to water-quality issues that affect entire communities.

Both of these projects highlight instances where, given the opportunity to participate in processes of creating and analysing locally relevant knowledge, communities who are (to varying extents) ‘marginalised,’ can use their knowledge not only to address a pertinent local challenge, but also to alter the way that they feel about themselves, as active and informed citizens within their respective communities. In particular, given the notable voluntary participation from women (in Lebanon) and female school children (in Kyrgyzstan), our research may suggest that a local, exploratory approach to open science could have implications for increasing the representation of women and girls within scientific initiatives.

In the cases of two projects in Costa Rica/Colombia and South East Asia, citizen science was explored and negotiated in different ways. In the Latin American case, the team sought to bring together local community members and academic researchers to discuss and negotiate how the “Model Forest” approach to sustainable development may be adapted and negotiated in the context of open science. While the project did not employ the collection of formal, quantitative data, the input from both parties was used to observe opportunities for collaboration and knowledge-sharing towards achieving local development goals. In the end, seven locally-driven open science initiatives were devised around the theme of local environmental adaptation - including a farming agroecology network, rainwater harvesting, a tree nursery and ecotourism awareness.

On the other hand, within the South-East Asian project, a much more subtle version of citizen science was seen to facilitate and assess project activities, through what the team refers to as '*a small science.*' Through this approach, science (and particularly the design of new tools and hardware) was envisioned as a gently facilitated process of creative engagement between diverse participants (including artists, designers, students, teachers, etc.), oftentimes without a tangible social or development objective in mind. The idea was that through bringing diverse individuals into a shared, physical space and with access to a wide range of tools and materials, there could be the potential to stir and foster creative innovation beyond the scope of an intricately planned workshop agenda.

These four projects highlight the deep nuances of characterizing 'citizen science' in the context of open science in development, the specifics of which vary depending on the theory of change used by the individual project. In all instances, the framing of who constitutes a 'citizen scientist,' and what role they play within a given project has important implications for assessing who has power within the scope of the research cycle, and hence the power to create relevant, local knowledge. To varying degrees, all four of these OCSDNet projects were designed to provide increased power and opportunities for regular citizens to participate in processes of knowledge creation and discussions that could have implications for development challenges influencing their lives. Importantly, each project sought to challenge the traditional idea of who constitutes a 'scientist,' and to reimagine the tools and processes required for legitimate scientific discovery and local innovation. Finally, all of these projects position citizens as agents of change with important, pre-existing expertise, rather than merely as volunteers involved in data collection for a pre-established project agenda.

3.2. Analysing Existing Open Science Projects

Along with developing an 'on-the-ground' perspective of grassroots open science initiatives discussed above, two projects within OCSDNet sought to examine, at a meta-level, the challenges and opportunities for larger, complex and ongoing open science initiatives that extend beyond the two-to-three year funding scope of the network. These projects assist in extending the perspective of the network towards a more objective 'outsider' perspective regarding the complexities of initiating, sustaining and scaling-up open science practices in the longer term. Given that open science is a relatively new field that continues to be defined and taken-up in different ways and in different contexts, these projects provide valuable insight regarding the complexities and longer-term challenges of existing open science projects in the Global South, both for individuals and institutions, as well as the practical implications that these challenges could have for achieving sustainable development goals.

Table 2. Analysing Existing Open Science Projects

Project Name	*Key Words*
Evaluating Open science e-infrastructure in Brazil	Brazil, virtual herbarium, botany, interdisciplinary collaboration, e-database, open science infrastructure
Negotiating Open Science in Argentina	open science, Argentina, negotiating openness, opening process, boundary objects

In the Brazilian case study highlighted in *Table 2*, the OCSDNet research team sought to understand how diverse users were accessing a Brazilian-based open access e-database and for what purposes; as well as documenting any benefits to data providers themselves. Known as a 'virtual herbarium,' the open access database consists of pooled botany and fungi records from a large network of Brazilian research institutions. The initial idea behind the virtual herbarium was to create a centralised hub of information that could be easily accessed by any individual interested in research on Brazil's rich and diverse plant and fungi kingdoms. The herbarium was initiated in 2008 and is currently composed of 106 associated national herbaria, 25 herbaria from abroad, and 20 other herbaria that are not directly associated to the project but contribute their data through a shared provider. As a whole, the e-infrastructure combines over 5.5 million data records from 191 datasets and more than 1.4 million images [29].

The OCSDNet research team encountered impressive results around the usage of herbarium records, documenting not only the surprising frequency with which data is accessed and used (1.7 *billion* records accessed between 2012 and 2017), but also the diversity of the users, who ranged from Masters and PhD students, to government representatives, local research organisations, NGO workers, the private sector, and younger students. Importantly, 94% of users were residents from Brazil, highlighting the immense importance of providing access to local knowledge through accessible, online tools and in local languages.

Perhaps most surprising for the team, however, was around the complex negotiations and cultural shifts that needed to occur, throughout the years, to ensure the project's success. For instance, while preliminary requirements for data providers demanded complete openness, through a series of negotiations, the parameters have since changed to allow data providers the flexibility to decide, on their end, which records are made openly available and how. On the other hand, all decisions regarding the technological aspects of the network's architecture and e-infrastructure are left to the technologists. Thus, in this case, it was important for key actors to have some degree of power regarding their contributions towards maintaining the herbarium; while simultaneously having appropriately defined roles to allow for efficient, longer-term planning and governance of the infrastructure. Communication, transparency and participation, according to the team, were indispensable for building trust, understanding and ownership amongst all actors.

In an Argentinean study, the team chose to assess four locally initiated open science case studies encompassing a broad range of disciplines, namely: the New Argentinean Virtual Observatory - NOVA (*astronomy*); Argentinean Project of Monitoring and Prospecting the Aquatic Environment - PAMPA2 (*limnology*), e-Bird Argentina (*ornithology*), and the Integrated Land Management Project (*Geography, Chemistry and Environmental Science*). The team sought to understand *what* is being "opened" within the specific cases; *how* it is being opened; and *who* is participating in the opening process. The team was particularly interested in understanding the consequences of 'scaling up' open initiatives, noting that while some institutional models of open science do exist, there is less emphasis on the initiation of openness at a 'laboratory level,' and how the transition from the laboratory to institutional level occurs in practice.

Through their analysis, the Argentinian team noted that while the four case studies employed different methodologies and actors for the collection of data, all had the overlapping consequence of making collected data more accessible to the general public. Furthermore, their findings suggest that as each open science initiative progressed to encompass different aspects of the research cycle (project planning to data collection to analysis to dissemination, etc.), there was a need to reflect on and reconsider the tools, resources and infrastructure required for each new phase. From a sociocultural perspective, this process of transition puts new strains on open science practitioners, as each new phase may entail a new form of contradiction and hence negotiation with traditional institutional norms and structures.

Looking at both the Brazilian and Argentinian case studies, several key lessons emerge regarding the complexities of sustainable, longer-term open science initiatives. First and foremost, open science is not merely the design of new "tools" that can allow for easier collaboration between individuals. Instead, an effective open science demands complex negotiations around roles and responsibilities; principles and priorities; timelines and resources. It may require new and innovative thinking at each stage of the research cycle and a reflection on how such practices may coincide with existing cultural and institutional norms. From a practical perspective, large-scale initiatives also imply a comprehensive consideration of long-term funding - particularly when multiple institutions are involved. Indeed, despite the success of the Brazilian virtual herbarium and its deployment since 2008, the infrastructure is still described as a 'project,' since the sustainability of future funding is by no means a guarantee [29].

From a development perspective, large, longer-term open science projects have an important role to play in providing the general public with knowledge and information that is useful for informing local decision making and determining development priorities. However, at the same time, due to a lack of access to viable, long-term funding and resources experienced by many Southern institutions, feasibility and timelines are critical considerations and potential hurdles to the success of such initiatives. Given the "project-based" timeline of the majority of funders, it may be difficult to plan and implement long term and larger-scale open science initiatives that seek to tackle complex development challenges and that inherently demand flexibility, reflection and adaptation at all stages of the research cycle. Thus, funding institutions who are interested in seeing real impact around open science in development initiatives must take these considerations into account while defining their priorities and criteria for funding allotment.

3.3. Exploring the potential of Open and Collaborative Science through new Tools and Frameworks

Beyond the creation and analysis of open science initiatives, other OCSDNet teams have taken the perspective of the network a step further by imagining the potential of open science through a variety of new tools and frameworks. As mentioned earlier in this paper, over the course of the past two years, all members of the network have been actively involved in a participatory process of designing an 'open science manifesto,' which presently consists of seven key principles that reflect the network's understanding and practice of open science, as a collective. Beyond the importance of this document for establishing trust and understanding between network members, it also provides a useful lens through which to consider localised development challenges, and whether an open science approach could be an effective way towards addressing such challenges. Importantly, the manifesto seeks to encourage plural forms of knowing and the collaboration of diverse actors across disciplines, languages and geographic boundaries.

In this regard, through the course of their research, two teams applied network principles of OS towards the creation of new, practical and usable 'tools' to negotiate complex development issues within their specific contexts, while four other teams used these principles to develop new ways of framing the discourse and possibilities around OS for addressing particular local challenges.

Table 3. Exploring the potential of Open Science through new Tools and Frameworks

Project Name	*Key Words*
Researcher contracts for Indigenous knowledge in South Africa	South Africa, indigenous knowledge, climate change, intellectual property rights, research contract, decolonising research methodologies, terra nullius
Disaster Management Tools for Small Island States	Disaster recovery plans, Small Island Developing States, Design Science, regional collaboration, knowledge broker artifact
Commercialisation & Open Science in Kenya	Kenya, IP laws, open science, universities, private sector, collaboration, research partnerships, commercialisation
Sustainable development and the potential for OCS in Brazil	Ubatuba, social change, sustainable development, potential of open science, participatory action research, diverse actors
Social problems and the potential of OS in Latin America	Latin America, openness, non-hegemonic countries, social problems, collaborative science, cognitive exploitation
Building Open Science Social Networks in West Africa & Haiti	West Africa, Haiti, open science networks, science shops, open repository, open research, participatory research, cognitive justice

In the case of tools, an OCSDNet team in South Africa sought to employ the principles of open science to negotiate a community-researcher contract in order to safeguard the knowledge of indigenous communities around climate change and other topics. Originally, the team had planned their project agenda to understand what knowledge exists within indigenous communities in this regard, and hence what knowledge might be openly shared, to promote shared learning around adaptation to climate change. However, after becoming increasingly cognisant of the historical and present-day cognitive exploitation that tends to occur during research with indigenous communities, the team changed their focus to be more reflective of the community's needs. Thus, they set-out to develop an innovative research contract, developed in close consultation with community members and legal professionals, that could be used as a tool for negotiating community rights in all future knowledge collaborations [30].

Similarly, in attempting to address the challenges of limited resources for climate change adaptation and disaster response, a Caribbean-based OCSDNet project developed a 'knowledge broker artifact,' to create and mainstream a common vocabulary across Small-Island Development States (SIDS) for improved collaboration during disaster-management responses. Using a “design science” approach, the team engaged with diverse stakeholders to negotiate the creation of an “artifact,” that could be used to efficiently plan and streamline a coordinated response. Similar to previous case studies, the team suggested that beyond the intricate debates associated with the development of shared terminologies, a more important challenge was in regards to negotiating the diverse institutional and social arrangements between collaborating stakeholders.

Both of these examples demonstrate that ‘open science’ can be imagined as a flexible philosophy or mindset, rather than a fixed set of practices. Imagining the concept in this way allows for increased flexibility in solving complex development challenges and issues, without relying on a one-size-fits-all protocol. However, at the same time, this process of negotiation can be deeply complex and time-consuming, particularly when working across heterogenous communities, with different socio-cultural or institutional arrangements.

Beyond tools, four other projects use a case-study approach to examine the potential of applying an open science research framework to various, complex development challenges. In Kenya, the team sought to understand how open science may be harmonised with commercialisation practices, which tend to prioritise IP protection and personal property, while the Brazilian team applied an open science lens towards a complex social situation in Ubatuba, seeking to examine whether OS can be applied to facilitate and achieve sustainable development outcomes across a broad range of actors and activities. The Kenyan example reveals the deep complexities of sustaining and scaling-up open science initiatives within academic and policy environments that have on-going relationships with the private sector, who tend to value the protection of data and collaboration that offers value-for-money. Particularly in many Southern contexts, financially constrained research institutions face enormous pressure to procure research funding, often through systems of IP protection, including copyright and patenting. On the other hand though, the team found that most Kenyan institutions make use of both open access tools (such as repositories) where possible, as well as pursuing partnerships with the private sector. In this way, an institutional environment must be flexible to both ‘open’ and ‘closed’ systems of knowledge production, but remains largely driven by external funding agendas and possibilities.

In the Ubatuba case study in Brazil, the team raises the fundamental question of ‘development for *whom*?’ in determining to what extent open science can be used as a tool for achieving sustainable development outcomes. Using participatory methodologies, the team looked at environmental conservation issues in Ubatuba Brazil, through engagement with stakeholders from diverse sectors. The authors suggest that while open and collaborative science does create new spaces and methods for traditionally marginalised groups to engage in scientific discussions and local problem-solving, the complexity of some development problems demands the strategic involvement of larger institutions.

Similarly, acknowledging the historic bias whereby the production and legitimisation of scientific knowledge has been dominated by the North, an OCSDNet team in Argentina draws on four diverse case studies throughout Latin America, in order to look critically at the roles and outcomes of collaborative knowledge creation through an open science lens. The project concludes by suggesting that different "types" of development challenges may be more or less amenable to collaborative practices of open science.

Thus, within a southern context, the Brazilian, Kenyan and Argentinian case studies demonstrate the immense importance of building *partnerships* across diverse sectors and with different actors in order to maximise the potential of open science in development. While this involves complex negotiations and the establishment of trust and defined roles, it is necessary not only for understanding and addressing the complexity of some situations, but also from the perspective of project sustainability and resource sharing.

Finally, using a somewhat different approach, another OCSDNet research team sought to define and promote open science and open access in French-speaking West Africa and Haiti using a network-building and advocacy approach, through the assistance of social media tools, surveys and workshops. Acknowledging the lack of access to academic journals experienced by many institutions within the regions, the team engaged university students and staff in discussions about access to research and the proportional lack of representation of Southern (and particularly French-speaking African and Haitian) researchers in the production of scientific knowledge. This group also helped to promote the concept of 'cognitive justice' within the network - a concept which acknowledges the right of human beings to participate in the creation of knowledge that is relevant to their own lives, experiences and worldviews.

As a whole, these six projects represent the way in which open science, as a concept, can be adapted and applied to promote collaboration, knowledge sharing and innovation to tackle a wide range of development questions and issues. Particularly in Southern contexts, where independent institutions may lack access to funding and resources, these cases highlight the power and complexity of multi-actor collaborations in order to take advantage of diverse skillsets, limited resources and to find innovative solutions to complex development challenges.

4. Cross-Cutting Lessons & Conclusions

For the purposes of this paper, OCSDNet projects have been divided into three categories, with the intention of viewing open science 1) from the local, 'grassroots' level using an insider approach; 2) from a meta-level 'outsider' perspective to understand the challenges of scaling and sustaining larger open science projects; and 3) by practically and theoretically exploring the potential of open science principles through the creation of new tools and frameworks for addressing local development issues.

Despite the diversity of these projects, an overlapping set of themes and conditions emerged across all or many of the projects, which demonstrate some important aspects to consider when implementing an open science agenda that is inclusive, and which aims to meet development goals. First and foremost is the importance of *building a common language* amongst open science practitioners. As we have seen with the disaster management artifact in the Caribbean, the harmonisation of OS and commercialisation and OS in Kenya, the virtual herbarium in Brazil, and the community conservation project in Colombia and Costa Rica, the engagement of diverse stakeholders in processes of collaboration requires a deliberate and reflective process around shared principles and goals, to ensure that everyone is striving towards a common objective. Within OCSDNet, we have found that the creation of our OCSDNet manifesto was indispensable, in this regard, amongst our own membership.

Secondly, *a contextual or situational framing of open science is key f*or encouraging local buy in and ownership of a project. As we have seen through the diversity of projects within the network, there is no one-size-fits-all approach to open science, but it is instead a flexible concept that should be adapted to reflect local norms and realities. In this way, a contextual approach to open science is one that encourages the inclusion of diverse actors and ways of knowing, and hence cognitive justice.

A third theme to arise throughout our analysis is the need to *be critical of the processes and information to be shared* within the design and negotiation of open science architectures. Complete openness is not always feasible nor desirable in all situations for historical or socio-political reasons, or merely due to differing work priorities of diverse collaborators. Evidence of this was clearly demonstrated in the South African case, in which the team worked to safeguard the traditional knowledge of indigenous communities, as well as more practically within the Brazilian virtual herbarium project, which recognised that data providers should have a say in deciding what data is made openly accessible to the public. When research or data contributors have no say in whether their data is made open or not, the result could be a disempowering, rather than empowering one.

Another cross-cutting lesson to emerge from all projects is the importance of not 'remaking-the-wheel,' in the creation of discourse and practices around open science. Certainly in regards to smaller-scale citizen science initiatives, there is a considerable amount of development literature that has been written since the 1970's, with lessons and best practices for facilitating inclusive and participatory processes of community engagement, which position problem solving and social transformation as key objectives[7]. Gender and critical race theorists [31] [32] [33] have likewise produced invaluable work that must be taken into consideration for the development of a situated and inclusive open science. Thus, new advocates and practitioners of open science must not work in silos, but explore relevant work that has been done to provide the groundwork for an emerging way of conceptualising and practicing a more inclusive and collaborative science.

In a similar vein, there is a need for *increased interdisciplinary and cross-sector collaborative research, particularly between Southern actors.* As described throughout the paper, collaboration, in varying forms, is essential when combining open science and social needs. Whether this is at the local level, between teachers and students (as seen in Kyrgyzstan), between communities, government and the private sector (as seen in Ubatuba, Brazil), or between different students from different institutions and regions (as seen in the West African project), collaboration allows for the sharing of skills, ideas and resources for tackling complex development issues over the longer term, as well as generating the necessary momentum and ownership to work against institutional norms which could potentially limit the potential of 'open' partnerships.

In sum, OCSDNet teams have recognised that open science has the potential to transform the foundational structures of knowledge creation in new and important ways. In particular, open science has the potential to offer spaces, tools, opportunities and principles that facilitate opportunities for historically marginalised groups to participate in knowledge production, and to validate new and existing forms of local knowledge. At the same time though, there is a subsequent realisation that powerful actors continue to resist OS narratives that situate knowledge as a public good to which everyone should be able to access and contribute towards. However, as Sillitoe [13] explains:

"The idea is not that the small local knowledge stone should knock Goliath

[7] For example, see [35]

science over...It is that we should create space for others' ideas. This is necessary not only because it should continue to add to global science's awesome fund of knowledge, but also because it might help us to manage this knowledge more effectively for the planet and humankind."

To position this conclusion more broadly, Goal #9 of the Sustainable Development Goals recognises the need to "build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation" [34]. In this regard, open science has a key role to play towards ensuring that infrastructure and innovation are locally appropriate, inclusive and hence sustainable in the longer term. This calls for local participation and inclusive dialogue at all levels, including resources and policies from the 'top,' which must be grounded in and designed by knowledge from local communities. It is only through the inclusion and consideration of diverse human actors and experiences that open science might offer the opportunity for transformational human development.

References

[1] Bartling, S. & Friesike S. (eds). (2014) Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing. Berlin: SpringerOpen. http://www.openingscience.org/get-the-book/

[2] Nielsen, M. (2013). *Reinventing Discovery: The New Era of Networked Science* (Reprint edition). Princeton University Press.

[3] Fecher B., and Friesike S., 2014. Open science: One term, five schools of thought,. In: Sönke Bartling and Sascha Friesike (editors). Opening science: The evolving guide on how the Internet is changing research, collaboration and scholarly publishing. Cham, Switzerland: Springer International Publishing, pp. 17–47. doi: http://dx.doi.org/10.1007/978-3-319-00026-8_2

[4] OECD. (n.d.). Open science. Retrieved Dec. 7, 2016, from https://www.oecd.org/sti/outlook/e-outlook/stipolicyprofiles/interactionsforinnovation/openscience.htm

[5] Orth, A., Pontika, N., & Ball, D. (2016, June). FOSTER's Open Science Training Tools and Best Practices. In Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing (p. 135). IOS Press. http://ebooks.iospress.nl/publication/42908

[6] Schmidt, B., Orth, A., Franck, G., Kuchma, I., Knoth, P., & Carvalho, J. (2016). Stepping up open science training for European research. Publications, 4(2), 16. http://www.mdpi.com/2304-6775/4/2/16/htm

[7] Chan L, Okune A. & Sambuli N., 2015. "What is open and collaborative science and what roles could it play in development?" In: Sarita Albagli, Maria Lucia Maciel, and Alexanfre Hannud Abdo (editors). Open science, open issues. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), pp. 87–112, and at http://livroaberto.ibict.br/bitstream/1/1061/1/Open%20Science%20open%20issues_Digital.pdf

[8] Czerniewicz, L. (2015). Confronting inequitable power dynamics of global knowledge production and exchange : feature - opinion. *Water Wheel*, *14*(5), 26–28. Retrieved from http://journals.co.za/content/waterb/14/5/EJC176212

[9] Graham, F., De Sabbata S., & Zook M.A.. 2015. "Towards a study of information geographies: (Im)mutable augmentations and a mapping of the geographies of information," Geo: Geography and Environment, volume 2, number 1, pp. 88–105. doi: http://dx.doi.org/10.1002/geo2.8

[10] Moletsane, R. (2015). Whose Knowledge is It? Towards Reordering Knowledge Production and Dissemination in the Global South. Educational Research for Social Change (ERSC) Volume: 4 No.2, pp.35-47. http://ersc.nmmu.ac.za/articles/Vol_4_No_2_Moletsane_pp_35-48_October_2015.pdf

[11] Okune, A., Hillyer, B., Albornoz, D., Sambuli, N., & Chan, L. (2016). Tackling Inequities in Global Scientific Power Structures. The African Technopolitan Vol. 4, Issue 1. Retrieved from https://tspace.library.utoronto.ca/handle/1807/71107

[12] OCSDNet. (2016). *The Open and Collaborative Science in Development Network: About the Network.* http://ocsdnet.org/about-ocsdnet/
[13] Sillitoe, P. (2007). Local science v. global science: An overview. In Sillitoe, P. (ed). *Local Science vs. Global Science: Approaches to indigenous knowledge in international development*. (1 - 22). Bergahn Books.
[14] Hall, B. L., & Tandon, R. (2017). Decolonization of knowledge, epistemicide, participatory research and higher education. *Research for All*, *1*(1), 6–19. https://doi.org/10.18546/RFA.01.1.02
[15] Ostrom E. Governing the Commons: The Evolution of Institutions for Collective Action. unknown edition. Cambridge ; New York: Cambridge University Press; 1991. 298 p.
[16] Ostrom E. Understanding Institutional Diversity. Princeton: Princeton University Press; 2005. 376 p.
[17] Hess, C. & Ostrom, E. (2005). A Framework for Analyzing the Knowledge Commons : A chapter from Understanding Knowledge as a Commons: from Theory to Practice. Libraries' and Librarians' Publications. Paper 21. http://surface.syr.edu/sul/21
[18] Hess, C. (2012). The Unfolding of the Knowledge Commons. St Antony's International Review, 8(1), 13–24.
[19] Frischmann, B. M., Madison, M. J., & Strandburg, K. J. (Eds.). (2014). *Governing knowledge commons*. Oxford University Press.
[20] Appadurai, A. (2006). The right to research. *Globalisation, Societies and Education*. 4(2): 167 - 177
[21] Albornoz, D., Hillyer, B., Posada, A., Chan, L. & Okune, A. (2017 In Press). Co-constructing an Open and Collaborative Manifesto to Reclaim the Open Science Narrative. Proceedings of the 21st ELPUB Conference: Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices. IOS Press.
[22] Sen A. (1989). Development as Capability Expansion. *Journal of Development Planning* 19: 41–58.
[23] Dolamo, R. (2013). Botho/Ubuntu: The heart of African ethics. *Scriptura.* 112: 1 - 10. Retrieved from: https://journals.co.za/content/script/112/1/EJC148117
[24] Escobar, A. (2002). Reflections on 'Development:' Grassroots approaches and alternative politics in the Third World. *Futures*. 24 (5): 411-436 https://doi.org/10.1016/0016-3287(92)90014-7
[25] Lewis, D. (2005). Anthropology and development: the uneasy relationship. London: LSE Research Online. Available at: http://eprints.lse.ac.uk/archive/00000253
[26] Monni, S. & Pallottino, M. (2015). A new agenda for international development cooperation: Lessons learnt from the Buen Vivir experience. *Development*. 58(1): 49 - 57
[27] Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution.* 24(9): 467 - 471
[28] Cohn, J. (2008). Citizen Science: Can volunteers do real research? *BioScience.* 58(3):192-197 doi: http://dx.doi.org/10.1641/B580303
[29] Canhos, D. A. L., Sousa-Baena, M. S., Souza, S. de, Maia, L. C., Stehmann, J. R., Canhos, V. P., … Peterson, A. T. (2015). The Importance of Biodiversity E-infrastructures for Megadiverse Countries. *PLOS Biology*, *13*(7), e1002204. https://doi.org/10.1371/journal.pbio.1002204
[30] Traynor, C. (2017) Contracting Justice Workshop – Exploring socially just research processes. (2017, March 30). https://ocsdnet.org/contracting-justice-workshop-exploring-socially-just-research-processes/
[31] Harding, S. (2006). *Science and Social Inequality: Feminist and Postcolonial Issues*. Urbana: University of Illinois Press.
[32] Harding, S. (2015). *Objectivity and Diversity: Another Logic of Scientific Research*. Chicago: University Of Chicago Press.
[33] Haraway, D. (2008). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*. 14 (3): 575 - 599
[34] United Nations. (2016). Sustainable Development Goal 9 https://sustainabledevelopment.un.org/sdg9. Accessed May 1, 2017
[35] Chambers, R. (1994). The origins and practice of participatory rural appraisal. World Development. 22(7): 953 - 969

Annex 1. OCSDNet Project Names & Keywords

Short Title	**Key Words**
Practicing Open Science at the 'Grassroots'	
Water Quality and Social Transformation in rural Kyrgyzstan	Kyrgyzstan, rural communities, citizen science, environmental conservation, water quality, participatory action research, open science motivation, teachers and students
Water quality and community development in Lebanon	Citizen science, participatory research, community-based environmental management, water quality, empowering conservation, bottom-up policy making
Community-driven environmental conservation in Costa Rica and Colombia	participatory action research, citizen science, Model Forests, Costa Rica, Colombia, human capabilities, adaptive capacity, sustainable development, biodiversity
Open Science Hardware for Development in Southeast Asia	open science hardware (OSH), transnational networks, little science, citizen science, do it yourself (DIY), Indonesia, Thailand, Nepal, tools, participation, tinkering, Right to Science
Analysing Existing Open Science Projects	
Evaluating Open science e-infrastructure in Brazil	Brazil, virtual herbarium, botany, interdisciplinary collaboration, e-database, open science infrastructure
Negotiating Open Science in Argentina	open science, Argentina, negotiating openness, opening process, boundary objects
Exploring the potential of Open and Collaborative Science through new Tools and Frameworks	
Researcher contracts for Indigenous knowledge in South Africa	South Africa, indigenous knowledge, climate change, intellectual property rights, research contract, decolonising research methodologies, terra nullius
Commercialisation & Open Science in Kenya	Kenya, IP laws, open science, universities, private sector, collaboration, research partnerships, commercialisation
Disaster Management Tools for Small Island States	Disaster recovery plans, Small Island Developing States, Design Science, regional collaboration, knowledge broker artifact
Sustainable development and the potential for OCS in Brazil	Ubatuba, social change, sustainable development, potential of open science, participatory action research, diverse actors
Social problems and the potential of OS in Latin America	Latin America, openness, non-hegemonic countries, social problems, collaborative science, cognitive exploitation
Building Open Science Social Networks in West Africa & Haiti	West Africa, Haiti, open science networks, science shops, open repository, open research, participatory research, cognitive justice

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-34

Claims About Benefits of Open Access to Society (Beyond Academia)

ElHassan ELSABRY[a,1]
[a] *National Graduate Institute for Policy Studies, Japan*

Abstract. This study tries to systematically identify claims about societal benefits of Open Access by analyzing different documents written by Open Access supporters. Three types of documents are used: key declarations and statements in support of Open Access, Open Access policies issued by public funding agencies and journal editorials announcing the adoption of Open Access. Analysis shows these three types emphasize different benefits for Open Access as they address different audience. There is strong support of the idea that Open Access has benefits to different groups of people outside side the university/credentialed research institutes. It is not clear how much evidence is available to support these claims, but identifying them would suggest new stakeholders to involve in the conversation and perhaps also inform the ongoing debate about who should bear the cost of Open Access..

Keywords. Open access mandate, open access policy, societal impact of research, funding agency, advocacy

1. Introduction

This paper presents the results of a study attempting to identify the different benefits of Open Access as anticipated by its supporters. Focus is mainly on claims about societal benefits of Open Access, i.e. those beyond the research community (outside the university and credentialed research institutes). As mentioned above, recent reviews of literature about Open Access [1] [2] have emphasized the lack of enough research to investigate the potential of Open Access to benefit individuals or groups of people who do not belong to universities and credentialed research institutes. This is in contrast to the abundance of studies about other aspects of Open Access (e.g. citation advantage). Davis & Walters [2] noted that “almost no studies have evaluated whether free access to the scientific literature has had an impact on the use of scientific information in non-research contexts such as teaching, medical practice, industry, and government”. The reason for this has been speculated (by a recent Research Information Network (RIN) report [3]) to be that it is currently not possible to “gather systematic data on the demographics of users either on publisher platforms or via repositories”.

This study comes in the context of a larger project [4] aiming to identify the societal benefits of Open Access and to devise new ways to measure and document this impact. Identifying the “claimed” societal benefits of Open Access (which is the aim of

[1] Corresponding Author, ElHassan ElSabry, Science, Technology & Innovation Policy Program, National Graduate Institute for Policy Studies, 7-22-1 Roppongi, Minato-ku, 106- 8677 Tokyo, Japan; E-mail: doc14153@grips.ac.jp

this study) will support the larger project in two ways. First, these claims can be tested against the currently available evidence about OA benefits to assess their credibility, which would be a useful exercise to guide the Open Access movement. Second, a deeper understanding of the discussion on societal benefits of OA will give insights about which stakeholders to include in the conversation and perhaps also inform the current debate on who should bear the cost of Open Access.

The paper is organized as follows. An overview of the study design is presented at the beginning. Then, owing to the different types of documents analyzed in the study, the data used for each group of OA supporters is presented in a separate section along with a brief background on the issue and the analysis results. A discussion of the overall outcomes and their implications then follows and the paper ends with some suggestions for future research based on the study findings.

2. Study Design

A total of 164 OA-related documents were chosen to represent the views of a wide range of OA supporters. Focus was not on the entire content of these documents but only the one or more key statement(s) within them, where the purpose behind supporting Open Access was stated. While many of these "statements of purpose" were mentioned in a straightforward manner under a separate section of the document (e.g. "Why Open Access", "Advantages of Open Access"), some were spread all around the document and were inferred from the context. Three types of documents were analyzed for the purpose of this study. First, a selection of prominent statements and declarations about Open Access was used to represent the views of Open Access advocates worldwide. Second, policies in support of Open Access issued by government bodies were used to understand how policymakers perceive OA's potential benefits. Third, it was important to include the perspective of researchers. Hence, editorials announcing the launch of open access journals (or conversion of traditional ones to OA) were examined as a possible source for journal editors' beliefs about Open Access. While it can be argued that there is some overlap between these three groups of OA supporters, it is also important to note that the chosen documents were written for different purposes and address different audiences.

3. Claims about OA Benefits by Group

3.1. Open Access Advocates

It is very difficult to define "OA Advocates" as a coherent group of people. It is a group that includes researchers, librarians, university administrators, research funders (both public and private), some scholarly publishers and even university students. Nonetheless, since what characterizes all of them is their outspoken support for Open Access, statements and declarations they produce can be a good representation of how they see Open Access and the benefits they expect from it. Declarations and statements in support of Open Access have played different roles at different points in the history of the Open Access movement. They were written to define the movement and lay out its main goals, to respond to related developments on the scholarly publishing scene or even to impose certain agendas on the debate. Many of them were used as tools to gather support for OA and were usually accompanied by large scale campaigns to call on people to sign them. While there is a wide range of documents fit the "statements in support if OA" description (e.g. one can consider every OA mandate or policy as such), a representative list of key documents had to be chosen for this study. The Open Access Directory (OAD) was consulted for this purpose. OAD is a community-sourced database aiming to document the Open Access movement. It is administered by a group of prominent OA advocates and hosted by the Simmons College.

Eight declarations were selected from the OAD list of “Declarations in support of OA” [5]. They were selected owing to their significant influence on the Open Access movement, global nature and their representation of different stakeholders of the scholarly communication system (librarians, publishers, researchers, funders, students and prominent advocates). Table 1 lists the chosen statements, their respective years of adoption and keywords pointing to benefits of Open Access as believed by the statement authors/signatories.

Table 1. Key declarations supporting OA and benefits of OA according to them

Statement	Year	Beneficiaries of OA
Tempe Principles for Emerging Systems of Scholarly Publishing	2000	researchers, industry, professors, students, informed citizens, the public
Budapest Open Access Initiative	2002	researchers, teachers, students, other curious minds
Bethesda Statement on Open Access	2003	researchers, developing countries, the public
Access to Scientific Information (by the Inter-Academy Panel)	2003	researchers, developing countries
Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities	2003	researchers, society
IFLA Statement on Open Access to Scholarly Literature and Research Documentation	2004	researchers, disadvantaged researchers
The Student Statement on The Right to Research (R2RC)	2009	researchers, students, patients, informed citizens, developing countries
Washington D.C. Principles for Free Access to Science	2004	not researchers , not patients

As expected, all declarations assume that researchers are the main beneficiaries from Open Access. These benefits take two forms, either through direct gain (i.e. visibility and citations for one’s own work) or as a general enhancement of the quality of research through the transparency and democratization offered by Open Access. Four declarations have made reference to the subgroup of “disadvantaged researchers”. Those are researchers whose institutions could not (or have never been able to) cope up with the rising costs of access to journals, especially in developing countries. The Budapest Declaration later specifically argued that Open Access should not be understood as a one-way communication tool (i.e. from the knowledge-rich north to the knowledge-poor south) but as providing mutual benefits for both. The declaration argues that removing barriers to scholarly literature will “share the learning of the rich with the poor and the poor with the rich”.

References to the benefits of Open Access to the educational process was made in three of the declarations. Needless to say, students’ frustration with the lack of proper access to research was the main driver of the R2R Statement. One of the signatories (The European Federation of Psychology Students' Associations) has even tried to systematically study this a few years later [6].

Beyond the academic/research community, three declarations made reference to “society” and “the public”. It is not clear if that was intended to mean specific benefits to laypersons from access to scholarly literature, or indirect benefits from an enhanced body of knowledge. One of these declarations, the Tempe Principles, also mentioned “an informed citizenry and a healthy global economy”. However, these were mentioned as the outcomes of “creation, dissemination, and application of new knowledge”. In other words, they are not direct beneficiaries from access to literature. One could also argue that Budapest Declaration’s reference to “other curious minds” is intended to encompass any groups of potential users outside academia. Nonetheless, none of the sources later cited in the declaration (as a proof of efficiency of OA) provide information about usage outside academia.

The R2RC statement was the only one to make explicit reference to some groups of society that could benefit from access to research. They mentioned patients who would “have access to the latest medical research” and citizens who could “evaluate scientific information on environmental impacts”. The statement does not offer any supportive evidence in this regard. However, as the most recent declaration among the ones in this study, it is possible that some results of research about of OA’s impact of society was already available to those who drafted it. Such a possibility is strengthen by the fact that four of the six Open Access policies cited by the R2R statement make similar claims about groups of societies that could benefit from access to scholarly literature. These include clinicians, policymakers (CIHR policy), families, patients (Autism Speaks policy), media (Canadian Cancer Society policy) and educators (Stanford GSE motion).

While many would consider the “Washington D.C. Principles for Free Access to Science” not a statement “in support” of OA as defined by this study, it was nonetheless important to include it here. This is mainly because this particular declaration argued against the societal benefits of Open Access. In addition to denying the need for access (even among researchers) by claiming that “published literature is routinely and readily available to all who need and want it”, the declaration asserted that “[it] is debatable whether members of the general public can actually benefit from reading the original research literature, as its arcane and specialized reporting is intended primarily for other researchers,”. This was also extended to imply that OA’s benefits to clinicians is also debatable given that “many findings are not relevant for immediate clinical application”. A few years later, some still maintained that there is no evidence for “unmet demand for the primary medical or health sciences literature among the general public”, albeit this “does not necessarily reflect the absence of unmet demand” [2].

Examining these key declarations shows that the issue of societal impact of Open Access was not strong on the agenda of most OA advocates who drafted them. Their main contention appeared to have been that it was not wise to ignore the value internet can add to scholarly communication. Open Access was the most efficient way to ensure the freedom of knowledge and internet’s contribution to enhancing research quality and reach.

3.2. Policymakers

There have been several case studies published to examine the impact of Open Access policies and mandates on individual institutions. They examined things like policy efficiency [7] or the researchers' reaction to these policies [8]. Only a small amount of studies aimed at studying government OA policies though. PASTEUR4OA (Open Access Policy Alignment Strategies for European Union Research) is probably the most extensive research project concerned with OA policies. It is based on the same database of OA policies used in this study and analyzes different policies. The aim was to encourage EU member states to align their OA policies to ensure best practice and to make compliance easier for researchers funded by grants from multiple source. Addressing policymakers about the anticipated benefits of Open Access, PASTEUR4OA researchers referred in several documents (for example [9]), to OA as a strategy to cut publishing costs and a way to foster innovation by giving SMEs access to the scholarly literature. In fact, one of their policy briefs was dedicated to present a framework of how knowledge transfer (via Open Access) has "spillover" effects on many segments of society outside the research community [10]. Policy guidelines developed by UNESCO have also echoed similar arguments but also emphasized the impact of access to biomedical literature on patients and healthcare practitioners [11]. In this regard, Waltham noted that in the US this tendency (to encourage public access to research) is a result of pressure by patient advocacy groups, in the UK it stems from a more general mission to raise the public understanding of science [12].

The Registry of Open Access Repository Mandates and Policies (ROARMAP) is a well-known, comprehensive resource for OA policies from organizations all over the world. However, the majority of listed policies are issued by universities or research units requesting (or requiring) their faculty and research staff to make the outcomes of their research openly available on the internet. As of December 24, 2016, only 136 (of around 800 listed polices) were issued by organizations described as funders (82) or organizations the both perform and fund research (54). These were either government bodies (e.g. ministries, parliaments), national research councils, national academies, or other smaller units. Twenty-nine private research funders (e.g. Wellcome Trust) were excluded as well as four entries that were not government bodies but partnership programs or universities.

Of the remaining 103 public research-funding organizations listed in ROARMAP, some fell under policies of larger organizations (16), issued policies that were not about research papers (9, e.g. open data polices), or published other types of documents (4, e.g. not a policy but workshop recommendations). Three policies could also not be found. Therefore, analysis for this study was based on 72 policies that fit the initial criteria. The majority of policies had some English version available online. For those that did not (12 policies), online automatic translation was used to identify and translate the statement of purpose in the policy. A native speaker was consulted in cases where the automatic translation was not clear. Table 2 presents the overall results of analyzing the 72 valid policy documents.

Table 2. Beneficiaries of Open Access according to government policies

Benefits of OA	Sample Keywords	Frequency	Percentage
research	quality, impact, reproducibility, duplication of efforts, open science, globalization, pace	44	61
industry	economy, growth, (open) innovation, valorization	34	47
public	awareness, culture, public understanding of science, taxpayer right, public accountability, scrutiny	29	40
professionals	users, deployment of research, uptake, clinicians	10	14
government	policymakers, public sector	10	14
education	OER, educators, fast percolation to high education	9	13
credibility	evaluating program managers, government transparency, M&E, efficient use of funds	8	11
visibility	intellectual gap, global recognition	6	8
NGOs	charities, NPOs	4	6
no mention		17	24
TOTAL		72	100

The majority of policies (61%) make at least one claim about the positive impact of Open Access on the research community. This does not seem to be different whether the organization issuing the policy is only a funder or also conducts in-house research. This is understandable given that benefits to the research enterprise can safely be considered the main purpose behind all of these policies. As mentioned before, much evidence has been piling up over the years to support the belief in OA benefits to the research enterprise. Examples of this positive impact include enhancing the quality of the research, allowing for more reproducibility, avoiding duplication of efforts and supporting the globalization of science with more reading and citations.

An interesting finding from analyzing these government OA policies is their consistent emphasis on the benefits of OA to the economy. This is not about Open Access being economically more efficient by some "system‐wide cost savings", as it was shown by Houghton [13] for example. Rather, it is about OA making more knowledge available to firms to build on, creating innovative products and services that would consequently boost the economy. The argument is well summarized in the European Commission's position that "[fuller] and wider access to scientific publications and data … help to accelerate innovation" because "faster to market = faster growth" [14]. A similar sentiment can also be detected in the US government commitment to fund and make available research which "catalyzes innovative breakthroughs that drive [the American] economy" [15]. This is also consistent with Prosser's idea that the move to more knowledge-based economies is one of main drivers supporting the argument for Open Access among policymakers. He mentions that "[as] developed countries struggle with the transition to post-industrial economies, there is a growing belief that knowledge provides both power and economic growth" [16].

Policymakers concern about benefits to the taxpaying public is understandable. Of the 40% of policies that mentioned these benefits, some made broad claims like preserving knowledge and culture. Open Access would enhance the knowledge produced by researchers and allow for maintaining it, which consequently will make it more relevant and useful to society as a whole. What lacked evidence was the more specific claims made by other policies about Open Access making possible the public scrutiny of the research outcomes. It is not clear what mechanism this will happen through. Indeed, some policies mentioned that OA will allow for better evaluation of the funding programs and their managers, which will consequently result in more credibility for the organization. However, this kind of benefit was included separately under "credibility" because it is more about accountability to other (superior) bodies of government than to the public per se. The claim that Open Access will increase the public understanding of science is also one that lacks supporting evidence. Even participating in citizen science projects does not guarantee an increase laypeople's understanding of science [17], let alone the mere presence of scholarly literature online.

Three other benefits get nearly equal attention from policymakers. These are OA research usage by practitioners (e.g. doctors, lawyers, etc.), usage by public sector researchers (e.g. policy research units) and the Open Access as a form of Open Educational Resources (OER). Each one of these benefits is acknowledged in one sixth of the policies.

The relatively high percentage of policies (24%) that mentioned no specific purpose for supporting Open Access is mainly because some policies were not issued in a separate policy document (e.g. law, resolution) but as a changes to already existing documents that included more topics than just Open Access (e.g. national science law, guidelines for using research funds, etc.).

3.3. Leading Researchers

First, as mentioned above, when considering the "anticipated" benefits of Open Access, the opinion of researchers cannot be dismissed. This is especially true for researchers who have leading positions in their fields. Editorials are by definition a good venue where journal editors can express their thoughts of ideas about different issues related to their field. For this study, a set of 85 editorials were collected (where a new Open Access journal is announced or when an existing journal announces conversion to Open Access) to determine the views of leading researchers about the benefits on Open Access. These editorials come from journals across different fields of research (albeit with very strong presence of biomedicine).

The selection was based on searching the content of Scopus database as of January 21, 2016. Search was limited to publications of the type "editorial", which contain the expression "Open Access" either in the title or in the indexing (or author-provided) keywords. After excluding publications where "Open Access" was used to describe an unrelated concept (e.g. open-access endoscopy or open access railway infrastructure), a list of 517 editorials remained (including 15 duplicate entries). Titles of these editorials were then checked to classify the editorials into four groups:

- editorials announcing a new OA journal or a subscription journal's transition to OA (85)
- editorials announcing some new green or hybrid OA policy (60)
- editorials discussing Open Access without announcing OA-related decisions (257)
- editorials whose topic is unidentifiable based on title (100)

Only the first of these three groups was used in this study as a source of journal editors' views on Open Access benefits. The second group was excluded because only reading the editorial might not have been enough to know the real intentions of choosing Open

Access. It is not clear too if the purpose was anything more than compliance with funder requirements or the increased revenue associated with the hybrid model. The results for analyzing the first group of editorials are summarized in Table 3.

Table 3. Beneficiaries of Open Access according to OA journal editors

Benefits of OA	Sample Keywords	Frequency	Percentage
wider dissemination	exposure, impact, visibility, indexing, archiving, citations, author retains copyright,	62	73
efficiency	access for developing countries, disadvantaged researchers, freedom of knowledge	33	39
rapid publication	immediacy, competitiveness	26	31
professionals	practitioners, clinicians, stakeholders	18	21
rising trend	citing OA declarations, compliance with funder mandates, revolution of scholarly publishing	17	20
public	taxpayer right, public understanding of science, interested laypersons	15	18
government	evidence-based policymaking	7	8
other groups	amateurs, media, parents, teachers	3	4
industry	drive innovation, private sector R&D	4	5
education	university students, professors	2	2
no mention		10	12
TOTAL		85	100

The most significant result from analyzing the Open Access benefits as seen by journal editors is their consistent focus on benefits to the research community. Unlike the previous two types of documents, editorials refer very little to any "public" or societal benefit of Open Access (only 18%). Even groups of people who might not necessarily be part of the research community but are very close to it (e.g. practitioners 21% or students 2%) are mentioned relatively very little. Benefits to industry are also rarely mentioned, although most of the editorials come from the field of biomedicine, which is traditionally associated with the very "science-intensive" pharmaceutical industry.

Another interesting aspect is that only six editorials (7%) made reference to compliance with funder mandates. This suggests that (at least for gold OA journals) a move to Open Access in communicating research might have happened naturally even in absence of funder mandates, given that editors chose to emphasis other benefits of Open Access.

Otherwise, the great emphasis that the majority of editors put on benefits to researchers in their field as the primary reason for support Open Access is very plausible. This was especially true for new journals that tried to emphasize benefits like citations and exposure as a way to attract their initial submissions (sometimes in combination with other strategies like waived APCs). It is however important to consider that for some journals the move to Open Access was also the move to online publishing, which by itself can account for benefits like rapid dissemination or more global visibility (relative to print-only publishing).

Some editorials mentioned adopting Open Access would be “sponsored” by a parent organization, i.e. no APCs will be required. However, it remains a limitation of this study the inability to know if the perceived financial gain from APCs was the main reason behind choosing Open Access. This is especially important to consider in cases where the editorial mentioned Open Access as a way to support the "growth" of their journal.

4. Discussion

Comparing the position of each of the three groups of OA supporters signals two main differences. One difference is that ideas about what benefits Open Access has on the researching community seems to be more about the researchers themselves (e.g. citations, visibility, copyright ownership, etc.) as viewed by journal editors, while at the level of policymakers more "abstract" benefits are generally perceived (e.g. globalization of science, reproducibility, transparency, etc.). Declarations occupy a somewhat middle ground on this issue. Regarding the benefits to developing countries, policymakers are the least to refer to this point. However, five policies (coming from Ireland, France, Brazil, Belgium and Slovenia) make reference to the somewhat similar concept of bridging the intellectual gap by making their own research more visible.

The analysis has also shown that there is near consensus that benefits of Open Access go beyond the academic/research community. Still, there is a lot of variation among the three groups in how they perceive the extent and reach of these societal benefits. The little regard OA journal editors give to Open Access benefits beyond the research can be explained in two ways. It is possible that they do not believe those benefits exist. This is understandable given the very little research done on this issue. Supposedly, researchers are more inclined to make evidence-based claims than most activists and policymakers. The other possibility is that they believe those benefits exist but (in writing those editorials) chose to focus on benefits to researchers as a way to garner support for their decision to adopt Open Access. In both cases, more research is needed on this topic to inform researchers about any potential societal benefits for Open Access, which in turn might influence their decision to adopt it.

5. Agenda for the Future

Claims about the societal benefits of Open Access, as investigated in this study, necessitate more discussion into two vital issues.

First, how much evidence is available to support claims about the benefits of Open Access? Research is needed to identify, classify and compare any literature that investigated the impact of Open Access on society. As mentioned at the beginning, it would be also interesting to evaluate the claims of this study against any available related evidence. It might also prove necessary to add to this evidence base with more research projects targeting groups that might benefit from Open Access, but do not necessarily belong to universities or credentialed research institute. Examples of these groups can be extracted from claims in mentioned in this study (e.g. Clinicians, charities, industry researchers) and also from outside of it (e.g. citizen scientists, think tanks, people in legal practice, etc.)

Second, in the light of the emphasis on the benefits of Open Access to (research-intensive) industries, to what extend is it plausible to suggest that they also contribute to the cost of Open Access provision? Perhaps coordination of efforts towards Open Access can be much easier between entities that are somehow part of the research community (e.g. university libraries or research funders). However, corporate subscriptions represent 15-17% of the journal publishing revenue and some leading publishers already view this segment as an expanding market [18]. It might not prove difficult to involve large corporations (or even other entities like think tanks and resourceful government research units) in cooperative models of funding for Open Access [19]. Smaller firms can also contribute through representative unions (e.g. Biotechnology Industry Organization), which was previously suggested as way for them to manage subscriptions [20].

In conclusion, the issue of societal benefits of Open Access can prove to be very complex and manifold. However, approaching it in the right way can take the debate on access to research to a whole new level, by reframing it as a social issue, rather than one that is just relevant to researchers.

References

[1] S. Pinfield, Making open access work: The 'state-of-the-art' in providing open access to scholarly literature, *Online Information Review* **39**:5 (2015), 604-636.

[2] P. M. Davis, W.H. Walters, The impact of free access to the scientific literature: a review of recent literature, *Journal of the Medical Library Association* **99**:3 (2011).

[3] Research Information Network. *Monitoring Progress in the Transition to Open Access*, 2014.

[4] E. ElSabry, Mapping Open Access Societal Impact. *Research Ideas and Outcomes* **3**:e11743 (2017).

[5] Open Access Directory. Declarations in support of OA. [Online].; 2016 [cited 2017 3 10. Available from: http://oad.simmons.edu/oadwiki/Declarations_in_support_of_OA.

[6] I. Flis, J. Haslbeckband, C. Noone, *European Student Scientific Literature Access Study*. European Federation of Psychology Students' Associations, 2013.

[7] P. Vincent-Lamarre, J. Boivin, Y. Gargouri, V. Larivière, S. Harnad, Estimating open access mandate effectiveness: The MELIBEA score. *Journal of the Association for Information Science and Technology* **67:**11 (2016), 2815-2828.

[8] S. Teplitzky, M. Phillips, Evaluating the Impact of Open Access at Berkeley: Results from the 2015 Survey of Berkeley Research Impact Initiative (BRII) Funding Recipients. *College & Research Libraries*, **77:6** (2016), 568-581.

[9] V. Tsoukala, M. Angelaki, Open Access Policy Guidelines and Template for Funders. PASTEUR4OA Policy Guidelines, 2015.

[10] M. Picarra, Open Access to scientific information: facilitating knowledge transfer and technological innovation from the academic to the private sector. PASTEUR4OA Briefing Paper, 2015.

[11] A. Swan, *Policy Guidelines for the Development and promotion of Open Access. UNESCO*, Paris, 2012.

[12] M. Waltham, Open access - the impact of legislative developments. *Learned Publishing* **18** (2005), 101-114.

[13] J. Houghton, Costs and Benefits of Alternative Scholarly Publishing Models: Lessons and Developments. In Publishing in the networked world: transforming the nature of communication: *14th International Conference on Electronic Publishing* (2010).

[14] EU Framework Program for Research and Innovation. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 (Version 1.0), *European Commission,'* 2013.

[15] J.P. Holdren, Increasing Access to the Results of Federally Funded Scientific Research. ; 2013.

[16] D. Prosser, Public Policy and the Politics of Open Access. *LIBER Quarterly* **17:**2 (2007).

[17] Scientific knowledge and attitude change: The impact of a citizen science project. *International Journal of Science Education*. 2005; **27**(9): p. 1099–1121.

[18] M. Ware, M. Mabe, The STM Report (Fourth Edition), *International Association of STM Publishers,* The Hague, 2015.

[19] A.C. Jiménez, J. Willinsky, D. Boyer, G.D. Col, A. Golub, Why an open access publishing cooperative can work - A proposal for the AAA's journal portfolio. *Journal of Ethnographic Theory* **5:**2 (2015).

[20] Lyman S. Industry access to the literature (letter to the editor). *Nature Biotechnology* **29**:7 (2011), 571-572.

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-44

Open Access Policy and Funding in Cyprus University of Technology a Case Study

Marios ZERVAS[a,1], Petros ARTEMI[a] and Stamatios GIANNOULAKIS[a]
[a] *Library and Information Services, Cyprus University of Technology*

Abstract. Today an increasing number of researchers and scientists follow Open Access. Open Access is a movement that offers researchers and the academic community the opportunity to share and access academic information freely and immediately. The Library and Information Services at the Cyprus University of Technology (CUT) has defined Open Access as one of its main strategies. An author fund has been established since 2013, with the financial support of a pharmaceutical company, in order to promote and support Open Access. Statistics of this Fund have shown that funding covers publications mainly in health sciences. These findings have allowed us to implement a new policy for the financing of Open Access publications from our university's budget, which falls within the framework of its social responsibility. However, considering the "hybrid model" and the "double dipping" which favors 'big' publishers, we will examine the different possibilities and present our reflections and decision for the new policy. Finally we will describe the policy implemented.

Keywords. Open Access, author fund, author fund policies, hybrid model

1. Introduction

Open Access is the free, immediate, and online availability of research articles coupled with the rights to use these articles fully in the digital environment [1]. According to Peter Suber [2], Open Access removes price barriers such as subscription fees and permission barriers such as licensing restrictions, in order for authors to publish their work without expectation of payment and with minimal use restrictions.

There are about 28.000 peer reviewed journals which publish about 2.5 million articles per year [3]. As we understand, there is a huge number of peer-reviewed journals to which a library or even a consortium is unable to subscribe.

Today an increasing number of researchers and scientists follow Open Access (OA). Libraries, especially academic libraries, and promote Open Access by providing their resources to their users and implementing institutional repositories.

Furthermore, the Cyprus University of Technology (CUT) Library has defined Open Access as one of its main strategic goals and is the first academic research institution in Cyprus, which has implemented a Research Fund to promote Open Access and support scientific publications in Open Access journals.

The two basic Open Access models are referred to as "Gold road" and "Green road". In the first model, "Gold road", the authors make their work openly accessible through Open Access journals. The Open Access journals often charge processing fees

[1] Corresponding author, Arch. Kiprianou 31, 3036, Limassol, Cyprus; Email: marios.zervas@cut.ac.cy.

which need to be covered by the authors. In the second model the "Green road", the authors submit their work to an institutional repository [4].

The CUT Library after studying different policies, applied by SPARC members, decided to apply a restricted policy. Articles, to be eligible for funding, must be published without any access restrictions in a peer-reviewed electronic Open Access journal which is a member of the Directory of Open Access Journals.

The Senate of CUT approved the creation of the CUT Open Access Author Fund. Within the framework of its Corporate Social Responsibility, a local pharmaceutical company has undertaken the financing of the Fund, with a maximum funding amount.

Research carried out under the European program PASTEUR [5] showed that many institutions enhance the Open Access fund from their own fund or allowed APCs to be paid from a research grant.

Recently, the Cyprus government approved of the Cyprus National Open Access Policy which complies with the current provision and best practices of the European funding program Horizon 2020 and the policy for Open Access to scientific publications [6].

The Cyprus Research Council has not yet managed to provide coverage for OA fees. Therefore, the CUT Library Committee suggested to the CUT Research Committee to enhance the Open Access APCs which was approved by the senate at the beginning of 2017. Furthermore, the Library committee is going to decide on the policy for the new fund.

In this paper, we are going to analyze the policies for financing Open Access publications. In addition, we will investigate the results of the Open Access Author Fund initiative and provide numbers and figures about the applications for funding and publication production. Furthermore, we will investigate the Open Access Author Fund policies worldwide and we will describe the policy we will implement in the case of funding Open Access by the University budget. Also, we will discuss recent CUT enhancements and the implementation of ORCID in order to encourage our researchers at the beginning of their careers to establish a persistent researcher identifier. The upgrade of our Institutional Repository with CRIS functionalities will be discussed in order to give more value and become more attractive, because it is our belief that when users use institutional repositories, they will better understand the benefits of Open Access.

2. The Journey of the Library towards Open Access

As already mentioned, the CUT Library has set Open Access as one of its Strategic goals. In order to achieve the aforementioned, the Library and Information Services of CUT implemented a strategic plan and followed specific steps.

In 2007, the Cyprus University of Technology decided that the academic staff must deposit their scientific publications to the Library. Then, in 2008 the deposit of the undergraduate and postgraduate theses to the Library was made mandatory.

Moreover, the Library developed the institutional repository Ktisis with the use of the open source software DSpace. Also, since its operation, Ktisis provides the author who deposit its work the possibility to apply one of the available Creative Commons licenses.

The CUT Library supports the two basic Open Access models. The first one is the "Gold road", where the authors make their work openly accessible through Open

Access journals. The Open Access journals often charge processing fees which need to be covered by the authors. In this case, the CUT Library covers the APC fees according to the CUT Open Access Author Fund requirements.

The second model is the “Green road”, where the authors submit their work to an institutional repository. The CUT Library has designed and developed the first institutional repository in Cyprus named Ktisis. The deposit of the undergraduate and postgraduate thesis in Ktisis was made mandatory. Also, the CUT Library developed management policies which were approved by the academic senate for the undergraduate and postgraduate theses, and the scholarly publications.

Furthermore in 2011, the Cyprus University of Technology signed the Berlin Declaration on Open Access.

Moreover, the same year the CUT Library conducted a study to find out what the academic community and PhD students thought about Open Access, difficulties they faced and also to examine the degree of familiarity with the Open Access publishing model [8]. From the research, we concluded that the use of Open Access is at a good level, but more promotion is needed.

Since 2012 the CUT Library participated in Open Access Week and organized events in order to promote Open Access and inform the scientific community.

Another important action was the approval, by the senate, of the establishment and the creation of the Cyprus University of Technology Open Access Author Fund in 2013. A local pharmaceutical company has undertaken the financing of the Fund, within the frame of its Corporate Social Responsibility.

Furthermore, in September 2013, the CUT Library upgraded its open source software DSpace for the purposes of compatibility with the European project Open Infrastructure for Research in Europe.

The CUT Library, in 2015, managed to renew the funding for another two years until 2017.

Recently, the senate approved the funding of Open Access publications by the university budget.

In 2016, the Cyprus government approved the Cyprus National Open Access Policy which complies with the current provision and best practices of the European funding program Horizon 2020 and policy for Open Access to scientific publications. The National policy provides guidelines and support for the implementation of Open Access to research outputs that are funded locally, is aligned with the European policies and is based on the already established infrastructure at the European level [9].

The Cyprus Research Council has not yet managed to provide coverage for OA fees. Therefore the CUT Library Committee suggested to the CUT Research Committee to enhance the Open Access APCs which was approved by the senate at the beginning of 2017. Furthermore, the Library committee is going to decide on the policy for the new fund.

The latest action was the implementation of ORCID. Researchers can create a research profile portal where they can obtain an ORCID or connect to an existing one. Through the CUT’s system, they have access to authenticate, read, update, add and synchronize research outputs with the new version of the Ktisis repository which runs on DSpace CRIS.

DSpace CRIS enhances and extends the institutional repositories and is an important software component in making them better able to cover all research activities, projects and to better support the people of an organization/institution with their research. Additional important operations in the research world that need descriptions and tracking, are “grants”, “patents”, “organization units” and “researcher profiles (people)” [7].

Integrating and organizing all these operations in combination with publications adds value in terms of “visibility, discovery and the understanding of the complexity of the research domain” [7].

All objects such as publications, projects, supervised theses and patents are linked to the researcher profile. At the same time, all researchers belong to the organization’s units, i.e. Departments and Faculties.

3. Open Access Publishing Funds

A growing number of universities and research organizations worldwide require from their researchers to provide Open Access to their peer-reviewed research articles, using the “Green” or “Gold” models. Furthermore, many universities have established separate funds for covering Article Processing Charges (APCs) regarding the Open Access publications.

The CUT Library conducted a report to examine the different Open Access funding policies around the world. The purpose of this report was to evaluate the different policies used by other universities and then to adapt them in order establish the Cyprus University of Technology Open Access Author Fund.

The CUT Library report covered the requirements for a university to become eligible for Open Access funding and analyzed approximately ninety OA policies from different universities from Australia, Austria, Canada, Germany, Netherland, Norway, Spain, Sweden, Switzerland, USA and United Kingdom. The results of the CUT Library report of the various policies used for financing Open Access publications will follow.

According to the report, the first to be examined was the case of the UK universities. It showed that academic institutions in the UK are adopting the Research Council UK [10] policy that supports both “Gold” and “Green” roads to Open Access. Funding covers both OA and hybrid journals.

Also, the Research Councils UK established an agreement with Elsevier where RCUK-funded authors can choose to comply with RCUK policy by either gold or green Open Access.

In Australia, there are numerous universities and research funds that provide authors with the financial support to cover OA publishing charges both for Open Access and hybrid journals. Since 2013, any publications arising from the NHMRC [11] and the ARC [12] are required to be deposited into an Open Access institutional repository freely accessible to the public within a 12 month period from the date of publication. Both the ARC and the NHMRC allow for up to 10% of the project budget (direct costs) to be allocated for the payment of article processing fees.

In Austria, two universities that established an Open Access funding were examined. The eligibility requirements for funding are that the journals must be listed in the Directory of Open Access Journals (DOAJ). There is also the flexibility which allows for one of the two universities to cover publication fees and in the case that the journal is not listed in the DOAJ, the publisher must be a member of the Open Access Scholarly Publishers Association (OASPA).

From an analysis of fourteen universities in Canada, it is shown that funded publications must be published in full Open Access journals. Articles that are published in hybrid journals or journals with embargoes (limiting Open Access to content after publication) are not eligible for funding.

The CUT Library report also examined thirteen universities in Germany which showed that Open Access funds do not support the hybrid model. Regarding the eligible journals, it is stated that journals must be listed in the Directory of Open Access journals but this is not mandatory for the funded publications.

In contrast with the German universities, five universities in Norway established an Open Access fund that covers only the journals which are listed in the Directory of Open Access Journals (DOAJ).

In the Netherlands, two institutions were recorded that support Open Access funding. Both institutions encourage the option to cover magazines listed in the Directory of Open Access journals but only one of the institutions defines that the hybrid model is not eligible for funding.

In Spain, it was recorded that the University of Barcelona [13] had established a funding program. The requirements that apply to journals are, that the journals must be listed in the Directory of Open Access Journals. Also, publishers must be members of the Open Access Scholarly Publishers Association (OASPA) or adhere to its Code of Conduct. Articles in hybrid journals are not eligible for funding. Additionally, the University of Barcelona in 2016 established an agreement with the Multidisciplinary Digital Publishing Institute (MDPI), offering to researchers a discount of 25% on the standard fee they charge to publish accepted articles in any of their journals.

Finally, Open Access funding policies that are adopted in the United States were examined. Forty-seven universities have established Open Access funding, of which only seven cover the hybrid model. Hybrid journals are not eligible for funding according to twenty-two out of the forty-seven universities in the United States. Also, twenty-two universities define that eligible publications for funding must be listed in the Directory of Open Access Journals (DOAJ) and in some cases the publisher must be a member of the Open Access Scholarly Publishers Association (OASPA) or demonstrate its adherence to the Code of Conduct.

The findings of the CUT Library report indicate that the basic eligibility requirements for funding publications that are adopted by many universities around the word, require that the journals must be fully Open Access and listed in the Directory of Open Access Journals (DOAJ). Also the publisher must be a member of the Open Access Scholarly Publishers Association (OASPA).

4. Existing Cyprus University of Technology Open Access Author Fund

The Library of Cyprus University of Technology established the “Cyprus University of Technology Open Access Author Fund”[14] for the promotion of Open Access and the support of academics and researchers in the publication of their research in peer reviewed OA journals and books, where the author bears the publication costs. The CUT is the first academic research institution in Cyprus to implement the creation of a Financing Fund to support scientific publications in Open Access sources.

The creation of the fund was approved by the Senate in September 2013, aiming to fund academic publications in journals and Open Access books. The pharmaceutical company Remedica has undertaken the financing of the Fund, with a maximum funding amount of €14000 per year.

The CUT Open Access Fund is available to all the members of the academic staff, and to post-graduate students. The total requested funding amount per faculty member must not exceed € 3.000 and must be spent within the period of one academic year. Each faculty member, according to their discretion and prioritization needs, is required to not exceed the amount distributed and provided that the expenditure is eligible.

The basic criteria for eligible publications are:

- The article should be published in an Open Access journal without any access restrictions.
- Open Access journals must be a member of the Directory of Open Access Journals.
- The publisher must be a member of the Open Access Scholarly Publishers Association and adhere to the Code of Conduct (OASPA Code of Conduct).

Eligible Costs for funding:

- The «CUT Open Access Author Fund» covers only the costs and fees associated with Open Access publishing activity (Open Access publication fees).
- If there is more than one author in a publication, the funding covers the percentage corresponding to each author who is a member of the University and applying for funding.

Eligible costs for funding do not include:

- Costs for reprints, graphics and other charges. Such costs are covered primarily by the budget of the Research Activities.
- Published articles in electronic journals or books that charge an annual fee, including the magazines or books using the model of delayed Open Access, or provide an Open Access only option.
- Published articles or books that are the result of research funded by other External Research Projects covering expenses of publication/issue. This applies in cases where they are fully financed from external sources, mainly from the European Commission, the Research Promotion Foundation and other organizations/institutions.

5. The CUT Library Hybrid Institutional Repository

The CUT Library institutional repository is a hybrid repository. This is defined by our research articles published in different scientific journals. This is not something that we control, it depends on the Publisher's policy.

When an article is deposited in KTISIS we verify, through the SHERPA/RoMEO database, the Publisher copyright policies and self-archiving.

In the case where there are no restrictions of access we attach the pdf file to the article entry in KTISIS. Otherwise, we enter the DOI of the article. Access to the article depends on if your institution subscribes to this journal.

6. The New Complementary Cyprus University of Technology Open Access Author Fund

After a new Library initiative at the beginning of 2017, the Senate approved the suggestion of CUT Library Committee for the creation of an Open Access fund to enhance the Open Access APCs. The main purpose of the CUT funding is to offer flexible publishing options for university researchers to make their research available in OA journals that cover a wide variety of subject areas.

The eligible criteria that we propose for the new Fund are the following:

1. Funded publications must be peer reviewed in order to be eligible for funding.
2. Publications should be published in an Open Access journal without any access restrictions using the Creative Commons Attribution (CC BY) or (CC-BY-SA) license. The Creative Commons Attribution licenses are also being used by the FP7 post-grant Open Access publishing Funds pilot [15] and the Research Councils UK. The (CC BY) license gives the opportunity for the re-use of the publication and the dissemination the research.
3. Open Access journals must be a member of the Directory of Open Access Journals.

Non-eligible journals:

1. The fund does not cover hybrid journals.
2. Journals with embargoes limiting Open Access content are not covered by funding. However, the authors can publish the post-print of their paper at the institutional repository of the university.

7. Cyprus University of Technology Open Access Author Fund Results and Discussion

Table 1 presents the total amount spent by the CUT Open Access Author Fund for the years 2013-2016. The total spent on APCs per year 2013 -2016 was € 45422. According to the given data there was a rise from 3 applications made in 2013 to 19, where an amount of € 24110,68 APC payments was recorded in 2016.

Table 1. Total applications and cost of APCs per year 2013 – 2016

Year	Application	Approved	Rejected	Amount (Euro)
2016	19	17	2	24110
2015	11	9	2	12058
2014	7	7	-	8386
2013	3	1	2	866
Total	40	34	6	45422

The pie chart in Figure1 illustrates the top Publishers based on the number of APC payments of the Open Access Author Fund between 2013 and 2016. Looking at the chart, we can see that BioMed Central captures a substantial part with 47%, Frontiers

12%, Hindawi 11%, Copernicus Publications and Public library of Science (PLoS) 9%, MDPI AG 6% and Dove Medical Press and BMJ 3% of the total APCs payments.

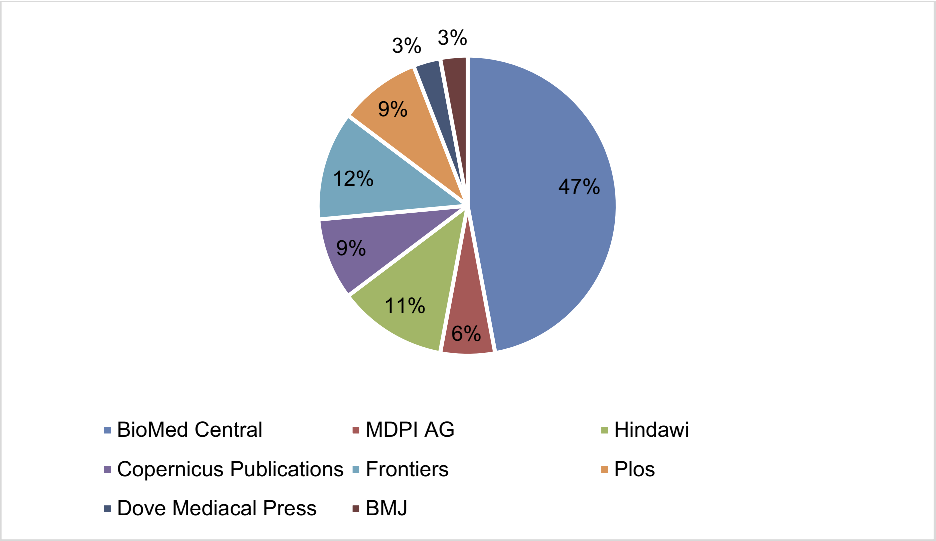


Figure 1. Top Publishers per year, 2013-2016

We can see clearly in figure 2 that the medical sciences have a higher publication number of OA journals. An important outcome is that the funding concerns are mainly health sciences departments with more than 70% of the total articles supported. As Tepltzky and Phillips said, “Nariani and Fernandez surveyed researchers at York University whose author-processing charges had been subsidized by institutional memberships in BMC, PLoS and Hindawi [16].

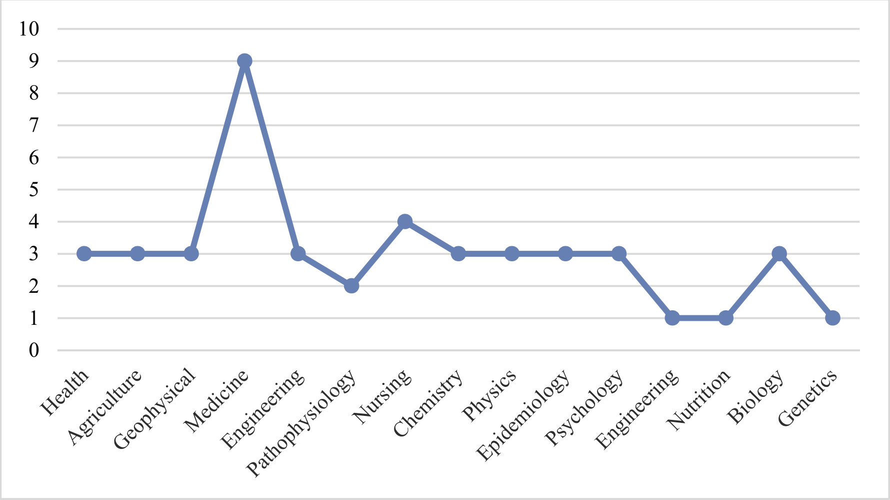


Figure 2. Publications in different subject areas

Table 2 gives information about the impact factor of Open Access journals. We can see that some OA journals have a high impact factor compared to subscription based journals. According to Bjork and Solomon's [17] research, the OA journals that fund publishing with article processing charges (APCs) are on average cited more than other OA journals. In medicine and health, OA journals funded in the last ten years receive about as many citations as subscription journals launched during the same period.

Table 2. Open Access journal impact factor

Journal	Impact Factor
Atmospheric Chemistry and Physics	5.114
BMC Plant Biology	4.085
BMC Public Health	2.264
Remote Sensing	3.036
BMC Infectious Diseases	2.825
BMC Psychiatry	2.891
Evidence Based Complementary and Alternative Medicine	1.931

8. The APC Model

Fully OA Publishers, such as Biomed Central (BMC) and PLoS, have been using APCs as their Prime Business model since 2002 [18]. In our case, BMC capture a substantial part with 50%, Hindawi 12.5%, Frontiers 9.5% and Public Library of Science (PLoS) 6.5% of the total APCs. Another important outcome is that the funding concerns mainly health sciences departments with more than 70% of the total articles supported. As Tepltzky and Phillips said, "Nariani and Fernandez surveyed researchers at York University whose author-processing charges had been subsidized by institutional memberships in BMC, PLoS and Hindawi [19, 20]. Their qualitative study showed that author funding can provide an incentive for researchers to publish in OA journals, but their cohort and selection of journals was limited". These findings have led us to think that we need to find another way to support OA which must be more flexible and cover more journals in various research areas" where a publisher offers an "author-pays" model for Open Access and simultaneously sells the end product to libraries with a subscription. Stephen Pinfield wrote, "commercial publishers such as Elsevier and Wiley, which dominate the subscription market, are now also capturing a substantial part of the OA APC market" [21]. We are troubled by this and we are not sure if the hybrid model is the right way for OA, especially if we take into consideration Björk and Solomon's findings that APC costs for "hybrid OA" journals are significantly more and provide further evidence to support their argument that the "hybrid market" may not be operating effectively in terms of price sensitivity [22].

However, the CUT Library aims to adopt "offsetting" agreements with hybrid publishers for fair prices in hybrid journals, so that the phenomenon of double dipping is avoided. Double dipping is the term used to describe a publisher gaining from two income streams, APCs and subscriptions, in a way that its overall income from the same customer rises [20]. Publishers have also recognized the need to review their subscription prices. Several publishers have given general "offsetting" undertakings of this sort [23]. Elsevier and Dutch universities have agreed on a new subscription deal which will begin transitions to Open Access publishing. The Association of Universities in the Netherlands (VSNU) has also made the unprecedented demand that Elsevier and other publishers must allow their academics to make their papers Open Access at no extra charge [24].

The following chart (Figure 3) presents a research carried out under the European program PASTEUR, which shows that many institutions enhance the Open Access fund from their own fund or allow APCs to be paid from a research grant.

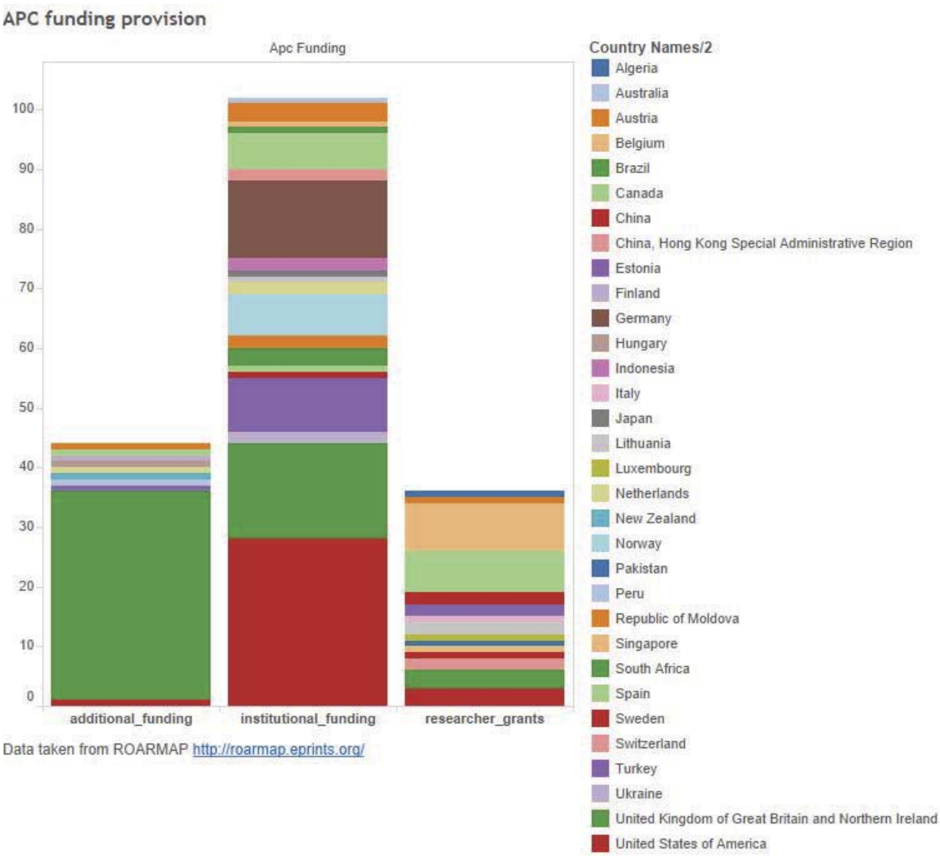


Figure 3. APC funding provisions (taken from: ROARMAP http://roarmap.eprints.org/)

On the other hand, the Research Councils UK's (RCUK) Open Access policy does not allow cost for Open Access publication fees in research applications, and funding for Open Access is raised by the RCUK through a block grant awarded directly to research organizations.

9. Conclusions

The CUT Library will continue to play a dynamic role in the development of the Open Access framework by supporting researchers, Open Access journals and publishers that provide full Open Access. Moreover, the CUT Library aims to make agreements with publishers, in order to motivate authors to publish their work on an Open Access basis. The agreements with different Open Access publishers are essential to the CUT Library for increasing Open Access publications of the university in wide research areas. More and more stakeholders, like the European Union, state that research organizations and universities support Open Access. The new funding supported by the university budget will be available for financing research output published in full Open Access journals, even if the publishers are not members of OASPA. Furthermore, research institutions must negotiate with publishers and find a solution to avoid double dipping.

References

[1] SPARC (2017). Open Access. Retrieved January 5 2017, Available from https://sparcopen.org/open-access/
[2] Suber, P. Open Access overview: focusing on Open Access to peer-reviewed research articles and their preprints, 2015. Available from: http://bit.ly/oa-overview.
[3] Stevan, H., Optimizing Open Access Policy, The Serials Librarian 69 (2015), 133-141.
[4] Guédon, J.-C. The “Green” and “Gold” Roads to Open Access: The Case for Mixing and Matching, *Serials Review* **30** (2004), 315–328.
[5] PASTEUR Retrieved April 10 2017, Available from http://pasteur4oa-dataviz.okfn.org/apcs.html
[6] Horizon policy Retrieved April 10 2017, Available from European Commission https://ec.europa.eu/digital-single-market/en/news/open-access-scientific-publications-horizon-2020-projects=
[7] DSpace-CRIS Retrieved April 10 2017, Available from https://wiki.duraspace.org/display/DSPACECRIS/DSpace-CRIS+Home
[8] Giannoulakis S., Zervas M., Artemi P. *Promoting Open Access at the Cyprus University of Technology: survey results*. In: 5th International Conference on Information Law 2012, Corfu, June 29-30, 2012.
[9] Koukounidou, S. (2016). Cyprus adopts National Open Access Policy. Retrieved January 10, 2017, from https://blogs.openaire.eu/?p=923
[10] RCUK Policy on Open Access and Supporting Guidance Retrieved April 10 2017, Available http://www.rcuk.ac.uk/documents/documents/rcukopenaccesspolicy-pdf/
[11] National Health and Medical Research Council (NHMRC) Retrieved April 10 2017, Available https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-open-access-policy?
[12] Australian Research Council (ARC) Retrieved April 10 2017, Available http://www.arc.gov.au/arc-open-access-policy
[13] University of Barcelona funding for OA Retrieved April 10 2017, Available http://crai.ub.edu/en/crai-services/open-access-ub/funding
[14] Cyprus University of technology author fund Retrieved April 10 2017, Available http://library.cut.ac.cy/en/node/1259
[15] FP7 post-grant Open Access publishing Funds Retrieved April 10 2017, Available pilot https://www.openaire.eu/postgrantoapilot
[16] Nariani, R. & Fernandez, L., 2011. Open Access Publishing: What Authors Want. Collage & Research Libraries, p. crl-203.
[17] Bjork, B.C. & Solomon, D. (2012). Open Access versus subscription journals: A comparison of scientific impact. BMC Medicine, 10, 73.
[18] Delamonthe, T. & Smith, R. (2004). Open Access publishing takes off. BMJ.
[19] Teplitzky, S. and Phillips, M. (2016) Evaluating the impact of Open Access at Berkeley: Results from the 2015 survey of Berkeley research impact initiative (BRII) funding recipients’, College and Research Libraries 77(5), pp 568-581. doi: 10.5860/crl.77.5.568
[20] Nariani, R. & Fernandez, L., 2011. Open Access Publishing: What Authors Want. Collage & Research Libraries, p. crl-203.

[21] Pinfield, S., Salter, J. & Bath, P., 2016. The Total Cost of Publication" in a Hybrid Open-Access Environment: Institutional Approaches to Funding Journal Article-Processing Charges in Combination With Subscriptions. Journal of the Association for Information Science and Technology.
[22] Bjork, B.C. & Solomon, D., (2014). Developing an effective market for Open Access article processing charges. Final report, (March), p.71. Retrieved January 10 2017, from: https://wellcome.ac.uk/sites/default/files/developing-effective-market-for-open-access-article-processing-charges-mar14.pdf
[23] Elsevier. (2014). No double dipping policy. Elsevier website. Retrieved April 10 2017, Available http://www.elsevier.com/about/open-access/open-access-policies/
[24] Dutch universities and Elsevier reach deal over Open Access. Retrieved April 10 2017, Available https://www.timeshighereducation.com/news/dutch-universities-and-elsevier-reach-deal-over-open-access

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-56

OpenAIRE: Supporting the H2020 Open Access Mandate

Vasiliki (Sylvia) V. KOUKOUNIDOU[1]
University of Cyprus/OpenAIRE NOAD, Cyprus

Abstract. This paper will focus on the practices used by the OpenAIRE Cyprus National Open Access Desk as part of a pan-European network, for the implementation of the European Union's vision and policies for Open Science and Open Access to knowledge. Furthermore, the purpose of this paper is to present the scope, the role and the actions of the European Project OpenAIRE – Open Access Infrastructure for Research in Europe, which since 2009 is working towards the support of the European Commission's policies for Open Science. By definition "Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. The idea captures a systemic change to the way science and research have been carried out for the last fifty years: shifting from the standard practices of publishing research results in scientific publications towards sharing and using all available knowledge at an earlier stage in the research process [1]".

Keywords. open science, Open Access, OpenAIRE, Horizon2020, Cyprus NOAD

1. Introduction

In the past few years, research in Cyprus has increased considerably due to the integration of the country into the European Union. The establishment of six more universities, both public and private, in addition to the University of Cyprus that already existed, has also contributed to this increase. The main funder of research in Cyprus is the EU, which has supported research by providing the majority of the funding. Horizon2020 found the research community of Cyprus to be supported by €73.608.521 (234 participations) [2], a considerably important amount for a member state struggling with economic difficulties. Since the EU has given a significant boost to research activity nationally by being the primary funder of research, its role is considered particularly important for further support of research both for researchers and the organizations that employ them. Having this in mind the policies announced regarding the research frameworks of the European Union, find Cyprus following and implementing any necessary actions for compliancy. Since Open Science and Open Access is one of the main policies of the EU [3], Cyprus had to establish and implement all the necessary infrastructure towards the successful support of these policies. Of course there are still a lot of things to work on, (e.g. promoting the benefits, linking Open Access to evaluation process, ethical issues etc.) but the influence of

[1] Vasiliki (Sylvia) V. Koukounidou, University of Cyprus / OpenAIRE NOAD, Cyprus 75 Kallipoleos Ave. Nicosia Tel.:+35722892168 Fax: +35722893435 E-mail: Koukounidou.vasiliki@ucy.ac.cy.

European champions of Open Access and all local relevant activities are expected to achieve the adoption of "Open Science Culture".

2. The EU Funded Projects for the Support of the Implementation of Open Access Policies

A project for Open Access Infrastructure for Research in Europe – OpenAIRE[2] launched in December 2009 as one of the largest funded project of the European Union. OpenAIRE aimed to support the implementation of Open Access in Europe. Its main scope was to provide the means to promote and realize the widespread adoption of the Open Access Policy, as set out by the ERC Scientific Council Guidelines for Open Access and the Open Access pilot launched by the European Commission. OpenAIRE was a three-year project, that established the infrastructure for researchers for their support in complying with the EC OA pilot and the ERC Guidelines on Open Access. It also provided an extensive European Helpdesk System, based on a distributed network of national and regional liaison offices in 27 countries, to ensure localized help to researchers within their own context. The OpenAIRE portal was built and e-Infrastructure for the repository networks and exploration of scientific data management services together with 5 disciplinary communities was created. It also provided a repository facility for researchers who did not have access to an institutional or discipline-specific repository. The importance of the subject, lead to two additional subsequent projects. 2011-2014 saw the project OpenAIREplus – 2nd Generation of Open Access Infrastructure for Research in Europe. A 30 month project, funded by the EC 7th Framework Programme, was extending the mission further to facilitate access to the entire Open Access scientific production of the European Research Area, providing cross-links from publications to data and funding schemes. This large-scale project brought together 41 pan-European partners, including three cross-disciplinary research communities. The project capitalised on the successful efforts of the OpenAIRE project which was rapidly moving from implementing the EU Open Access Pilot project into a service phase, enabling researchers to deposit their FP7 and ERA funded research publications into Open Access repositories. The publication repository networks expanded to attract data providers from domain specific scientific areas. Innovative underlying technical structures was deployed to support the management of and inter-linking between associated scientific data. Access to and deposit of linked publications via the OpenAIRE portal was supported by the Help Desk, and OpenAIRE's collaborative networking structure extended to promote the concept of open enhanced publications among user communities. Liaison offices in each of the project's 31 European countries worked to support the needs of researchers in Europe. The project also actively contribute to the international discussion for common standards, data issues and interoperability on a global level.

A year after the beginning of the European Commission's (EC) Horizon 2020 (H2020) Research Framework Programme, OpenAIRE entered a new phase of funding with a continuation project. OpenAIRE2020, started in January 2015 and until mid-2018 it will work towards the implementation of the H2020 Open Access policy for publications, being one of the key European infrastructures supporting the H2020 Open Research Data Pilot. OpenAIRE is a Pan-European project with a well-established

---

[2] https://www.openaire.eu/

network of 50 organizations, all Europe's member states and beyond. OpenAIRE supports the European Commission Open Access policy by providing infrastructure for researchers to comply with the corresponding requirements for Open Access to research results. Other activities of the project include: the collaboration with the national funders for reinforcing the infrastructure's research analytic services; an APC Gold OA pilot for FP7[3] [4] publications with collaboration from LIBER [13; novel methods of review and scientific publishing with the involvement of hypotheses.org[4] [4;, a study and a pilot on scientific indicators related to open access with CWTS's [12] assistance; legal studies to investigate data privacy issues relevant to the Open Data Pilot; and international alignment with related networks elsewhere with the involvement of COAR [11].

2.1. National Open Access Desks Network

Accepting the statement that "Open Access is global – but implementation is local, and consequently European diversity requires extensive knowledge of national research practices, languages, administrative procedures and technical infrastructures [4]" OpenAIRE established a reference point in every member state and associate country. 37 pan-European advocacy nodes, known as National Open Access Desks (NOADs), have among others, the task to reach out to researchers and project coordinators of EC-funded projects to inform them of the EC OA mandate and to align their local infrastructures with a common European platform. The European Commission, as one of the most significant funders of research and facilitator of collaborative and cross-disciplinary scientific activities, required beneficiaries of its previous funding programme, the 7th Framework Programme (FP7) to make their best effort to ensure OA to peer-reviewed articles (Open Access pilot). Its new funding programme, Horizon2020, with the investment of nearly 80 billion Euros into competitive research includes the mandate of the publication output of all EC-funded projects be made open. Furthermore, OpenAIRE supports the Open Research Data Pilot and mandate which aims to maximize access to and re-use of research data generated by projects. A post-project Gold Open Access Pilot is in implementation (until April 2017), and other scholarly communication topics such as peer review and metrics are explored. The project also supports efforts of individual partner countries for national Open Access initiatives and OA Policies. The perspective of the Cyprus NOAD participation in the project is reported in this paper.

3. Cyprus as a National Open Access Desk

Cyprus via the University of Cyprus Library, has been participating in OpenAIRE since the beginning of the project in 2009. The main objectives of a NOAD are among others to support, promote and disseminate all relevant information regarding the policies of Open Access to all possible stakeholders. Cyprus achieved this by establishing a collaboration mechanism among researchers, institutions, funding organizations, EC National Reference Points, repository managers, librarians and the librarians

[3] https://postgrantoapilot.openaire.eu/
[4] https://www.openaire.eu/hypotheses.org

association – a real human network. Three parallel approaches are in continuous process for this achievement.

I. Central approach: Cooperation with the Research Promotion Foundation and the Directorate General for European Programmes, Coordination and Development (the local NPRs for OA) in order to forward informative material via email for Open Access & OpenAIRE project. During the last two years this cooperation along with the creation of the OA working group, has played an important role to the implementation of the National OA policy.
II. Cluster approach: a) Co-organize/Participate in conferences of librarians/information scientists who acted as multipliers because they were able to disseminate the obligation derived by H2020 projects to their institutional researchers. b) Identification and participation with posters or papers in conferences / information days that took place in our region through which researchers were reached.
III. Individual approach: Emails to the Cypriot coordinators/partners of SC39 FP7 and H2020 projects and phone calls were answered for questions & help requests.

3.1. Relevant EU Projects Involvement

In conjunction with participation in the OpenAIRE network, Cyprus NOAD had use the expertise and the advocacy material produced by completed and acting relevant projects such as the RECODE - Policy (RECommendations for Open Access to Research Data in Europe) [5], PASTEUR4OA - (Open Access Policy Alignment Strategies for European Union Research)[6] (acted also as Key Node), FOSTER - (Facilitate Open Science Training for European Research) [7] (hosted training event). Cyprus NOAD is working for the fulfillment of all the above mentioned tasks. Resulting, in Cyprus, of all three repositories being compliant with the European infrastructure, so that all local researchers could have the possibility to deposit to an OpenAIRE compliant repository. Apart from the local infrastructure OpenAIRE provides Zenodo. Zenodo [8] a catch-all repository for EC funded research created by CERN, an OpenAIRE partner. Among others, Zenodo helps researchers deposit their research output (publication or data) and fulfil their obligations to their funders. They are also able to receive credit by making the research results citable and through OpenAIRE, integrate them into existing reporting lines to funding agencies like the European Commission. Citation information is also passed to DataCite and onto the scholarly aggregators.

4. The Implementation of the Cyprus National Open Access Policy

Open Access and open science awareness activities are also part of the NOADs [4] tasks. In Cyprus, activities are taking place around the year but especially during the International Open Access week, where usual dissemination activities take the form of one or two days conference, webinars, focused meetings etc. Important work has been done for the coordination of the Cyprus Open Access working group. The outcome of the excellent collaboration between the different stakeholders involved in the Working Group for Open Access (consisting of the National Point of Reference for Open

Access, namely the Directorate General for European Programmes, Coordination and Development, the Research promotion foundation, local academic institutions and research funders), coordinated by the OPENAIRE2020 Cyprus NOAD (University of Cyprus Library), and highly supported by the coordinators of the project PASTEUR4OA (Open Access Policy Alignment Strategies for European Union Research), was the approval of the document for the National policy for Open Access [9]. On the 25th of February 2016, the Council of ministers of the Republic of Cyprus, had finally approved the adoption of the National policy for Open Access in Cyprus.

The National policy provides guidelines and support for the implementation of Open Access for research outputs that are funded locally, aligned with the European policies and based on the already established infrastructure at European level (e.g. OpenAIRE).

Several activities are taking place in Cyprus in order to support the adoption of the national policy and ensure the success of its implementation. For example, the new funding schema of the National research funder "Restart 2016-2020"[5], announced late-2016, included the requirement for Open Access to Research outputs of the projects funded by the National Research Promotion Foundation [2]. Compliance of the researchers with the national open access policy, cannot yet be reported. Institutional policies will be the next step in order to be aligned to the European and National policy.

5. Conclusion

Open Access to research results is an essential part of Open Science, which aims to make science more reliable, efficient and responsive, in order to enforce innovation opportunities. Openness also improves the reproducibility of research results and it might introduce new and perhaps unexpected audiences to a researcher's work [10]. For Cyprus, as a small island, research and innovation is a crucial asset. Keeping in mind that findable, accessible, interpretable and re-usable (FAIR) knowledge is essential for the public sector, the financial sector, innovative businesses, academia and citizens –which all have an essential part to play in order to invent further and new aspects of knowledge – adopting and implementing all relevant policies and tools for the successful establishment of the "Open science culture" will only benefit Cyprus and Europe as a whole. OpenAIRE is important infrastructure that has played and will continue to play a key role in these strategic subject and local infrastructures via its local NOAD.

References

[1] Open Science (Open Access) - Horizon 2020 - European Commission [Internet]. Horizon 2020. 2017 [cited 2 April 2017]. Available from: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access

[2] Research Promotion Foundation [Internet]. Research.org.cy. 2017 [cited 5 April 2017]. Available from: http://www.research.org.cy/

[3] Research and Innovation-Open science European Commission, 2017 [cited 5 April 2017]. Available from: https://ec.europa.eu/research/openscience/index.cfm.

[5] http://www.research.org.cy/el/restart-2016-2020/

[4] OpenAIRE - OpenAIRE [Internet]. Openaire.eu. 2017 [cited 4 April 2017]. Available from: https://www.openaire.eu/

[5] RECODE [Internet]. Recodeproject.eu. 2017 [cited 7 April 2017]. Available from: http://recodeproject.eu/

[6] PASTEUR4OA [Internet]. Pasteur4oa.eu. 2017 [cited 7 April 2017]. Available from: http://pasteur4oa.eu/

[7] FOSTER – Facilitate Open Science Training for European Research [Internet]. FOSTER – Facilitate Open Science Training for European Research. 2017 [cited 1 April 2017]. Available from: https://www.fosteropenscience.eu/

[8] Zenodo - Research. Shared. [Internet]. Zenodo.org. 2017 [cited 1 April 2017]. Available from: https://zenodo.org/

[9] A snapshot of Open Data and Open Science policy in Europe [Internet]. Sparc Europe & Digital Curation Centre; 2017 p. 2. Available from: http://sparceurope.org/open-data-open-science-policy-europe/

[10] The Open Access Citation Advantage Service (OACA) list [Internet]. Sparceurope.org. 2017 [cited 4 April 2017]. Available from: http://sparceurope.org/what-we-do/open-access/sparc-europe-open-access-resources/open-access-citation-advantage-service-oaca/oaca-list/

[11] COAR – Towards a global knowledge commons [Internet]. Coar-repositories.org. 2017 [cited 3 April 2017]. Available from: https://www.coar-repositories.org/

[12] CWTS - Centre for Science and Technology Studies - Leiden University [Internet]. CWTS. 2017 [cited 4 April 2017]. Available from: http://www.cwts.nl/Home

[13] LIBER (Ligue des Bibliothèques Européennes de Recherche – Association of European Research Libraries) [Internet]. LIBER. 2017 [cited 7 April 2017]. Available from: http://libereurope.eu/References

[14] Research Promotion Foundation [Internet]. Research.org.cy. 2017 [cited 5 April 2017]. Available from: http://www.research.org.cy/el/restart-2016-2020

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-62

New Toolkits on the Block: Peer Review Alternatives in Scholarly Communication

Birgit SCHMIDT[a,1] and Edit GÖRÖGH[a]
[a] *Göttingen State and University Library*

Abstract. Peer review continues to play a central role in scholarly communication processes, however, over the last decade the concept has branched out in terms of methods, platforms and stakeholders involved. The paper demonstrates how alternative peer review tools and methods are instrumental in further shaping the communication of scholarly results towards Open Science. The analysis is based on the examination of various *review methods* (peer commentary, post-publication peer review, decoupled review, portable or cascading review) and review tools and services (publishing platforms, repository-based, and independent reviews). Besides the differences in operation and functionality, these new workflows and services combine common features of network-based solutions and collaborative research applications with varying degrees of openness (e.g. regarding participation, identities and/or reports). They, therefore, represent good examples of Open Science, in terms of transparency and networking among researchers.

Keywords. open science, open peer review, OpenUP, review alternatives

1. Introduction

Open access has by now become a core strategy for European research, aiming at wide knowledge circulation and fostering innovation. Embedded into a broader discourse about open science this represents a transformative approach to research, based on digital technologies and methods as well as new collaborative tools. There are still several challenges which have to be addressed – e.g. interoperability of infrastructures and services, intellectual property rights and quality assessment – and these in turn have an impact on all facets of the scholarly communication process.

The growing dissatisfaction with the traditional scholarly communication process and publishing practices has triggered a proliferation of alternative dissemination and assessment methods. In particular, scientific papers are increasingly publicly scrutinized by peers and several of these cases pointed out significant scientific flaws or even outright misconduct which has led to retractions of papers [1]. Considering the growing diversity of platforms and channels by which these comments and reviews are communicated, there is an urgency to assess the current status and gather best practices which can further guide developments in this field.

The EU-funded OpenUP project [2] addresses key aspects and challenges of the currently transforming research landscape and aspires to come up with a cohesive

[1] Birgit Schmidt, Corresponding author, Göttingen State and University Library; Platz der Göttinger Sieben 1, 37073 Göttingen; E-mail: bschmidt@sub.uni-goettingen.de.

framework for the review-disseminate-assess phases of the research life cycle that is fit to support and promote Open Science. The primary objectives of the project are (1) to identify ground-breaking mechanisms, processes and tools for peer-review for all types of research results (e.g., publications, data, software), (2) to explore innovative dissemination mechanisms with an outreach aim towards businesses and industry, education, and society as a whole, and (3) to analyze a set of novel indicators that assess the impact of research results and correlate them to channels of dissemination. The project employs a user-centered, evidence-based approach, engaging all stakeholders (researchers, publishers, funders, institutions, industry, and the general public) in an open dialogue through a series of workshops, conferences and training, while validating all interim results in a set of seven pilots involving communities from four research disciplines: life sciences, social sciences, arts & humanities, and energy.

Discourse on alternative dissemination and evaluation methods increases transparency, opens up the scholarly communication process, but most of all, encourages researchers to discuss research practices and share their results

2. Alternative Takes on Peer Review Processes

Peer review in the context of scholarly communication is a *concept* and not a narrowly defined *methodology*. As such it can be unbounded from the journal paper and applied to any research product [3]. In addition, peer review is very versatile: it can be employed for e.g. evaluating scientific results, research data, research proposals and the performance of projects. It is used in teaching to assess portfolio information about the teaching of an instructor, in pedagogy to enhance students′ critical skills, in medicine, as the process by which a committee of physicians examines the work of a peer and determines whether the physician under review has met accepted standards of care. In all these case, the common theme is the scrutiny of one's work by fellow workers/peers. However, although the primary goal is the same, the methods for putting peer review into practice vary across journals and disciplines [4].

In the age of the Internet and proliferation of communication channels, the printed and peer reviewed journals and books are no longer the principle vehicles by which research is disseminated [5]. The new tools, platforms and services enrich the academic publishing scene, and provide functionalities to continuously revisit and re-evaluate the process and the outcomes of the scholarly discourse. The publishing scene has come to include open access e-journals, independent publishing platforms, overlay journals based on repository content, blogs and journal clubs. As dissemination methods diversify the scholarly communication scene, the related review processes have become varied. Depending on the dissemination channel they are connected to, we can find review tools and methods from open peer review, revealing the reviewer's identity and/or the review report, through post-publication review, cascading or decoupled review to collaborative review and community based commenting. If we situate these methods of dissemination and evaluation on a scale, they show a definite move away from the traditional closed peer review process to a more open and transparent methodology with varying openness in identity, documentation, participation, and time. The new, innovative tools incorporate the basic principles of open science by employing open, collaborative and network-based publishing and review methodology.

In the following we will examine how the alternative peer review tools and methods shape the communication of scholarly results and how they contribute to the strengthening discourse of open science. The analysis is based on the examination of various *review methods* (peer commentary, post-publication peer review, de-coupled review, portable or cascading review) and the employed *communication tools* (commenting, annotation/tagging, reports, evaluation templates). Besides the differences in operation and functionality, these new workflows and services carry the common features of network-based solutions, employment of digital technologies and collaborative research applications.

3. Defining Open Peer Review

The concept of "open peer review" is rather controversial, because presently it is being used for several fairly different models of peer review. In most cases, open peer review refers to the review process in which the identity of the reviewers is disclosed (examples) or the review itself is accessible for the public [6]. However, there are studies which go beyond such simplified interpretations and include other attributes of the review process in the definition. The present analysis relies on the definition of Open Peer Review as it has been proposed by OpenAIRE [7]. Open Peer Review is considered as an umbrella term which comes in different flavors, where the traditional process is opened up by modifying one or more aspect to make it more inclusive, transparent and/or accountable. Based on a literature review seven distinct traits could be identified (Ibid.).

- **Open identities:** Authors and reviewers are aware of each other's identity.
- **Open reports:** Review reports are published alongside the relevant article.
- **Open participation:** The wider community to able to contribute to the review process.
- **Open interaction:** Direct reciprocal discussion between author(s) and reviewers, and/or between reviewers, is allowed and encouraged.
- **Open pre-review manuscripts:** Manuscripts are made immediately available (e.g., via pre-print servers like arXiv) in advance of any formal peer review procedures.
- **Open final-version commenting:** Review or commenting on final "version of record" publications
- **Open platforms:** Review is de-coupled from publishing in that it is facilitated by a different organizational entity than the venue of publication.

The possession of at least one of the first three traits is considered sufficient for qualifying as Open Peer Review (basically due to the fact that this covers all but one of 122 definitions considered in the literature review) (Ibid.). Based on a survey conducted by OpenAIRE it turns out that a majority of respondents provided support for opening up the discussion between authors and reviewers, to publish review reports and to allow open comments on final papers [8]. Over a third was reluctant to make papers available online before peer review and nearly every second respondent opposed against revealing reviewers' identities (Ibid.).

When it comes to peer review methods or processes there are further options to consider. We outline the ones which are relevant for our discussion:

- **Peer commentary** refers to the feature that other interested parties are explicitly invited and/or generally be enabled to provide comments at some stage in the publication process. This functionality may be offered based on the pre-publication and/or the final published version. Typically, peer commentary is considered as a *supplement* to peer review, but not as a *substitute* for it [9]. Examples include e.g. the 2-stage publication process as offered by Copernicus Publications [10] and peer-reviewed papers published by the journal PeerJ (peerj.com) (the functionality is not enabled for preprints).
- **Post-publication peer review** is based on the published version and can take the form of rather lightweight peer commentary or a more serious review report. However, such a process is not necessarily moderated and then merely takes place based on intrinsic motivations of the individual reviewer. Examples include e.g. the F1000 Research (f1000research.com) publishing platform where papers which pass an in-house quality check are published immediately. These publications are then subject to formal peer review and referees' reports are published on the site and all referees are named. Authors then have the opportunity to respondent to these reviews and are encouraged to revise their papers [11].
- **De-coupled review** refers to unbundling the review service from the publishing service [12]. In this case a paper may first be deposited in a disciplinary or institutional repository, then be subject to a review by an independent review service, followed by formal publishing in a scholarly journal. Examples for this model are e.g. Rubriq (www.rubriq.com) and Peerage of Science (www.peerageofscience.org). Typically these services partner with journals which after the completion of the review and revision process can make an offer to the authors to publish their paper.
- **Portable or cascading review** treats review reports (and revisions) as essential context information for submitted manuscripts which are then moved together through the whole peer review process. Examples of this model have e.g. been introduced by mega-journals which started to reuse reviews from journals which have rejected the manuscript in question [13].
- **Journal clubs** are a post-publication collective review exercise in which participating individuals come together to critique and keep-up-to-date with relevant literature. However, there is no standard process of conducting an effective journal club (for a systematic review cf. [14]). Today, publishing platforms have emerged which facilitate the publication of review papers which may have been emerged based on journal clubs (e.g. The Winnower (thewinnower.com)).

4. Evaluating Review Tools

Alternative review methods and services provide innovative ways for researchers to communicate their scientific results at smaller, communal level or at a wider, global level, and to evaluate each other's work. By assessing the problems and criticism which the peer review system faces, review tools and services can be categorized and evaluated on the basis of their functionality and sustainability within the current scholarly communication system. Overall, open peer review services and tools can be grouped in four categories: 1) publisher-based platforms or journals, 2) independent peer review services with openness functionalities, 3) repository-based solutions and 4) commentary/annotation tools.

4.1. Journal Editors and Publishers

Journal editors and publishers have been major drivers in introducing alternative peer review methods. Some moved away from the traditional method of reviewing by shortening the publication time and by making the review process partially or entirely transparent. The openness of the review process is ensured by publishing reports alongside articles and by strongly urging, but not necessarily mandating the disclosure of the identity of reviewers.

Another feature of open peer review is also incorporated in the operation of the majority of these publishers. The review process is turned into a collaborative effort either through the communication among editors and authors, or through initiating discussion within research communities. This starts early on from supplementing the traditional peer review process by peer invited commentary: S. Harnad coined the term “open peer commentary” in the 1970s and introduced commentary from a group of peers on selected articles in Behavioral and Brain Sciences [15]. The first journal to introduce open peer review was the British Medical Journal which requires reviewers to sign their report and publishes the papers together with review reports and reviewers’ names [16]. Several publishers followed in the early 2000s, introducing a range of pre- and post-publication open peer review workflows. They employ different degrees of collaboration: while eLife (elifesciences.org) ensures the discussion of the editor and the reviewers about the submitted manuscript, Frontiers (frontiersin.org) established a “Collaborative Review Forum,” which unites authors, reviewers and the Associate Editor [17]. Copernicus Publications allows the widest collaboration by involving the research community early on in the review process. Their “Interactive Peer Review” supplements the evaluation of the reviewers with the comments from the scientific community [18].

4.2. Independent Peer Review Services

Independent peer review services decouple the review process from the publishing platform(s). The review service is not affiliated with a journal or publishing house, thus the evaluation is not skewed by the standards of the respective publisher. The process allows different degrees of openness and involvement from authors and reviewers. Some of these platforms offer opt-in functionalities to publish review reports and/or disclose reviewer names in relation to publications (pre and post-publication). Some publishers provide reviewers with an actionable link which enables direct reporting of reviewer activity to such platforms (e.g. Publons (publons.com)). These services also allow for author-directed workflows, ranging from authors setting the time frame for the evaluation (Peerage of Science), through contacting reviewers to participate in the process (SciOR (science-open-reviewed.com)), to deciding the degree of openness they are comfortable with (PubPeer (pubpeer.com), Publons). These platforms, in general, advocate a network-based approach where collaboration between authors, editors and reviewers is strongly encouraged in order to improve the paper and the overall review experience. Community interaction can further step up the quality of scientific research by enabling innovative approaches [19].

As the quality of the review process, including both the quality control function of the review in regard to the materials passing through the system, and the quality of reports prepared by the reviewers, has become a major concern in the established review system, independent review services introduce various tools to contest this problem. Standardized evaluation forms (Rubriq Score card) guide the reviewers in their evaluation urging them to tackle major points of relevance and quality in their review report. Furthermore, evaluation of the reviewers (Peerage of Science: Peerage Essay Quality scores) by fellow reviewers provides a clearer picture about the researchers involved in the process. In some cases, anonymity of the author is requested (PubPeer, Publons, Peerage of Science) in order to keep the process bias-free. There is a service specializing on the quality control of the review process, offering journals an evaluation of the transparency and integrity of their review process: the Peer Review Evaluation (PRE) review services created a seal of approval in the form of a PRE badge which ensures quality publishing in regard to both the articles being peer reviewed, and to the publishers authors can choose from [20].

Although these review platforms operate independently from publishers, they may be connected to a chosen set of journals. The journals, the platforms are working with, accept articles for publishing based on the recommendations of the review platforms. Thus, besides the primary function of managing the review process for scientific outputs, the review services evaluate the fit of the paper to a variety of journals. The match between the article and the journal can be made even if the review service is not connected to the author's preference of publisher; the author is free to submit his/her peer-reviewed work to any journal with a link to the completed process (Peerage of Science).

Peer review platforms carry several benefits for reviewers. They employ a range of methods to recognize and reward review work. At Publons the peer review and post publication activity factors into the paper's Altmetric scores (new silver line in the Altmetric donut). Furthermore, the researcher's review activity is automatically exported to their ORCID ID adding a permanent record in their research history. Rubriq goes one step further and provides besides the academic reward forms, a financial compensation for the review work, alternatively they offer a contribution to the reviewer's organization fund or a donation to a charity in the research community. Thus, the methods and tools may vary in rewarding review work, but it is a common feature at these review services that the work and time of the researchers is acknowledged.

4.3. Repository-Based Solutions

Repository-based solutions are gaining momentum in the publishing discourse. The Internet facilitates immediate communication and dissemination of (preliminary) research results. In particular, uploading to and making preprints available in disciplinary and/or institutional repositories facilitate a rapid distribution of research findings. The pioneering and successful example of the arXiv which covers preprints in the field of physics, mathematics and further quantitative disciplines (launched in August 1991) found followers in other fields, in some only 20 years later. Due to more receptive audiences, bioRxiv, which is an arXiv-licensed but independent preprint server provided by CSHL Press for the life sciences, launched in 2012, was soon followed by AgriXiv in agriculture and allied sciences; engrXiv for engineering, and SocArxiv in social science [21], and by the most recent addition of paleorxiv, soon to be launched in 2017.

Preprints are increasingly being recognized by the publishing industry. Numerous journals and publishers exempt preprints from copyright restrictions allowing deposit of and access through repositories, institutional and/or personal websites (SHERPA/RoMEO database (www.sherpa.ac.uk)). However, in regard to open science and the re-use of open access materials, the free availability of these manuscripts does not necessarily imply a free re-use option. "Recent data show that authors uploading their work to bioRxiv choose the most restrictive license on offer – retaining full copyright – for their work," most probably to ensure full control over their work [22], Some publishers, like eLife, even allow the deposition of manuscripts on a preprint server while they are still under review. Even DOIs are issued for preprints by Crossref from late 2016 [21],

Funders also acknowledge the growing presence of preprint publishing in their policies: the Wellcome Trust allows researchers to cite preprints in their grant applications [23], and they are cooperating with an international group of research funders to explore the value and feasibility of establishing a Central Service for Preprints, which would set out to aggregate content from multiple sources and provide new ways for researchers and machines to search, access and reuse the content of preprint servers [24].

In order to facilitate a wide scientific discussion about preprint materials, a variety of forums and platforms were created which channel communication related to the uploaded materials. This way this body of literature gets more accepted and used as evaluated scientific content. The repository-based dissemination and review forums can take a variety of forms. There are platforms, such as PaperRater (www.paperrater.com) or SciRate (scirate.com), which are repository specific discussion forums, allowing for commenting on preprints in arXiv. ScienceOpen provides peer review to arXiv content by building collections where an editor or group of editors can group together articles that they find interesting, and open up all articles to post-publication peer review decoupling peer review and the communication of research from the formal publishing process [25]. Preprint servers facilitate communication on research results on a wider scale than traditional channels of dissemination and evaluation allowed for. Some platforms, like, biorXiv or PeerJ Pre-prints have a built-in commenting or peer review function on the platform. Others allow for crowd-sourced discussion on preprints in a specific field of study (Haldanes Sieve (haldanessieve.org)), or function as a multidisciplinary repository for articles and preprints (Self-Journals of Science (www.sjscience.org)). In addition, the overlay journal format allows managing preprints as journal content (episciences (www.episciences.org)), and there is a forum dedicated entirely to reviews on preprints (Academic Karma (academickarma.org)).

Repositories can also offer peer review functionalities. By turning repositories into evaluation platforms, the quality control aspect of the scholarly communication process is given back to the research communities. The open-source review plug-in, the Open Peer Review Module for repositories, developed by Open Scholar in association with OpenAIRE, adds overlay peer review functionalities to repositories using the DSpace software. OPRM on an institutional or other open access repository will enable the formal review of any digital repository content, including data, software code and monographs, by an unlimited number of peers. The review process is open and transparent, thus the full text of the reviews is available and the identity of the reviewers is disclosed. The system allows all interested peers to submit a review after creating a reviewer account and providing credentials certifying their qualification as peers. In addition to reviewing research objects, reviewers are also asked to evaluate previous reviews of each object they review. The OPRM includes a reviewer reputation system based on the assessment of reviews themselves where the reputation of the reviewer weighs on the importance of each review on the overall assessment of a research work. The primary objective of this system is to create reliable reputation metrics for research works, authors, reviews and reviewers. OPRM builds on the existing infrastructure offered by open access repositories. Besides providing novel metrics for the quantitative assessment of research quality, it promotes the use of relevant content that has been validated by reviewers using tags and advanced search filters. It advances an open and transparent dialogue about reliable and reviewed research material [26].

Preprint platforms typically do not employ much editorial functions beyond a check by moderators if content fits thematically and is scientifically sound. Additional value is added by overlay services which enable the management of a pool of reviewers. However, they all advocate open dissemination and enable open peer review (while not necessarily on the same platform): open identity of the reviewers, open report/commentary, and open participation from all research communities and public readers, as well.

4.4. Commenting Applications and Tools

Commenting applications and tools are not identified as peer review methods per se, however, they aim to provide complementary assessment of scientific content. They function as an application providing a layer of customized features on top of repository or journal content (PaperHive (paperhive.org)), or on materials disseminated through academic social networks (Research Gate OPR (www.researchgate.net/publicliterature. OpenReviewInfo.html)). These tools contribute to the network-based and collaborative aspect of research by opening up the discussion on published scientific results. In this way, they can be viewed as (light-weight) post-publication review tools.

Some tools allow sentence-level critique (Hypothes.is (hypothes.is), PaperHive) leading to contextual in-depth analysis of the content. Their operative features are based on annotation standards for digital documents (W3C Web Annotation standards): a new area of developments in digital content management [27]. These tools and platforms prepare for the next generation of read-write web application (Hypothes.is). TrueReview (www.theme-junkie.com) is an open-source tool with the motivation to provide reviews and evaluations. It organizes papers in venues, allowing different scientific communities to set their own submission and review policies. This tool offers benefits to the reviewers by ranking that can be prominently displayed alongside papers in the various disciplines, and provides reward to the authors of the most significant papers, both via an explicit paper ranking, and via increased visibility in search [28].

Table 1. Alternative review tools and services described on the basis of the attributes of Open Peer Review (defined by OpenAIRE)

	Platform	**Identity:** *reviewer's identity is published*	**Report:** *reviews and comments are published (alongside the relevant article)*	**Participation:** *by invitation and/or open to wider community to able to contribute to the review process*	**Interaction:** *specified discussion between authors and reviewers, and/or public, open interaction is allowed*	**Time:** *Open pre-review manuscripts/pre-publication review/ post-publication review or commenting*
F1000 Research	publishing platform	open	open	invited reviewers and open for commenting after registration	open	post-publication
The Winnower	publishing platform	open	open	invited reviewers and open commenting	open	open pre-review manuscripts
Science Open	publishing platform	open	open	reviewer: ORCID with 5 publications, comment: ORCID with 1 publication	open	post-publication
Frontiers	OA publisher	open	closed	invited reviewers	discussion of authors and reviewers	pre-publication
Copernicus Publications	OA publisher	opt in/out to sign	open	invited reviewers and research community commenting	closed	pre-publication
PeerJ	OA journal	opt in/out to sign	authors opt in/out to publish	invited reviewers	closed	pre-publication
eLife	OA journal	opt in/out to sign	authors opt in/out to publish decision letter	invited reviewers	discussion of editors and reviewers	pre-publication
Peerage of Science	standalone peer review platform	opt in/out to sign	opt in/out to publish review	registered, invited peers	closed	pre-publication
Publons	standalone peer review platform	opt in/out to sign	opt in/out to publish review	open	open	any point in the publication process
Rubrique	standalone peer review platform	double blind review	closed	closed	closed	pre-publication
SciOR	standalone review platform	open	open	open to registered authors and editors	discussions of authors and reviewers	pre-publication

PubPeer	standalone peer review platform/journal club	opt in/out to sign	open	open	open	post-publication
SciRate	scitation and commenting tool on arxiv content	open	open	open to registered users	open	open preprint manuscript
Self-Journals of Science	repository and evaluation system	open	open	open to authenticated scholars	open	open preprint manuscript
episciences	overlay journal platform to preprint servers	opt in/out to sign	closed	closed	discussion of author and copy editors	open preprint manuscript
Academic Karma	online peer review network	opt in/out to sign	opt in/out to publish	ORCID ID is needed to review	open	open preprint manuscript
Haldane`s Sieve	preprint commentary	open	open	open	open	open preprint manuscript
Hypothes.is	annotation /commentary tool	open	open	open	open	any point in the publication process
Research Gate OPR	review tool	open	open	open	open	post-publication
PaperHIVE	interactive platform	open	open	open	open	post-publication

Open pre-review manuscripts: Manuscripts are made immediately available in advance of any formal peer review procedures.
Pre-publication: review takes place before the publication of the final version of the manuscript is published.
Post-publication review: Review or commenting on publicly available version of the manuscript (revisions are allowed) or on published final-version manuscript.

The alternative review tools and services as discussed above offer various methods for review, such as open review, pre-publication or post-publication review, collaborative or decoupled review, and different degrees of openness in identity, participation and interaction among stakeholders. They might differ in their solutions, but they all carry several common features:

(1) they move away from the established publishing and review system by finding solutions to the problematic aspects of the traditional single/double blind review process (lack of transparency, potential bias, quality of review, etc.),
(2) the review process becomes more transparent either by opening up certain aspect of the process, or by providing detailed review policies,
(3) they urge a more conscious, collaborative participation from stakeholders either through invitation and dialogue within small circles between authors, editors and reviewers, or through crowdsourcing the process and allowing the public to add comments and reviews

These tools and services described on the basis of the seven attributes of Open Peer Review (defined in Table 1 by OpenAIRE), identify the main issues where intervention is needed in the traditional review system. The solutions they offer invigorate conversation among researchers about the functionalities of the review, their role and responsibilities in the process. Such dialogue, which is continuously reshaped by the exchange of ideas, new perspectives (Open Science approach) and tools (ORCID review tracking functionality) and emerging frameworks (pre-registration of research, uploading preprints for grant application), promises a more scholar-centric approach.

5. Conclusion

Several of the severe quality and transparency issues of scholarly communication can prospectively solved by a more active participation of researchers, reviewers and editors in the discussion on the opportunities and barriers, as well as the lessons-learned from existing and emerging services and experiments. As formal and informal knowledge sharing forums gain increasing significance within academic communities and their research activities, it is important to examine and discuss these alternatives to move towards a more structured and moderated dialogue about the underlying issues of research dissemination and evaluation.

Projects, such as OpenUP, set out to facilitate this discourse on scientific dissemination, peer review and assessment by mapping out the current scholarly environment and unraveling the underlying processes. The alternative peer review methods as discussed above contribute to a more democratic, transparent and community-based knowledge discovery and dissemination. However, as with many emerging new options of communication further evaluation of user roles and perceptions in open peer review settings is essential, including their change over time (based on good and bad experiences).

Recent surveys (OpenAIRE, 2016 [29], OpenUP, 2017) investigate some of these perceptions. In particular, researchers seem to be reluctant to fully embrace openness in the review process, but definitely see advantages of a transparent, collaborative review process. In order to make researchers less vulnerable to share their work and make their research open for comments, these alternative tools and services would benefit from further standardization and integration into the research cycle [30]. However, the formal acknowledgement of the viability and validity of these alternatives, such as independent review services, or review solutions for repositories and preprint servers, presupposes discussion on their sustainability, long-term availability, and their uptake by the researchers. Furthermore, as soon as review reports become a separate publication type and the reviewer's work is acknowledged as an academic activity on its own right, issues of authorship and copyright need to be taken into account (e.g. COPE) [31].

These and other practical challenges need to be considered when stepping up awareness and education efforts. In collaboration with other initiatives, OpenUP will set up a range of experiments which offer opportunities for further exploration and evaluation. Through the engagement with research communities, best practices, the most fitting methodologies and settings can be identified in different research areas (arts & humanities, social sciences, energy, life sciences). The concrete results from these experiments and pilot studies will provide insights into transforming research practices as well as challenges that need further investigation. The goal is to broaden the discourse and ultimately accelerate the uptake of Open Science solutions in the scholarly communication practices across all research disciplines.

References

[1] Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences, 109(42), 17028–17033. doi:10.1073/pnas.1212247109

[2] OpenUP. GARRI-4-2015 - Innovative approach to release and disseminate research results and measure their impact. Accessed on 18 November 2016 at: http://cordis.europa.eu/programme/rcn/665153_en.html

[3] Odell, J. D., Pollock, C. M. J. (2016). Open Peer Review for Digital Humanities Projects: A Modest Proposal. Accessed on 2 December 2016: https://goo.gl/JLPtlP

[4] Chowdhry, A. (2015). Gatekeepers of the academic world: a recipe for good peer review. Adv Med Educ Pract 6: 329–330. doi: 10.2147/AMEP.S83887

[5] Dineen, M. (2012). Time to rethink peer review: Evaluating scholarly work in the Internet age. University Affairs / Affairs Universitaires (UA/AU). Accessed on 2 December 2016: http://www.universityaffairs.ca/opinion/in-my-opinion/time-to-rethink-peer-review/

[6] Tattersall, A. (2015). For what it`s worth – the open peer review landscape. Online Information Review 39 (5): 649-663.

[7] Ross-Hellauer, T. (2016a). Defining Open Peer Review: Part One – Competing Definitions. OpenAIREblog. Accessed on 16 November 2016: https://blogs.openaire.eu/?p=1371

[8] Ross-Hellauer, T. (2016b). Defining Open Peer Review: Part Two – Seven Traits of OPR. OpenAIRE

[9] blog. Accessed on 16 November 2016: https://blogs.openaire.eu/?p=1410

[10] Callaway, E. (2016). Open peer review finds more takers. Nature, 539(7629), 343–343. doi:10.1038/nature.2016.20969

[11] Harnad, S. (2000) The Invisible Hand of Peer Review, Exploit Interactive, issue 5, April 2000. Accessed on 2 December 2016: http://cogprints.org/1646/1/nature2.html

[12] Pöschl, U. (2012). Multi-Stage Open Peer Review: Scientific Evaluation Integrating the Strengths of Traditional Peer Review with the Virtues of Transparency and Self-Regulation. Frontiers in Computational Neuroscience, 6. doi:10.3389/fncom.2012.00033

[13] Hunter, J. (2012). Post-Publication Peer Review: Opening Up Scientific Conversation. Frontiers in Computational Neuroscience, 6. doi:10.3389/fncom.2012.00063

[14] Priem, J., & Hemminger, B. M. (2012). Decoupling the scholarly journal. Frontiers in Computational Neuroscience, 6. doi:10.3389/fncom.2012.00019

[15] Björk, B.-C. (2015). Have the "mega-journals" reached the limits to growth? PeerJ, 3, e981. doi:10.7717/peerj.981

[16] Deenadayalan, Y., Grimmer-Somers, K., Prior, M., & Kumar, S. (2008). How to run an effective journal club: a systematic review. Journal of Evaluation in Clinical Practice, 14(5), 898–911. doi:10.1111/j.1365-2753.2008.01050.x

[17] Harnad, S. (1979). Creative disagreement. The Sciences 19: 18 - 20.

[18] Smith, R. (1999) Opening up BMJ peer review. A beginning that should lead to complete transparency. British Medical Journal, 318, 4-5.

[19] Frontiers. Collaborative Peer Review. Accessed on 06.03.2017: http://home.frontiersin.org/about/review-system

[20] Pöschl, U. (2012). Multi-Stage Open Peer Review: Scientific Evaluation Integrating the Strengths of Traditional Peer Review with the Virtues of Transparency and Self-Regulation. Frontiers in Computational Neuroscience, 6. doi:10.3389/fncom.2012.00033

[21] Akst, J. (2015). PubPeer Founders Revealed. The Scientist. Accessed on March 20, 2017: http://www.the-scientist.com/?articles.view/articleNo/43877/title/PubPeer-Founders-Revealed/

[22] PRE White Paper. (2016). Peer Review: Transparency, Standards and Training - Key Findings from Peer Review Evaluation (PRE) 2016 Survey

[23] McGlashan, D., Hadley C. (2017a). The rising tide of preprint servers. Inlexio blog. Accessed on 06.03.2017: https://www.inlexio.com/rising-tide-preprint-servers

[24] McGlashan, D., Hadley C. (2017b). Preprint servers: challenges and consequences. Inlexio blog. Accessed on 06.03.2017: https://www.inlexio.com/preprint-servers-challenges-consequences/

[25] Bourne, P. E., Polka, J. K., Vale, R. D., & Kiley, R. (2016). Ten simple rules for considering preprints. PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.2669v1
[26] Wellcome Trust. (2017). Preprints: we're supporting calls for a Central Service. Accessed on 10 March 2017: https://wellcome.ac.uk/news/preprints-were-supporting-calls-central-service
[27] Tennant, J. (2017). What if you could peer review the arXiv? ScienceOpen.com blog. Accessed on March 20 2017: http://blog.scienceopen.com/2016/04/what-if-you-could-peer-review-the-arxiv/
[28] Open Scholar. (2016). Open access repositories start to offer overlay peer review services. Open Scholar. CIC. Accessed on March 20, 2017: http://www.openscholar.org.uk/institutional-repositories-start-to-offer-peer-review-services/
[29] W3C. (2017). Web Annotation Data Model. Accessed on March 28, 2017: https://www.w3.org/TR/annotation-model/
[30] de Alfaro, L. (2016). TrueReview: A Platform for Post-Publication Peer Review. arXiv:1608.07878v1
[31] Ross-Hellauer, A., Schmidt, B., Deppe, A. (2016c). Novel Models for Open Peer Review. OpenAIRE deliverable (not published).
[32] Tattersall, A. (2016). Open Peer Review. In: Tattersall, A. (ed.) Altmetrics: A Practical Guide for Librarians, Researchers and Academics. Accessed on 2 December 2016: http://eprints.whiterose.ac.uk/100874/
[33] COPE Ethical Guidelines for Peer Reviewers 2013. Available at http://publicationethics.org/files/Peer%20review%20guidelines_0.pdf

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-75

Imparting Knowledge in Humanities. About Some Practices of Scientific Blogging on *Hypothèses*

Ingrid MAYEUR[1]
Université de Liège

Abstract. My contribution aims to explain how a platform of electronic publishing such as *OpenEdition* might reach a larger public outside the academic sphere, and contribute to the revealing of research in the Humanities to civil society. Practices of academic blogging have expanded since the 2000's. Its potential as a vulgarization tool was acknowledged early. However, most initiatives in scientific blogging are concerned with the field of STM (*Science, Technology, and Medicine*). A quick overview of the classical literature about vulgarization shows that a great deal of attention is being paid to STM. As such, I propose to examine what the practices of communication are towards civil society in Humanities through *Hypothèses*, the blogging platform of *OpenEdition*.

Keywords. citizen science, scientific blogging, humanities, OpenEdition, vulgarization

1. Introduction

It is commonplace to say that the digital environment has deeply transformed scientific communication, which for a long time has been diffused through conferences, seminars, monographs and journals. In regards to the written field, forms of books and periodicals have been transformed, first by digitization and then by their digital native feature that allows new accessibility, new dimensions, or even interactivities and connections. But the technical possibility of an unlimited diffusion can be restrained by an editor's authorizations. The Open Access movement has then intervened with another strategy that is promoting a broader diffusion of knowledge, by making available and appropriable scientific results that are still for the most part financed by public funds [1–3]. Open Access can be considered as an opportunity to systematize *direct scientific communication* [4], which is not submitted to a peer-reviewed process and takes place in specific infrastructures such as archival depots, platforms, personal websites, etc.[2] Scientific blogging belongs to this category, and many researchers have

[1] Université de Liège, Dpt., *Sciences du Langage et Rhétorique* (Bât. A2). Place Cockerill 3-5, 4000 Liège, Belgium; E-mail: Ingrid.Mayeur@ulg.ac.be.

[2] According to Martin Eve, Open Access can be defined as "the removal of price and permission barriers to scholarly research. Open access means *peer-reviewed* academic research work that is free to read online and that anybody may redistribute and reuse, with some restrictions." [2] (p. 1). In this sense, blogging cannot really be considered as belonging to Open Access. But in Peter Suber's view, blogging is well included in this context [1] (p. 65). I have choose to follow him on this point, due to the fact that Open

invested in it to communicate about their research in progress to the scientific community – but also to a broader public, by popularizing science.

What is popularizing science (particularly in regards to the Humanities)? How can a scientific blog help such a project? My contribution aims to explain how a platform of scientific blogging such as *Hypothèses* might reach a broader public than the academic sphere and contribute to the diffusion of Humanities research to civil society.

2. Short State of the Art and Problem

2.1. Popularizing Knowledge

In a synthetic note, Jacobi, Schiele and Cyr offer to talk not about vulgarization (in the singular) but about vulgarizations (in the plural), defined as "the attempts at the socio-dissemination of science outside the educational framework" [5] (p. 82, *my translation*). This term should encompass others like *popularization*, *divulgation*, *scientific public communication*, etc. that can also be used on this matter. Such a definition is established from the basis of the Anglo-Saxon concept of *informal education* [6] that underlines the role of vulgarization in general scientific culture outside educational institutions.

Indeed, practices of the diffusion of scientific knowledge through civil society are many and vary over time. The project of imparting science to a non-technical public is linked to the formation of the modern public sphere and the ideal of an individual emancipation carried out by the Enlightenment [7]. The first journals that circulated in the literate Salons during the classical Century, such as *Le Journal des Sçavans* or *The Philosophical Transaction* played a big role from this point of view [8]. According to Bensaude-Vincent [9], France in the 19th Century experienced, on one hand, a conception of *popular science*, which means an alternative science for a public of skilled amateurs that can be illustrated by Camille Flammarion's *Astronomie populaire*; and on the other hand, a conception of scientific vulgarization, which means that the scientist will "translate" science for an ignorant audience.

This last conception persisted into the 20th Century, even if the bad connotation attached to the term had decreased. Journalists are then called upon as mediators to build bridges between the scientific world and the non-scientific world, conceived as separated fields. In the 70's, the approach of vulgarization was questioned in a sociological perspective [10,11] and its relevance criticized by Roqueplo as a self-legitimization practice from the socio-intellectual elite[3]. Sharing knowledge would be an illusion, it is just like a show and the gap between scientists and non-scientists is impassable. Daniel Jacobi's works in the 80's [13,14] reintroduced vulgarization in the field of the scientific activities, through the idea that there is a continuum between specialist discourses and vulgarized discourses. According to him, the act of producing knowledge can't be distinguished from communicating it. Popularizing science is thus the researchers' assignment and not the journalists' charge. Thereafter, Fayard [15] notes the increasing demand from the audience for scientific information, and pleads to replace the old term of *vulgarization* with *public scientific communication*, a term

---

Access has given an impulse to open science in general, and has thus raised the need to take into account new avenues for communicating research.

[3] For this point as well as others, Yves Jeanneret's book *Écrire la science* [12] can also be referred to.

favoured by the emerging new medias. More recently, Stilgoe writes about the *citizen scientists*, “people who intertwine their work and their citizenship, doing science differently, working with different people, drawing new connections and helping to redefine what it means to be a scientist” [16] (p. 11). Imparting knowledge would thus contribute to democracy, by leading research in accordance with the common interest as well as by allowing citizens to understand issues that could be a matter of choice in their society. Moreover, it does not exclude contributions from citizens to build science (for instance with a call for crowdsourcing in research projects), and bring a dimension of feedback not existing before.

As we can see, imparting knowledge to a non-scientist public can be done with different purposes and depends on the intellectual and technical context. In summary, using Jacobi, Schiele and Cyr’s terms, we have passed from a paradigm of *rupture* with a scission between scientific and non-scientific spheres to a paradigm of *continuity* – and now even *action* or *interactivity* that can be added with *citizen science*. Therein Open Access should constitute a huge opportunity since scientific communication is inserted into the reticularity of the web and nourish what J.-C. Guédon named the “big scientific conversation” [17]. Scientific blogging spreads in such an environment, and completes other forms of diffusion.

2.2. Scientific Blogging

Publishing a blog requires a researcher to implement other means of scientific communication than traditional means. A blog post is characterized by its relative brevity, its link to the actuality (as we know, posts are usually displayed in reverse chronological order) and the possibility for the reader to engage in a dialogue by posting comments. Blogging implies the use of specific modes of writing due to the technical substrate of a digital text/discourse that enables to use of hypertextuality, intermediality, etc.[4]. The practice of academic blogging has expanded since the 2000’s [21,22]. There are several reasons for publishing an academic blog [23,24]. For a scientist, it could be the place to introduce himself and promote his work by increasing its visibility. Doing so, he positions himself in a professional network [25] and maximizes his chances of being read. An academic blog is also an interface that can show research in progress, record scientist’s actualities, or diffuse new from a disciplinary field. Finally, writing on a blog can help a scientist elaborate a reasoning and express it clearly; in this sense, this is a formative exercise that supports the process of reflection.

However, the value of such scientific texts that are not peer reviewed can be questioned. More specifically, what are their contribution to science? The arguments that are most given are that they provide publicity for research in progress and enable interactivity with readers. From this point of view, it reintroduces the dialogism of oral scientific communication like conferences or seminars. Such an idea has been developed by Melissa Gregg [26] for whom blogging is a “conversational scholarship”. Interacting with peers in a community gives the opportunity to improve research results, or even interact with a non-technical audience in a bottom-up dynamic [22]. This conversation thus exceeds the academic sphere, and the potential of blogs as a tool for

[4] Such issues have been treated in the field of discourse analysis by Marie-Anne Paveau [18,19]. About hypertext as a *figure of reading*, we can also mentioned Alexandra Saemmer’s book *Rhétorique du texte numérique* [20].

popularization [27,28] or for scientific journalism was acknowledged early. While the expected audience for a scientific book or journal is a research community, blogs might be consulted by a broader public because of the Open Access element. As a result subjects and lexicons can be adjusted in this particular context.

2.3. What about Humanities?

As Marin Dacos (who is *OpenEdition*'s director) has underlined [29], there remains a research gap about blogging in SSH (Social Science and Humanities) and its infrastructures while platforms of blogging in STM (Science, Technique, Medicine) like *ScienceBlog, ResearchBlogging* or, for popularization, *C@fé des sciences* are well known. In the English-speaking field, we can mention the platform Hastac, which is interdisciplinary and mixes "hard" and "soft" science. The development of *Hypothèses* (2010) on the platform *OpenEdition*, after *Revues.org* (1999, dedicated to journals) and *Calenda* (2000, which is a calendar for scientific activities) is an attempt to fill this gap. *Hypothèses* began in French, and has since become multilingual (blogs can be read in German, English, Spanish or even Portuguese).

Such a research gap can similarly be observed in studies about the popularization of science. References that have just been quoted speak a lot about popularizing or imparting knowledge in the physical, medical or technical sciences, but not so much about Humanities (which Stilgoe does not even mention in *Citizen Science*). This is quite surprising, since Humanities are funded to question and explain the human environment and culture; in this way, they are basically linked to the concerns of citizens[5]. Through an overview of the practices that has been developed in the "non-specialists blogs" on *Hypothèses,* I wish to ask, (i) what paradigms of imparting knowledge have been used and, (ii) what means are used to impart knowledge towards a non-specialist audience. I should indicate that in the framework of a short paper, my purpose is to present an overview and not an accurate analysis of these blogs[6].

3. About the Corpus from Hypothèses

To establish my corpus, I have taken as a starting point the blogs' catalogue on *Hypothèses,* activating the "non-specialist blog" filter[7]. Through research led in March 2017, I have obtained 57 results. I then reduced these results to the disciplines in Humanities, excluding psychology, economics, politics, sociology, and also geography that are sometimes include in[8]. At the end, I removed the blogs which had not been

---

[5] The observation of a gap concerning Humanities in citizen science was also made by Dobreva & Azzopardi [30] – but in this case, citizen science was conceived as an active contribution from citizen to science, and focused on the uses of crowdsourcing in Humanities research. I will focus more on the scientists' practices.

[6] I cannot make such an analysis anyway due to my very poor knowledge of German, Spanish and Portuguese.

[7] Such categorization is suggested by the blog's editor when he submits his project to the editorial team, who can then review it. This process can be discussed, since some other blogs of the platform are undeniably non-specialist blogs but not categorized as such (ex. *Mondes sociaux*). On the other hand some blogs of this category are maybe not really accessible to a non-specialist audience. However, it offers a working basis in the framework of this short paper.

[8] The Humanities can be described in a nutshell as "the study of how people process and document the human experience", cf. Liu [31].

updated since 2013 as well as those where Humanities were much too marginal. In doing so, I obtained a final corpus of 37 blogs. Most are in French (18) or in German (12). Some are in Spanish (4), English (2) or even Portuguese (1)[9] [see Table 1].

Table 1. Corpus Tab

Blog	**URL**	**Nr of visits in 2016**[10]
The Recipes Project: Food, Magic, Science, and Medicine	https://recipes.hypotheses.org/	525 452
El vellocino de oro, Blog de contenido sobre cultura griega y romana.	https://vellocinodeoro.hypotheses.org/	167 340
1914-1918: Ein rheinisches Tagebuch, Quellen aus Archiven des Rheinlands	http://archivewk1.hypotheses.org/	135 056
Historisch denken | Geschichte machen: Anmerkungen zu Geschichte | Vergangenheit | Geschichtsunterricht | Geschichtsdidaktik Blog von Christoph Pallaske | @pallaske | @segu_geschichte	https://historischdenken.hypotheses.org/	122 705
Weber World Café: Transregionale Dialoge zwischen Wissenschaft, Gesellschaft und Kultur	http://wwc.hypotheses.org/	64 917
Geisteswissenschaft im Dialog	https://gid.hypotheses.org/	63 503
Dipnot: Note de bas de page en turc - Réflexions des chercheurs de l'IFEA	https://dipnot.hypotheses.org/	58 424
Krosworldia: Geschichte, Archäologie und die Welt der Geisteswissenschaften in Medien und Gesellschaft	https://kristinoswald.hypotheses.org/	54 433
Sottovoce. Espacio virtual de divulgación científica en español sobre la voz humana	http://sottovoce.hypotheses.org/	42 334
Actualité des études anciennes: Activités scientifiques autour de la Revue des Études Anciennes	http://reainfo.hypotheses.org/	40 671
Angles droit: actualités scientifiques du droit et de la science politique de l'Université de Bordeaux	https://anglesdroit.hypotheses.org/	33 202
Publier une correspondance: Méthode et contenu	https://puc.hypotheses.org/	32 193
En route to a shared identity [Zu einer gemeinsamen Identität: Quellen zur Geschichte Mitteleuropas im digitalen Zeitalter]	http://dighist.hypotheses.org/	29 631
Conserver, enseigner, chercher, Réflexions autour du patrimoine scientifique d'Aix-Marseille Université	http://tresoramu.hypotheses.org/	28 296
Philosophie - Phisolophie, Besser durch den Alltag mit Reflexion	https://philophiso.hypotheses.org/	28 181
The French Revolution Network, Revolutionary transitions from the eighteenth century to the present	http://revolution.hypotheses.org/	27 367
|Marginalien: Religionswissenschaftliche Randbemerkungen	http://marginalie.hypotheses.org/	26 605
E-pigraphia, Epigrafía en Internet	https://epigraphia.hypotheses.org/	26 143
Bling: Blog de linguistiques illustré	http://bling.hypotheses.org/	23 665
Mes langues aux chats: Limpide linguistique et analyse de discours 3.0	https://lac.hypotheses.org/	22 951
Dada, Merz and Co. Historische Avantgarde im	http://merzdadaco.hypotheses.org/	21 685

[9] It is in fact the main language of the blog, since some blogs can publish posts in another language (for instance *Dipnot* is also published in Turkish, Weber World Café displays posts both in German and English, etc.).

[10] According to the public statistics available on http://logs.openedition.org/

Spiegel der Gegenwart		
Histoires lyonnaises: Des érudits vous racontent leur histoire de Lyon !	http://lyonnais.hypotheses.org/	20 226
Erinnern in Speyer 1933-1945: Das Blog stellt eine Dokumentation über Zwangsarbeiter in Speyer und Umgebung in den Jahren 1933 bis 1945 dar.Erinnern in Speyer 1933-1945	https://speyermemo.hypotheses.org/	20 028
Ecdotique	http://ecdotique.hypotheses.org/	19 877
La bitácora de Carriazo: Una bitácora sobre las actividades académicas del profesor José-Ramón Carriazo	http://carriazo.hypotheses.org/	19 510
Archivum Rhenanum: Archives numérisées du Rhin supérieur	https://archives-fr.hypotheses.org/	14 449
Onomatique: Traitement automatique et onomastique	https://onomastique.hypotheses.org/	11 047
Les langues à l'EHESS: Langues étrangères appliquées aux sciences sociales : allemand, anglais, français langue étrangère (FLE)	http://bdl.hypotheses.org/	8016
Janvier / novembre 2015. Réfléchir après…	http://apres2015.hypotheses.org/	6 678
O "case" Santa Rita (Rio de Janeiro): Arqueologia da transição | Arquitetura da persuasão | Antropologia da devoção	http://santarita.hypotheses.org/	6 316
Call me, Sprachen lernen mit digitalen Medien	http://callme.hypotheses.org/	5 783
NDig, Neuburg und der große Krieg. Ein Pilotprojekt zur Stadtgeschichte im 19. und 20. Jahrhundert	http://neuburg.hypotheses.org/	5 624
Gedenkstättenpädagogik-Blog	http://gedenkpaed.hypotheses.org/	5 183
Plongée avec Pline l'Ancien: Carnet de recherche sur la biodiversité aquatique décrite au livre IX de l'Histoire Naturelle de Pline L'Ancien	http://hstpline.hypotheses.org/	3 835
Mémoires méditerranéennes: Association pour la sauvegarde et la valorisation du patrimoine documentaire méditerranéen	http://memoirmed.hypotheses.org/	2 712
Anchora: Médiation culturelle des textes scientifiques latins et grecs de l'Académie de marine de Brest (culture matérielle et mer)	http://anchora.hypotheses.org/	273
Les archives et le web: Blog de recherche d'un étudiant de master	http://archiveweb.hypotheses.org/	245

3.1. What Ways for Imparting Knowledge in the Humanities?

In the corpus, three ways to impart knowledge to a non-specialist audience can be identified:

- **Applied Humanities** i.e. using knowledge in Humanities to explain a given subject in the realm of SSH, creating a clearer understanding of some phenomena. For instance, explaining the sense or the uses of French expressions (*Bling* or *Mes langues au chat*), or the function of the vocal apparatus (*Sottovoce*); giving a historical perspective on revolutions (*The French Revolution Network*); introducing methods and results of names studies (*Onomatic*); providing a reading of social or politic actualities through legal systems (Angles droits), etc.

- **Cultural mediation** i.e. enhancing and promoting a historical/cultural/ patrimonial issue that could be avant-garde (*Dada, Merz and Co.*), monuments (*O "case" Santa Rita (Rio de Janeiro)*), cities (Lyon, Neuburg), ancients texts (*Anchora*), cultures (*El vellocino de oro, Actualité des études anciennes*); letters (*Publier une correspondance*), or artifacts (*Conserver, enseigner,* chercher) etc.. This can also take the form of a virtual exhibit: *1914-1918: Ein rheinisches Tagebuch*, *Erinnern in Speyer 1933-1945*.
- **Publishing activities** i.e. making visible some activities of an academic team or institution, by promoting an academic program (*Les langues à l'EHESS*); by publishing proceedings as well as videos or texts of lectures (*Geisteswissenschaft im Dialog*, *En Route for a Shared Identity, Weber World Café*). It can also occur through publishing actualities of a disciplinary field, a research group, or a project by scientific watching (*E-pigrafia*, *Dipnot, Archivum Rhenanum*).

These applications could be non-exclusive. As an exemple, the blog *Janvier-décembre 2015. Réfléchir après…* is concerned with the Paris attacks of 2015 and its implications. It wishes to clarify "the sense and the geopolitical, political and social factors of these events"[11] (*my translation*). In this sense, it contributes to the treatment, with methods in SSH (such as history, philology, medias studies etc.), of phenomena – such as how religion may or may not play a role in terrorism, what the Rule of law becomes under terrorism, how the media could build an interior enemy, etc. But this is also a place to publish conference texts or make a scientific watch linked to this subject, because this blog is linked to a conference cycle. Texts are thus adequately published for non-specialists.

3.2. Expected Audiences

What does "non-specialist" mean or, in other terms, what is the expected audience for these blogs? It is not an easy issue, since this is, for the most part, not specified on the blog. On the other hand, some of these blogs can obviously not be consulted by the layman, despite their categorization as a "non-specialist blog". However, in certain cases the expected audience is clearly specified. Some blogs are directed at students, like *El vellocino de oro* (which is very popular). It can also be professionals or scholarly (even if other audiences are not excluded): for instance, *Ecdotique* or *E-Pigraphia* are concerned with publishing ancient sources; or *The French Revolution Network*, that plans to gather scholars from all disciplines. Without surprise, some other blogs claim their accessibility to the general audience, with a specific interest on a given subject, for instance, religion in *Marginalien*, or recipes of all kinds for *Recipes projets* – as the short title indicates "Food, Magic, Art, Science, and Medicine". This last blog has to be highlighted because it is the most consulted in the corpus. It federates a large community by investigating a folk topic that is transversal since it exists in every cultural area. Moreover, such an issue is treated in an interdisciplinary perspective that can interest a lot of people. The blog also has a significant presence on social media (Facebook, Twitter) that could favor its success.

---

[11] « À propos », *Janvier-décembre 2015* [cited 27 March 2017]. Available from: https://apres2015.hypotheses.org/a-propos.

4. Conclusion

I will now link the overview of my corpus with the question raised earlier on what paradigms of vulgarization are used within this corpus. It is clear that *Hypothèses* takes place in the paradigm of *continuity* due to the fact that researchers communicate their research themselves. Popularization is done by scientists and not by a third party (journalist, mediator, etc.). Moreover, popularization is brought together with other scientific practices (eg. *Hypothèses* is a platform to be read by peers as well as a broader audience. Research blogs and non-specialist blogs coexist on the same platform). OpenScience is taken as an opportunity to make scientific communication public. However, there remains a subject I have no time to address in this short paper, which is the big difference that might exist between providing access to scientific knowledge, making it visible, and making it appropriable by non-scientists. I think some examples in the corpus could testify to some confusion about this subject.

Are these practices related to *citizen science*? In some ways, they are. One cannot say there is a real bottom up conversation, since in general very few comments are left after the posts. As an example, the very popular blog *1914-1918: Ein rheinisches Tagebuch* expects that the readers comment and, if they could, provide more information on the archival documents that are exhibited[12]. But, in practice, they don't – while the blog is one of the most visited in the corpus[13]. The impact should then be measured by other indicators such as browsing statistics, backlinks or shares on social media that may better capture what Marin Dacos named the *silent conversation* [32]. But citizen science is not just science made with the help of citizens; more importantly, it is science that serves the common interest. How can Open Science in Humanities serve the common interest? As Peter Suber said in his introduction to Martin Eve's *OpenAccess and the Humanities*, "In the case of the sciences, that can mean new medicines and useful technologies, and in the case of the humanities it can mean enriched education, politics, compassion, imagination and understanding." [2] [p. ix]. Some practices of imparting knowledge to non-specialists are directly related to this project of a better understanding of social matters in order to act as a cultivated citizen[14]. In that way, imparting knowledge in Humanities contributes to making the world intelligible and we may consider that the practices, such as those we have overviewed, fit fully into that framework.

Acknowledgements

I would like to thank the *CLEO* (*Centre pour l'Édition électronique ouverte*) team and in particular Marin Dacos, for exchanges and documents concerning *OpenEdition* and its activities.

---

[12] "Editorial", 1914-1918: Ein rheinisches Tagebuch, s.d.. Online: http://archivewk1.hypotheses.org/uber.

[13] This is also the case with a blog like Geisteswissenschaft im Dialog, which attempts to open a discussion following the posts (some rules of good behaviors are even specified in the section "Über das Blog ").

[14] This is also the sense of Marin Dacos' speech "Le savoir est une arme" (2016) given at the awarding of his Medal of Innovation (CNRS).

References[15]

[1] Suber P. Qu'est-ce que l'accès ouvert ? [OpenAccess]. Marseille: OpenEdition Press; 2016 [2012] Available from: http://books.openedition.org/oep/1600.

[2] Eve MP. Open Access and the Humanities: Contexts, Controversies and the Future. Cambridge University Press; 2014. Available from: http://ebooks.cambridge.org/ebook.jsf?bid=CBO9781316161012.

[3] Chartron G. Stratégie, politique et reformulation de l'open access. RFSIC. 2016 (8). Available from: https://rfsic.revues.org/1836.

[4] Beaudry G. La communication scientifique directe : un nouveau champ éditorial. Hermès 2010/2 (57): 51‑7.

[5] Jacobi D, Schiele B, Cyr M-F. Note de synthèse [La vulgarisation scientifique et l'éducation non formelle]. Rev Fr Pédagogie. 1990; 91(1): 81‑111.

[6] Lucas AM. Scientific Literacy and Informal Learning. Stud Sci Educ. 1983; 10(1): 1‑36.

[7] Bertrand G, Guyot A, editors. Des « passeurs » entre science, histoire et littérature : Contribution à l'étude de la construction des savoirs (1750-1840). Grenoble: UGA Éditions; 2017 [2011]. Available from: http://books.openedition.org/ugaeditions/566

[8] Beaudry G. La communication scientifique et le numérique. Paris: Lavoisier; 2011.

[9] Bensaude-Vincent B. Splendeur et décadence de la vulgarisation scientifique. Quest Commun. 30 juin 2010; (17): 19‑32.

[10] Roqueplo P. Le partage du savoir, science, culture, vulgarisation. Paris: Seuil; 1974.

[11] Boltanski L, Maldidier P. La vulgarisation scientifique et son public: enquête sur « Science et vie ». Paris: Éditions de l'École des hautes études en sciences sociales; 1977.

[12] Jeanneret Y. Écrire la science: Formes et enjeux de la vulgarisation. Paris: Presses Universitaires de France; 1994.

[13] Jacobi D. Auteurs et lecteurs de la recherche. Bull Bibl Fr BBF. 1 janv 1984;(6):484‑91.

[14] Jacobi D. Diffusion et vulgarisation: itinéraires du texte scientifique. Besançon: Presses Univ. Franche-Comté; 1986.

[15] Fayard P. La communication scientifique publique: de la vulgarisation à la médiatisation. Lyon: Chronique sociale; 1988.

[16] Stilgoe J. Citizen Scientists: reconnecting science with civil society. Demos London; 2009. Available from: http://creias.ipleiria.pt/files/2010/08/Citizen_Scientists_-_web.pdf.

[17] Guédon J-C. Chapitre 7. Le libre accès et la « Grande Conversation » scientifique. In: Vitali-Rosati M, Sinatra ME, éditeurs. Pratiques de l'édition numérique. Montréal: Presses de l'Université de Montréal; 2014 [cité 9 mars 2017]. p. 111‑26. (Parcours numérique). Available from: http://books.openedition.org/pum/324.

[18] Paveau M-A. Technodiscursivités natives sur Twitter. Une écologie du discours numérique. Épistémé Rev Int Sci Hum Soc Appliquées. 2013;9:139–176.

[19] Paveau M-A. Présentation. Les textes numériques sont-ils des textes? Itinér Litt Textes Cult. 1 févr 2015;(2014‑1). Available from: https://itineraires.revues.org/2312.

[20] Saemmer A. Rhétorique du texte numérique: figures de la lecture, anticipations de pratiques. Lyon: Presses de l'Enssib; 2015.

[21] Dacos M, Mounier P. Les carnets de recherche en ligne, espace d'une conversation scientifique décentrée. In: Jacob C, editor. Lieux de savoir. 2. Gestes et supports du travail savant. Paris: Albin Michel; 2010. p. N/A. Available from: http://archivesic.ccsd.cnrs.fr/sic_00439849/document.

[22] Blanchard A. Quand la culture scientifique s'affranchit sur le web : l'exemple des blogs de science en français (2003-2014). In: Poirrier P, éditeur. Histoire de la culture scientifique en France : institutions et acteurs. Dijon, France: Presses universitaires de Dijon; 2016. Available from: https://halshs.archives-ouvertes.fr/halshs-01242707.

[23] Bukvova H. Taking new routes: Blogs, Web sites, and Scientific Publishing. ScieCom Info. 2011;7(2). Available from: http://www.sciecom.org/ojs/index.php/sciecominfo/article/view/5148.

[24] Blanchard A. Science blogs in research and popularization of science: why, how and for whom? In: Cockell M, Billotte J, Darbellay F, Waldvogel F, editors. Common Knowledge: The Challenge of Transdisciplinarity. Lausanne: EPFL Press; 2011. p. 219–232. Available from: http://www.peerevaluation.org/read/libraryID:21374

[25] Bersenev A. Scientific blogging as a model for professional networking online. Cell Ther Transplant. 2010; 2(7). Available from: http://www.ctt-journal.com/2-7-en-bersenev-2010aug2.html

[15] All the links were checked on March 27th 2017.

[26] Gregg M. Feeling Ordinary: Blogging as Conversational Scholarship. Continuum. 1 juin 2006; 20(2): 147‑60.
[27] Ducet P. Les blogs de vulgarisation scientifique. Université Paris-X Nanterre; 2008. Available from: http://association.cafe-sciences.org/public/Memoire_blogs_scientifiques.pdf.
[28] Collectif. Comment écrire pour un carnet de recherche ? : Proposé par Mélodie Faury et Pierre Mounier. In: THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques. Paris: Éditions de la Maison des sciences de l'homme; 2012. Available from: http://books.openedition.org/editionsmsh/324.
[29] Dacos M. Un angle mort ? Les infrastructures pour les SHS en général et pour les carnets de recherche en particulier. Blogo-numericus. 2013 [cited 27 March 2017]. Available from: http://bn.hypotheses.org/11261.
[30] Dobreva M, Azzopardi D. Citizen science in the Humanities : A promise for creativity. In University of Cyprus, Nicosia; 2014. Available from: https://www.um.edu.mt/library/oar//handle/123456789/987
[31] Liu A. What Are the Humanities?. 4Humanities. 2014 [cited 27 March 2017]. Available from: http://4humanities.org/2014/12/what-are-the-humanities/.
[32] Dacos M. La conversation silencieuse. Blogo-numericus. 2009 [cited 27 March 2017]. Available from: https://bn.hypotheses.org/191

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-85

Grey Literature Publishing in Public Policy: Production and Management, Costs and Benefits

Amanda LAWRENCE
Swinburne University of Technology

Abstract. Public policy and practice, and policy research, relies on diverse forms and types of information and communication, both traditional publications and a myriad of other documents and resources including reports, briefings, legislation, discussion papers, submissions and evaluations and much more. This is sometimes referred to as 'grey literature', a collective term for the wide range of publications produced and published directly by organisations, either in print or digitally, outside of the commercial or scholarly publishing industry. In the digital era grey literature has proliferated, and has become a key tool in influencing public debate and in providing an evidence-base for public policy and practice. Despite its ubiquity and influence, grey literature's role is often overlooked as a publishing phenomenon, ignored both in scholarly research on media and communications and in the debate on the changing nature of open access and academic publishing. This paper looks at the production of grey literature for public policy and practice where the changes enabled by computers and the internet are causing a hidden revolution in the dissemination of knowledge and evidence. It explores the production, dissemination and management of publications by organizations, their nature, purpose and value, and investigates the benefits and the challenges of publishing outside of the commercial or scholarly publishing enterprises. The paper provides estimates of the economic value of grey literature based on online surveys and valuations and considers the costs and benefits of self-publishing by organisations which provides both a dynamic, flexible and responsive publishing system and one in which link rot, duplication and highly varying standards abound. The findings are part of a broader research project looking at role and value of grey literature for policy and practice including consumption, production and collection. It will be of interest to a wide range of policy makers and practitioners as well as academics working in media and communications, public administration and library and information management.

Keywords. grey literature, public policy, electronic publishing

1. Introduction

While digital technologies have radically increased our capacity to produce and disseminate knowledge and information, many of the social and economic benefits are being lost as researchers and policy makers struggle to filter search results, find relevant material and evaluate the huge variety of resources being published online by a wide range of organisations [1]. Given that a great deal of money and resources (much of it public funds) are spent creating knowledge and information to improve outcomes on public interest issues, governments and producers should seek to maximise its benefits for the community. One way in which evidence for policy is produced is by academic researchers publishing in books and journals. However an even larger and arguably more influential source of knowledge for policy and practice are the documents produced and disseminated by organisations, outside of the commercial or scholarly publishing industry, such as technical and research reports, working papers, policy documents, evaluations and briefings. These are sometimes referred to as grey literature, a term which often seems to obscure more than it illuminates but which is a useful collective noun for a wide variety of formats and publishing approaches which do share some fascinating common properties.

Grey literature may seem to be an issue for a by-gone age of small print runs, fugitive documents and specialist clearing houses set up to try to collect them, of interest only to historians and of little importance in the internet age. In reality the opposite is the case. Like other forms of communication, grey literature has moved online and this has had a huge impact on the way it is able to be produced, disseminated, discovered and used. Despite this, grey literature as a form of electronic publishing is a phenomenon that is often overlooked. As Thompson puts it about the publishing industry as a whole, "A revolution *has* taken place in publishing and *is* continuing to take place, but it is a *revolution in the process* rather than a *revolution in the product*" [2, emphasis in the original]. Quietly but steadily, the number of organisations, departments and centres producing and disseminating policy related grey literature has grown and perhaps without widespread awareness it has become central to public debate, informing and contributing to a wide range of topics and disciplines.

Prior to the advent of the internet, one of grey literature's defining characteristics was that it was costly to print and distribute and difficult to find and access, in contrast to journal articles which may have imposed a charge but were professionally managed and indexed and had stable subscriptions and distribution channels [3, 4]. The internet turned this on its head, allowing individuals and organisations to cheaply and easily send their reports and documents around the world to peers, partners and other organisations. Grey literature has been at the forefront of the digital publishing and open access revolution ever since, but it is often not perceived in this way and has been undervalued as a resource and under-appreciated as an influence on policy and practice decisions. Grey literature is coming into its own as a form of electronic publishing that is prolific, heavily used and highly valued, particularly for public policy but also for policy related scientific issues such as climate change, for example the IPCC reports [5]. Yet at the same time grey literature creates many problems, not only for users but also for the producers themselves and any collecting services, due to its lack of management and publishing standards and a serious under investment in long term management of public interest digital content – our collective public knowledge commons.

2. Grey Literature Research

The concept of grey literature as an object of academic research has mainly been of interest to the discipline of library and information science [6] and somewhat overlooked by sociology or media and communications researchers. Perhaps this is due to a general lack of focus on media's relationship to society by sociologists until the 1990s [7] or a lack of research on contemporary book and journal publishing industries [2, 8]. It is certainly the case that there is little use of the term grey literature in either field and it is not generally recognized as a type of media or communication.

Historically it is clear that publishing in all its manifestations is a form of media since the birth of the printing press, with occasional pamphlets not only being a precursor to more regular newspapers from the 17th century [9] but also leading to other document types that would now be considered grey literature such as discussion papers and reports. The approach taken here is to consider media in a broad sense following Couldry [7] who defines media as being much more than newspapers, radio, television and film, it is the "institutionalized structures, forms, formats and interfaces for disseminating symbolic content" [7].

On this basis it is argued that grey literature is not only of interest as a challenge for information professionals but that it operates as a type of media. It is symbolic content produced in a variety of forms and formats in institutionalised structures, however these are not the institutions that we are used to consider as media producers. It is because grey literature is, by definition, produced by organisations whose main purpose is *not* commercial or professional publishing, which leads to its distinctive interest as a form which requires focused and specialized analysis. As Lobato and Thomas point out, the way that media economies are organized and regulated has important social consequences: "Systems of communication shape our understanding of the world and help us define who we are, as individuals and as communities…Media economies – as systems that organize this communicative capacity – are gateways for power, politics, and pleasure… " [10].

At the same time the infrastructure that has enabled the explosion in grey literature production and that is required for its long term access and preservation is also part of what needs to be studied. The 'computational turn' in the social sciences seeks to examine the structuring aspects of the search engines, platforms and software that we all use when interacting online [11]. 'Infrastructure studies' or 'knowledge infrastructure' [12, 13] involves an examination of how they shape and define our world: "As knowledge infrastructures shape, generate and distribute knowledge, they do so differentially, often in ways that encode and reinforce existing interests and relations of power" [12].

There are some similarities that can be drawn between grey literature production to the discussions of user-generated content or what Castells calls "mass self-communication" – such as its speed, flexibility, targeted audience and open dissemination and circulation, and highly variable standards. Yet they are also significantly different in their institutional basis and the scale and sophistication of their outputs. While some include social media as a form of grey literature, this article is not concerned with social media specifically. Nor will it focus on the widespread production of technical reports in engineering, or field notes in archaeology or other major grey literature production sources. This paper is concerned primarily with public policy and practice oriented organisation-based publications, or grey literature, or what could be described, rather clumsily, as "public publishing", as in public broadcasting and public libraries – given that much of it is funded by governments and made public in the public interest or with the aim of participating in public debate.

3. Methodology

The impetus for this research has been my 11 year involvement with Analysis and Policy Observatory (APO – previously known as Australian Policy Online) a digital library and alert service established in 2002 by researchers at Swinburne University of

Technology. APO collects and catalogues policy and practice grey literature resources and as the service has developed over the years the need to better understand the proliferation of policy reports and papers and the diverse ecosystem of producing organisations led to an Australian Research Council funded project to investigate the production, use and collection of grey literature for policy and practice. This research has involved online surveys and interviews with producing organisations, research users and collection services and a survey of digital collections. This paper focuses on the findings for producing organisations. Other results from this research have already been published (see 1, 14, 15) and further publications on users and digital collections will be forthcoming.

To establish who is participating in grey literature production, how and why, we conducted an online survey of producing organisations. To complement and probe the survey results we also undertook semi-structured interviews with representatives from organisations engaged in production of research publications on public policy issues in Australia. The survey and interviews asked about the kinds of resources produced, how important they are, how many staff are involved in production, sources of income, and the reasons for direct publishing.

Producing organisations were asked to identify themselves as being in one of four sectors: Government: Federal, State, local government departments or agencies; Education: Universities and TAFES particularly university research centres and institutes; Non-government organisations (NGOs): associations, interest groups, think tanks, charities; Commercial or private businesses: large, medium or small companies; business groups; consultants and research companies, lobbyists. Respondents from all four groups were sought via direct emailing to some of the 3,500 organisations listed as sources on APO at the time of the survey as well as general online promotion on APO and other websites and newsletters including via Linkedin and Twitter. In total 155 organisations responded to the online questionnaire, with 144 from Australia (93%), 7 from New Zealand and 4 from other countries. As the vast majority were from Australia, only these responses have been included in the analysis for this paper (N=144).

Estimate of the population of producing organisations were needed to provide estimates of the scale and value of grey literature production. Calculating the grey literature producer population is difficult as it is likely to be a subset of organisations across all sectors of society. Based on various data sources such as the Australian Bureau of Statistics it is estimated that across government, education, civil society and industry there could be at least 30,000 organizations producing policy-oriented grey literature in Australia. This figure is a conservative estimate based on the following figures: at least 2000 departments and agencies across the Australian federal government, state and territories and local governments; 41,008 not-for-profit organisations [16]; 56,894 registered nonprofit institutions (NPIs) [17]; 250,000 businesses in Professional, Scientific and Technical Services [18] where at least 10% of these (25,000) may well be producing policy-oriented grey literature; 11,770 management consulting firms [19]; and an unknown number of large companies who may be producing grey literature relating to policy issues affecting their industry.

Survey respondents came from all four sectors with 38% from non-government organisations, 35% from education, 21% from government departments and agencies and 7% from commercial companies. About half of the organisations responding were small, having 10 staff or less, and a further 17% had up to 20 staff. About 10% had over 500 staff, these being mainly government organisations. About two thirds of organisations had 10 staff or less engaged in producing research and information publications, which given their overall size indicates considerable focus on knowledge production. Direct publishing is a key method used by organisations across all sectors of society to produce and disseminate new research, policy information or opinion. Amongst the surveyed organisations, the production of research and information occurs at a fairly steady pace with a third (38%) publishing on a weekly basis or more often, and two-thirds (62%) producing material quarterly or more often. Staff estimated that they spend around a third of their weekly work time creating grey literature each year.

4. Why Organisations Engage in the Research and Publishing

Despite the different drivers and incentives that exist across government, academia, civil society organisations and private companies, the survey showed overwhelming agreement about the motivation for organisations to produce their own publications. As Table 1 shows, the most important reasons for producing research and information are to provide an evidence base for, and inform public policy and practice (92%), translate knowledge for public use (84%), and maximise public access to research and information (79%). Those in education and the NGO sectors also rated sharing findings with peers, raising their organisation's profile and attracting media coverage highly, more so than those in government or commercial sectors.

Table 1. Reasons organisations produce their own publications (grey literature)

Producer N= about 109 Important/ Very important	**Gov %**	**Edu %**	**NGO %**	**Com %**	**All %**
Provide an evidence-base for policy or practice	90	92	95	78	92
To inform public policy or practice	95	92	96	63	92
Knowledge translation, ie making research findings clearer for public use	75	95	82	75	84
Maximise public access to research and info	80	81	84	38	79
Share findings with peers	61	78	77	25	71
Raise organisation or staff profile or position	47	72	78	50	69
Media coverage and public debate of an issue	58	69	80	25	68
Advocacy or lobbying tool	56	36	87	50	62
Meet organisation or funder requirements	56	71	55	33	59
Internal purposes or analysis	72	37	61	75	56
Flexibility i.e. of formats, content etc.	33	44	71	38	53
Control the timing of publication	61	46	61	14	53
Other (N=31)	33	42	38	40	39
Comply with regulations	56	9	17	25	22
Sales and other financial benefits	6	6	14	25	11

NGO's strongly value using research as an advocacy and lobbying tool (87%) and appreciate the flexibility of grey literature formats (71%). NGOs and governments most valued being able to control the timing of a publication (61%), significantly more than the other two sectors. The education sector also values meeting organisation or funder requirements (71%). Two thirds of government and commercial organisations rated using publications for internal purposes as important, higher than NGOs or education organisations. This corresponds with the arguments made about the contested nature of policy evidence and the way in which competing interests participate in evidence production [20].

Table 2. Sources of income for production of research and information materials

N= about 96 Important/Very important	**Gov %**	**Edu %**	**NGO %**	**Com %**	**All %**
Funding agreements/contracts	31	74	68	50	63
Grants	31	77	47		52
Commissions and contracts	13	64	32	67	43
Support from other parts of the org	50	41	39	25	40
Sponsors/partners	19	28	31	13	26
Memberships	6	7	35	13	19
Philanthropic/private benefactors		10	33		17
Donations		7	28		14
Revenue from sales/subscriptions/licenses		7	3	13	4
Other (N=27)	70			33	30

Notably, financial gain was not highly valued with only 11% of respondents overall identifying this as an important or very important consideration and only a quarter of commercial respondents. The production of grey literature is funded mostly from the public purse and not via sales or subscriptions, unlike commercial or scholarly publishing. Only 10% of respondents identify income as an important reason for producing grey literature, and more than 70% report that they do not try to sell their content. And only a tiny four percent actually report earning revenue from sales or subscriptions. If selling content is not a major motivation in the production of grey literature, how is income generated. Table 2 shows that funding agreements and contracts are the most important source of income for two-thirds (63%) of organisations, followed by grants, which are important for around a half (52%), but important for three quarters (77%) of those in the education sector. While some producing organisations are based in the private sector or represent business interests, it is probable that most of the material produced by government, NGOs and education is paid for through public funds.

When asked who is the target audience for their material, it is probably unsurprising that the government sector was rated as the most important for 96% of respondents across all sectors. Even those in government are trying to communicate with their colleagues. To reinforce this finding, politicians are the second most important audience for 80% of organisations. The third key audience group is practitioners (74%), indicating that producing organisations are trying to influence not only policy, but also its implementation. Income sources

5. What Gets Produced and How Important Is It?

The types of resources that are produced, circulated and consumed for policy and practice work are many and varied. Some can be characterised as research, defined as creative work undertaken on a systematic basis in order to increase the stock of knowledge and to use this knowledge to devise new applications [21]. Since the 1990s the demand for research and particularly 'evidence-based policy' has steadily increased and is often associated with promoting the use of peer-reviewed journal articles, systematic reviews and more recently a movement to promoting the use of randomized controlled trials in public policy in the UK [22] . However academic research and peer-review can both exist in publications produced directly by organisations. A range of other kinds of information and contextual knowledge also plays a part such as procedural information, policy or political statements, practitioner experience and so on [23]. Less rigorous types of investigation may provide new and useful insights on public interest issues, such as project reports, discussion papers, case studies, submissions or evaluations. And some grey literature is more informational knowledge - the translation of research as information sheets, reviews or guidelines, or the production of non-research content, such as procedures, policies, plans and strategies, stakeholder views and advocacy documents.

Table 3. Importance of materials for the organisations that produce them - %

Producers (N=144) **Material**	**Impt/ Very impt %**	**Produce Materials %**
Reports	93	76
Submissions	91	63
Evaluations	90	41
Data sets	90	37
Discussion papers	89	77
Briefings, guides, reviews	89	67
Policies, standards etc.	85	35
Websites	78	57
News reports, media releases	76	54
Working papers	74	35
Journal articles	73	43
Conference papers	69	82
Information sheets	68	54
Essays and articles	68	49
Book chapters	53	33
Social media, talk back	49	44
Audio/video material	46	35

Producers were asked to indicate from an extensive list of 25 resource types (see Table 3) including journals, books, data, reports, briefings, evaluations, news reports and many more, what their organisation produces and how important these are for the organisation. Overall conference papers are the most common resource produced by 82% of organisations, followed by discussion papers (77%), reports (76%), briefings/reviews (67%), and submissions (63%). Over half of all respondents also produce webpages/websites, news reports or media releases and information sheets. NGOs tend to produce more discussion papers and submissions, indicating the importance of advocacy work. They are also the highest producer of news reports and media releases. Government (80%) and NGOs (61%) are much higher producers of information sheets and summaries indicating the very important role of government as a translator of research and policy information for the wider community.

In a slightly different order, the *most important* resources overall for producing organisations are reports (58%), discussion papers (56%), briefings/guides (48%), conference papers (47%), and submissions (47%). If we look at results only for those that produce them, the most important materials, produced by over 50% of respondents are reports (93%), submissions (91%), evaluations and data sets (90%), discussion papers and briefings/reviews (89%), and news articles/media releases (76%). Journal articles are produced by 43% of producers and are considered important by 75% of these organisations. One reason for this is that although a small number of organisations in sectors other than education do produce journal articles, there is often little incentive or reward for this material. As an interviewee from a research company stated:

> "*We want to be held in high regard and for our work to be academically rigorous. Our clients expect that of us. But keeping people publishing in academic journals is hard. The priority is the contract, the client, the deliverable. If, at the end of the day, there's a bit of time, great, write up some articles. It doesn't feature as high as it perhaps it should.*"

Interestingly conference papers, which are produced by 82% of organisations are only important for 69% of them. It may be that conference papers are produced more as a requirement of participating or organizing conferences, which supports targeted research and organisation promotion and policy networking, rather than as an output in their own right. This is supported by the figures showing conferences are seen as an important dissemination method for 75% of producers.

6. Production and Dissemination

While many organisations have excellent reputations as producers of high quality research and policy materials, grey literature overall can be highly variable and is often considered to be not as credible as journal articles and books [5, 24]. One reason for this is a lack of standards and transparency around the way in which research has been conducted, a lack of detail about whether a publication has been peer-reviewed and in what way, and poor bibliographic details and amateur publishing practices. This makes evaluation of grey literature time consuming and fraught, and citation based metrics of grey literature difficult. It also provides opportunities for marketing and advocacy materials to be disguised as research.

Some definitions of grey literature state that it is not peer-reviewed, causing further confusion (often to be found in online guides produced by university libraries). In fact a significant amount, but certainly not all, grey literature is formally peer-reviewed or reviewed in some other way such as by an expert advisory group. While concerns about the quality of grey literature are often raised, and may well be justified for some material, just under two thirds of organisations surveyed indicate that they 'often or always' conduct an internal review or use an advisory group (60%) or have their work peer-reviewed (55%), with about a quarter (26%) using an external board to review prior to publication (26%) (Figure 1). Almost all organisations surveyed (90%) undertake basic editing and formatting of their content in-house and professional editing is used by 39% of producing organisations. While this is good news, and an indicator of a fairly high level of scrutiny and credibility going into some policy grey literature, the issue remains that even professionally produced and peer-reviewed material may not include all the information required by information users and collectors to evaluate the document they have discovered through search engines, social media, email or some other method.

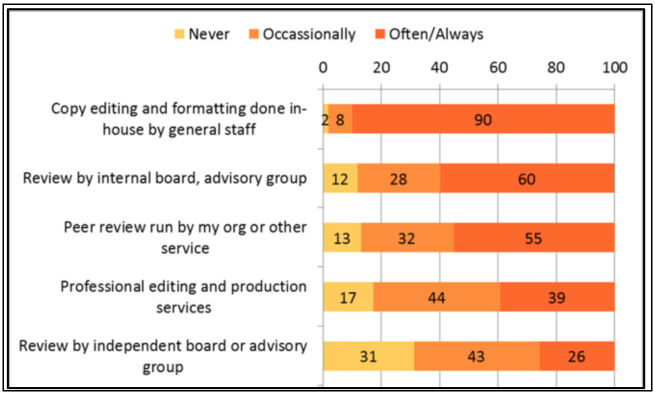


Figure 1. Quality control carried out on materials produced by organisations - % frequency

Once content is written, revised and ready, documents have been desk-top published and the ubiquitous PDF prepared, organisations need to make them publicly available and distributed to their audience. Publishing and particularly distribution were amongst the most difficult and expensive tasks for organisations in the print era however the transformative difference of the internet means that once the creation and production costs have been covered, dissemination costs are minimal. Producing organisations overwhelmingly choose to publish content directly on their organisation website with free to access to anyone with the link, and this is possible thanks to almost zero copy costs. Seventy percent of the organisations surveyed made between 50-100% of their material available online or in print for free in the last 12 months.

Posting a publication on a website does not necessarily ensure an audience and to achieve policy impact may require a more active campaign of promotion and attempts to alert users to the new material, either directly or via third parties such as the media and information services. In the digital world, publishing, promotion, dissemination and access blur as roles converge and traditional activities and business models are disrupted.

As Table 4 shows, the most widely used methods of dissemination, rated important or very important by over three quarters of organisations are: providing access via an organisation's own website (89%), sharing information directly with contacts (85%) sending email newsletters (83%), presenting at conferences or events (75%) and sending out details via other organisations' newsletters (73%). Over 90% of NGOs rated all these methods as important or very important.

Table 4. Most important methods used for alerting audiences to new material

N= about 92 Important/Very important	Gov %	Edu %	NGO %	Com %	All %
Own organisation's website and/or intranet	83	88	95	75	89
Sharing information directly with contacts	81	77	93	86	85
Organisation's email newsletter/list	88	73	98	38	83
Events or conferences	44	73	93	57	75
Email newsletters/lists	71	61	93	25	72
Media releases	44	58	68	14	57
Social media e.g. Twitter, Facebook, etc	50	50	63	43	55
News reports and articles	39	53	63	14	52
Post on partner websites	44	23	53	25	39
Print copies sent to key contacts	31	34	49	25	39
Subject databases and clearing houses	27	55	34		37
Institutional repositories	13	47	21	13	28
Alerts/RSS	20	28	27	43	27
Blogs	13	20	24	14	20
Libraries incl Trove	33	27	3		16
Journal subscriptions	14	31	3		14
Advertising	13	4	11	14	9
Other/No response (N=20)	14	20	33		20

7. Storage, Access and Long Term Management of Grey Literature

In the print era publishing was often described in terms of a chain, where a publication worked its way from an author's manuscript, to an agent, then a publisher, printer, distributor, bookstores and libraries and finally to a reader [2]. Other participants might assist in this progress such as newspapers and magazines publishing reviews but access to the actual item was channeled through key points. The print grey literature chain had some similarities, beginning with either an author or an organization producing content then having this printed. From there things got harder, distribution of print materials was costly, often involving direct mail to key people and special events. Collection and management was also problematic with many items not having ISBNs and therefore not coming to the attention of the legal deposit agencies. Special libraries, 'clearing houses' and private collections were required to bring together key resources and publications for niche audiences.

In the internet era, the publishing business may still involve a range of commercial and public services including authors, agents, publishers and organisations, distributors, booksellers, ebook vendors, libraries, subscription or open access databases, and promotion through media outlets. Or one or all of these roles may be done by the publisher or producing organization or any other element. Roles have converged as publishers and producing organisations can commission, produce, publish, distribute, promote, provide access and discovery systems and store for long term preservation.

Organisations that are not professional publishers may, or may not, have the experience, personnel, resources or motivation to meet best practice publishing standards or adequately manage all of the requirements and responsibilities to ensure effective and long-term discovery, access and preservation of their publications. In a number of ways electronic online publishing by organisations has made this situation worse, as more and more organisations now produce content and publish it online without adequate publishing or standards and little or no plan for long term management. Online publishing has also conflated activities that were previously separate such as promotion, discovery, access and storage with many organisations now undertaking or being responsible for all four of these to varying standards.

Discovery and access which was previously the role of the bookseller or library could now be provided by an organization directly as a ‘retail service’. From the survey of producing organisations we found that storage and access to publications and resources is usually managed by organisations themselves with most surveyed producing organisations (85%) ‘often or always’ using their organisation website to store and provide access to their content. Unfortunately websites are notoriously unstable with redesigns often causing hosted files to be removed from online. As an interviewee summarized things: *“The problem with the website is that it's hard to maintain so once you put something up there it's this static document that then if the website is changed the document could go.”* (Producer interview, government sector)

Less than half (46%) report having their own repository software to provide a stable hosting platform or long term management. Just over half of the surveyed producers based in education deposit their material with their institutional repository, suggesting that these systems could be better utilised for grey literature produced by universities’ centres and departments. Beyond this there is little take-up of external databases, libraries or other curatorial services, even though many are free (for example APO.org.au). Only 20% of organisations surveyed comply with their obligation under legal deposit to provide a print copy to the National Library or a state library.

8. Linkrot

In preferring the ease and immediacy of their own websites rather than more stable options such as repositories, producing organisations are major contributors to the proliferation of linkrot across the internet and within the reference lists of many publications. Linkrot or reference rot is the loss of access to online content when it is moved or removed, often as a result of website upgrades or changes. Studies estimate that the rate of loss of digital content is around 30% within a few years of publication online [25, 26]. Despite posting most of their content on their own website, only 26% of the producing organisations we surveyed have a strategy in place to prevent linkrot. Of the rest, 42% know they have no strategy in place and a third don’t know either way.

While it may seem like many producers are mainly focused on the short term this isn’t necessarily the case but neither individual authors or the producing organization are in a position to be able to provide long-term management. This is particularly the case for academic research projects or other funded projects. “Every project that we've done we've had a website built... Then the project ends, the website you have to keep paying for or you have to archive and if it gets archived after a couple of years the material disappears. It’s so frustrating.” (Producer interview, Education sector)

This is no better for government, in fact possibly worse. The survey data shows that many government organisations have little knowledge of what plans are in place to ensure ongoing access to public sector information, despite recent campaigns for open governments and policies to improve government management of information [27, 28]. A recent investigation into the withholding of government research in the UK was surprised to find that there is no comprehensive account of how much research is commissioned by government or if it has ever been published [29]. Despite the poor management and lack of strategies for dealing with deadlinks, 55% of producing organisations agreed with the statement that 'Providing long-term access to our online content is not an issue for my organisation'. It would seem that while many organisations are aware of the potential to lose content they are not actively doing much about it. So if it is not a priority for over half of all producers, who is going to provide long-term access to policy research?

When asked why they don't have a strategy, the most common reason given by around a third of producing organisations is that they hadn't had the time or resources to deal with it. Another third either hadn't thought about it or didn't consider it important. The loss of online content resulting from poor resource management is seen as a serious issue by only a third (37%) of producers. A further 37% were somewhat concerned, while a quarter thought it was not that much of a problem. It is interesting to compare this attitude to that of commercial publishers who have realized that their backlist of electronic books and journals is a potential asset that they can continue to sell or even resell to libraries and subscribers that had already purchased print copies. This comes back to the issue of the financial model of public interest publishing which has public funding supporting production but does not adequately address access, discovery and management of publicly funded grey literature.

9. Improving Production Practices

From these survey results and interviews we can begin to see that grey literature operates at a scale and importance that needs to be taken seriously as part of the public policy debate and as a form of communication and knowledge production. Yet while this content is of great value, usually paid for through public funds, it is not produced in any systematic way that will ensure it is available to discover or access much beyond the media cycle in which it is intended to feature. Publicly funded research should include provision for the sustainable management of outputs and collection strategies. Expectations tied to funding create incentives for large-scale change that can be managed flexibly and with discretion at the individual and organisation level. For example grey literature should be integrated into future assessments of research impact and quality currently being reviewed by the Australian Research Council [30]. Similarly copyright reform to introduce fair use principles will support great sharing and reuse of policy resources [31].

With so many organisations producing material, evaluating the credibility of their work often requires knowledge of organisations in the field and their role and legitimacy. The task is often made harder because many organisations do not include adequate bibliographic information in their publications, and do not work with collecting services to improve discovery and long-term accessibility. As we have seen from the survey results most organisations manage their own production and publish directly online. Despite the use of some review systems and professional editors professional publication standards are often overlooked. This is an issue that is raised constantly by those using or collecting grey literature.

Clearly there is considerable scope for producers to improve standards. Simple steps would be to ensure basic bibliographic information is included in all their work, together with a clear statement of any reviewing process. Many organisations may simply be unaware that certain information is essential for users and collectors to be able to make an assessment of a document. Given this, it would be relatively easy to improve publication standards of grey literature with the development and adoption of clear publishing guides. By including essential bibliographic information — date of publication, authors, producing organisation, a copyright or creative commons statement, page numbers, and a web address — producing organisations could reduce use and collection costs and have a major impact on the accessibility and credibility of their work.

Producers can make sure their publications look professional and can be discovered by search engines and by readers, curated – by information services, clearing houses and libraries, evaluated – by anyone who wants to use it, and cited and measured by including basic bibliographic information on every resource published. A mnemonic to help remind authors and producers is T.A.P. D.A.N.C.E. [32].

10. The Economic Value of Organisation Publishing

As part of this research we have done some estimates of the economic value of grey literature production, based on reported time spent producing publications calculated to the national level based on the estimated population of grey literature producers discussed earlier. Australian respondents reported spending an average of 31% of their working time per week creating grey literature during the last 12 months (N=92). Similarly, Australian organisations reported spending an average of 12 hours per week creating grey literature during the last 12 months. That is a total of 1,122 hours for the 92 respondents. At average weekly wages plus on-costs, average annual grey literature creator costs amount to around $29,385 per person per annum, or $377 million per annum across the survey respondents.

If the population of grey literature producer organizations is 30,000, and their average staff numbers are similar to those of our survey respondents and their grey literature activities are, on average, one-third those of respondent organizations, then total national grey literature creation cost might amount to some $30 billion per annum. Respondents reported their organization or department spends a total of $234 million per annum on projects that result in the production of grey literature, an average of $3.3 million per annum per respondent. On this basis total national grey literature related *project* spending could be around $33 billion per annum. National R&D spending in Australia is $28 billion per annum so this seems plausible.

Australian respondents reported generating almost $80 million per annum from the sale and/or distribution of grey literature, an average of almost $1 million per annum per respondent organization (N=80). Scaling reported revenues generated from the sale and distribution of grey literature suggests national revenue of $10 billion. This implies a 32% cost recovery, with much grey literature made freely available.

11. Conclusion

If we are to have a more evidence-informed public sector in Australia or any other country, there needs to be greater recognition and long term support for the diverse range of data and publications used for policy analysis as well as investment in technical and managerial skills for producing, managing and providing access to multiple sources. This requires a policy culture that is supportive of transparency and knowledge sharing but also a recognition from academics that public policy is a highly contested space with diverse producers of knowledge and that policy makers rely on a wide range of sources [33]. Given the scale and significance of grey literature in public policy and the level of public investment, we need to ensure that it is produced to a standard that supports easy evaluation, correlation and analysis and that there is adequate investment in its management and collection so that it can be discovered and accessed today and into the future.

References

[1] Lawrence A, Houghton J, Thomas J, Weldon P. Where is the evidence: realising the value of grey literature for public policy and practice: Swinburne Institute for Social Research; 2014 [10 Nov 2015]. Available from: http://apo.org.au/node/42299.
[2] Thompson JB. Merchants of culture: the publishing business in the twenty-first century. Cambridge, UK: Polity; 2010.
[3] Luzi D. Trends and evolution in the development of grey literature: a review. The International Journal on Grey Literature. 2000;1(3):106-16.
[4] Lawrence A. Electronic documents in a print world: grey literature and the internet. Media International Australia (8/1/07-current). 2012(143):122-31.
[5] Macdonald BH, Soomai SS, De Santo EM, Wells PG. Science, Information, and Policy Interface for Effective Coastal and Ocean Management: CRC Press; 2016.
[6] Farace DJ, Schöpfel J, editors. Grey literature in library and information studies. Berlin: Walter de Gruyter; 2010.
[7] Couldry N. Media, society, world : social theory and digital media practice. Cambridge: Cambridge : Polity; 2012.
[8] Thompson JB. Books in the digital age. Cambridge UK: Polity; 2005.
[9] Briggs A, Burke P. A social history of the media: from Gutenberg to the internet. 3rd ed. Cambridge, UK: Polity Press; 2009.
[10] Lobato R, Thomas J. The Informal Media Economy. Cambridge: Polity Press; 2015.
[11] Burgess J, Bruns A. Easy data, hard data : the politics and pragmatics of Twitter research after the computational turn. In: Elmer G, Langlois G, Redden J, editors. Compromised Data From Social Media to Big Data. 1st ed.. ed. New York: New York : Bloomsbury Publishing; 2015.
[12] Edwards PN, Jackson SJ, Chalmers MK, Bowker GC, Borgman CL, Ribes D, et al. Knowledge infrastructures: Intellectual frameworks and research challenges. Working Paper. University of Michigan School of Information, 2013 2013-05. Report No.
[13] Edwards PN, Jackson SJ, Bowker GC, Knobel C. Understanding infrastructure 2007 [25 January 2016]. Available from: http://cohesion.rice.edu/Conferences/Hewlett/emplibrary/UI_Final_Report.pdf.
[14] Lawrence A, Thomas J, Houghton J, Weldon P. Collecting the evidence: improving access to grey literature and data for public policy and practice. Australian Academic and Research Libraries Journal. 2015.
[15] Thomas J, Lawrence A. The power of observation: making research visible, accessible, and usable for the Asian digital economy [Text]. Hong Kong: Digital Asia Hub; 2015 [updated 2015-11-27]. Available from: http://apo.org.au/node/60499.
[16] ABS. Not-for-profit Organisations, Australia, 2006-07 Re-issue. Australian Bureau of Statistics, 2009 Cat No 8106.0.
[17] ABS. Australian National Accounts: Non-Profit Institutions Satellite Account, 2012-13. Australian Bureau of Statistics, 2014 5256.0.
[18] ABS. Counts of Australian Businesses, including Entries and Exits, Jun 2008 to Jun 2012, Cat No 8165.0: Australian Bureau of Statistics; 2013. Available from: http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/4461C2EEC98E3A47CA257B7100149A76/$File/81650_jun20200820to20jun202012.pdf
[19] Ibis World. Management consulting in Australia: market research report: Ibis World; 2015 [15 October 2015]. Available from: http://www.ibisworld.com.au/industry/default.aspx?indid=1896.
[20] Wesselink A, Colebatch H, Pearce W. Evidence and policy: discourses, meanings and practices. Policy Sciences. 2014;47(4):339-44.
[21] OECD. Frascati Manual: Proposed Standard Practice for Surveys on Research and Experimental Development Paris: OECD; 2002 [cited 2014 11 November 2014]. 6th:[Available from: www.oecd.org/sti/frascatimanual.

[22] Pearce W, Raman S. The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance. Policy Sciences. 2014;47(4):387-402.
[23] Nutley SM, Walter I, Davies HTO. Using evidence: how research can inform public services. Bristol, UK: Policy Press; 2007 2007.
[24] MacDonald BH, De Santo EM, Quigley K, Soomai SS, Wells PG. Tracking the Influence of Grey Literature in Public Policy Contexts: The Necessity and Benefits of Interdisciplinary Research. Grey Journal (TGJ). 2013;9(2):61-8.
[25] Bugeja MJ, Dimitrova DV. Vanishing act : The erosion of online footnotes and implications for scholarship in the digital age. Sacramento, USA: Litwin books; 2010.
[26] Klein M, Van de Sompel H, Sanderson R, Shankar H. Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. PlosOne. 2014;9(12).
[27] OAIC. Towards an Australian Government information policy Canberra: Office of the Australian Information Commissioner; 2010 [15 June 2011]. Available from: http://apo.org.au/node/23119.
[28] OAIC. Open Public sector information government in transition Canberra: Office of the Australian Information Commissioner; 2013 [4/4/2016].
[29] Sedley S. Missing evidence: an inquiry into delayed publication of government-commissioned research [Text]. Sense about Science; 2016 [updated 2016-06-022016-06-02]. Available from: http://apo.org.au/node/64609.
[30] ARC. Engagement and impact assessment consultation paper [Text]. Canberra: Australian Research Council; 2016 [updated 2016-05-02 00:00:00]. Available from: http://apo.org.au/node/63886.
[31] Commission P. Intellectual property inquiry 2016. Available from: http://apo.org.au/node/72016.
[32] Lawrence A. Publishing online: the Tap Dance guide: Swinburne Institute for Social Research; 2014. Available from: http://apo.org.au/node/42303.
[33] Head B. Toward more “evidence-informed” policy making? Public Administration Review. 2015;76(3).

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-100

Arxiv-Based Commenting Resources By and For Astrophysicists and Physicists: An Initial Survey

Monica MARRA[1]
INAF – Osservatorio Astronomico di Bologna

Abstract. This paper results to be the first, though absolutely initial, overview of commenting platforms and other web 2.0 resources which were born for and within the astrophysical research community, from 2004 to 2016. Additional experiences, chiefly in the physics domain, were added for a total of twenty-one tools, inclusive of four items in the specific area of epijournals – plus four supplementary resources which have been simply mentioned or anyway much more synthetically described due to their specific features –, thus casting some light onto an unexpected richness and consonance of endeavours. These experiences rest on the contents of the pioneering database ArXiv, which adds to its universally recognized merits that of setting the grounds for web 2.0 resources, and research behaviours, to be put in place. These resources were surveyed substantially through the method of empirical evidence, partly routed by the web resources examined and by some of the literature, and are accounted for in a time sequence for their essential features. Most of the experiences retrieved are UK- and US-based, but other countries have been involved, such as Italy, the Netherlands and China. Final remarks are sketched. The results integrate the previous studies according to which the web 2.0 is presently of limited use for scholarly communication within the astrophysical community. Collaterally, some aspects of ArXiv's recent pathway towards partial inclusion of web 2.0 features are touched upon. The centrality of the scholarly literature for web 2.0 interactivity in astrophysics and – more presumably – in some other branches of the physics domain emerges as a plausible hypothesis and as a promising research suggestion. Further investigation is not only needed, but also absolutely hoped for.

Keywords. open commenting, open science, web 2.0, ArXiv, astrophysics.

1. Methodology

The research that follows can be estimated to have required about thirteen months of activity (FTE); it was partly conducted alongside with the preparation work for different projects. The first documentation (both literature and web resources) was retrieved in late 2014, the last one in March 2017.

Search engines have proved to be of limited usefulness in order to let these resources come to light. Queries have been executed with the phrases: 1) "arxiv comment*", 2) "arxiv discuss*", 3) "arxiv peer review*". The first three pages of results (30 items) for these queries yielded a total of five of the items here presented,

[1] Monica Marra, INAF-Osservatorio Astronomico di Bologna, Via Gobetti 93/3, I – 40129 Bologna, Italy; E-mail: monica.marra@oabo.inaf.it.

meaning that only 23.8% of the twenty-one main resources here described were retrieved directly through the search engine.

Other important sources for the present findings have been previous online compilations, such as the wide shared database *400+ Tools and innovations in scholarly communication* ([1], last visited March 20, 2017), first published in March 2015 by Bianca Kramer and Jeroen Bosman of Utrecht University Library and then constantly updated. As at March 21, 2017, it lists as many as 668 resources. This unique collection has been thoroughly consulted in spring 2016, with subsequent inspections later in 2016 and in 2017. At the moment of writing, it contains seven of the resources in the present survey (33.3% of the main group), only one of which - *ViXra* - can be retrieved also through the search engine above, plus *Discrete analysis* among the simply mentioned ones. The utility of this resource has been concrete and unquestionable; anyway, due both to the continuous update of the database and to the prolonged and multiple-source documentation activity which has brought to the present results, it would be difficult to reconstruct exactly, and retrospectively, the percentage of *400+ tools* which represented an actual source for our findings.

Also very important in order to build up the present survey were (social) media mentions, with special referral to blog comments suggesting resources. In fact, tracking comments on appropriated blog threads for as much as it has been possible, as well as practising web browsing to some extent starting from the resources already retrieved, have proved to be fruitful strategies for getting to the present findings. Actually, it can be realistically estimated that they made it possible to retrieve eleven out of the twenty-one resources in the main list (52.38%), plus one of those simply mentioned. A colloquium with an astrophysicist was the original source for a further platform, *Cosmocoffee*.

Precious details about some of the projects surveyed came from email exchanges with some of the researchers involved, as will be detailed below.

The majority of the bibliographic references was found in 2016 (~61%), with a further 31.7% being filed between 2013 and 2015. In fact, though, the literature was more useful for giving a profile to some of the themes involved than for providing concrete examples that be useful to the building of this survey. In fact, the literature was the original source for only two of the resources retrieved (*Naboj* and *The RIOJA Project*) – which doesn't mean that some more of these resources haven't received attention by researchers, journalists or bloggers, as indicated in the reference list.

The criteria for selecting the resources in this survey were: (a) having been created by researchers, (b) for their same scholarly communities, and (c) relying on ArXiv contents entirely or almost entirely.

2. The Importance of Arxiv Beyond Preprints Provision

The creation of ArXiv, the first and foremost preprint server in 1991, has been recognized as "the most significant change in scientific communication since the establishment of the journal in the 17th century" [2]. The importance of this novel way of circulating scientific papers much exceeds that of enhancing papers' availability in a peculiarly early stage of their customary disclosure. In fact, Arxiv has pioneeringly explored all the main changes in XX and early XXI century's scholarly communication practices – among which the progressive diffusion of the open access movement. The latter found ArXiv giving researchers the opportunity to upload accepted or published versions of papers, thus putting those principles into practice for the communities involved, while the massive hosting of preprint papers let ArXiv be perceived as an implemented source for open contents, in spite of the conceptually specific nature of this task.

In fact, ArXiv's fruitfulness went beyond. As early as in 1994 - two years before it's often stated to have happened - Paul Ginsparg envisioned the possibility for ArXiv to act as a starting platform for add-on tools fostering not only dissemination but also validation practices, the latter through the birth of a network-based scholarly interactivity centered on ArXiv papers ([3], [4]). The classic article by Rodriguez *et al.* [5] shows how cleverly these suggestions could be seized just after the landmarking debut of the web 2.0 around 2005 [6]. At the same time, traditional peer-review's features and role in the science production chain were increasingly questioned ([7], [8]; a review is in [9]), while the milestone phenomenon of web 2.0 slowly began transforming academic practices – as acknowledged even in cautious scholarly perspectives ([10]). Useful studies have aimed at tracking the process and at casting light on a variety of 2.0 tools for the scholarly communities, as well as on patterns of their use ([11], [12], [13], [14], [1], [15]).

The role of ArXiv within this global, substantial paradigm change doesn't result to have been fully investigated yet. Also, to the best of our knowledge there aren't any comprehensive studies about how the web 2.0 attitude has progressively affected the astrophysical field, although some particular aspects were conveyed. Polydoratou and Moyle [16] have interestingly surveyed astrophysicists' attitudes towards ArXiv overlay journals in 2007, in the context of a specific project that will be accounted for *infra*. Valuable observations based on narrative interviews on this community's attitudes about communication and publication practices have been made by Harley *et al.* [17]. These authors maintain that "astrophysicists have limited engagement with Web 2.0 technologies". They also highlight the role of "email networks" for communication, and stress that "face-to-face interaction remains an essential part of the collaborative process", which has later been confirmed by Marra [18] in a specific context, and by Delfanti [19] for high energy physicists.

The use of Twitter among astrophysicists has received considerable attention in more recent years ([20], [21], [22]); the conclusions seem anyway to downplay its role for internal scholarly communication, although from the present point of view it has been interestingly noted that "most tweets refer to the ArXiv instead of the publishers' versions" [21] .

Ritson [23] has examined socio-scientific aspects of the "trackback system" connecting ArXiv papers and scientists' blogs since 2005, with an account of the science-and-technology-studies (STS) literature on the subject. From the present perspective, three points result to be fruitful: a) blogs, although peculiar in type, may well be considered means for providing papers with scientific feedback, included peer-review; b) in 2006, one year after the debut of the trackback system, blogs ArXiv had approved for trackback were 51 and trackbacks were 5132. If considered that (c) the high-energy physics community has long been discussing in order to find consensus on how to practically identify members enabled to have their blogs trackbacked to the ArXiv, these numbers cast light on a phenomenon that may well be considered potentially wider and significant.

Within the scientific communities, the topic of providing ArXiv with commenting or even peer-review-type capacities – or not – has long been debated, as researchers' blogs and forums can witness. An almost randomatic sampling – including the threads https://www.physicsforums.com/threads/a-peer-review-system-for-the-arxiv.568276/ (2012; last visited March 8, 2017) and http://academia.stackexchange.com/questions/32367/why-doesnt-arxiv-have-a comment-section (2014, last visited March 8, 2017) – can provide an interesting insight into the views of shrewd and lively scientific communities.

2.1. Arxiv and Its Present Situation With Respect to the Web 2.0 Setting

It may appear somehow paradoxical that ArXiv, whose creator had so impressingly timely envisioned his database's web 2.0 potential, hasn't been equipped with tools for the new web ecosystem so far - notwithstanding ArXiv's persisting role as a pillar resource for astrophysicists. Paul Ginsparg's explanation for this slow pace has been the database's organizational framework due to budget and personnel constraints [24]. In fact, things seem now to be changing to some extent, as in April 2016 ArXiv conducted an online survey among its users in order to "improve arXiv and think of future directions for the service" [25]. One of the questions was aimed at measuring users' perception of importance for possible new services, among which the addition of a rating system for recommending papers and an annotation feature allowing commenting. The inclusion of these question looks meaningful and the subsequent development of a "next-generation arXiv" project in the next three years [26], although still not detailed, may let interested people envisage a renewed approach to these themes in the medium term.

As for now, the situation of ArXiv, jointly with its persisting overall popularity confirmed by the 2016 survey - 52.92% "very satisfied"; 42.43% "satisfied" – could bring to the supposition that the present, limited web 2.0 evolution of the database goes well with the astrophysical community's still prevailing inclination to tendentially preserve its scholarly practices.

In fact, there is significant evidence of commenting practices to ArXiv papers much beyond these traditional channels, with proper involvement of the web 2.0 setting.

3. Commenting on ArXiv

A largely practised mode of online interaction is represented by researchers' blogs and forums, which may comment on ArXiv papers. This specific channel is being barely mentioned here as the complexity of the scenario and the relations with ArXiv through the so-called "trackback system" [23] would require a dedicated and very extended analysis. A single experience will anyway be accounted for and it's the one of the astrophysical forum *Cosmocoffee* (http://cosmocoffee.info/, last visited March 1, 2017), born in September 2004, which had 2769 registered users as at March 1, 2017 (http://cosmocoffee.info/index.php, last visited March 1, 2017). Although in fact it results to be a multi-purpose information resource as it includes different-type information such as (but not limited to) conference announcements, job vacancies, discussions on three specific topics (e.g. "cosmological model"), founders "hope that it can also become a useful reference resource, complementing the arxiv itself" (http://cosmocoffee.info/faq.php#0, last visited March 1, 2017). "Daily we discuss work and new papers with colleagues, either at our local coffee break or via email with colleagues all over the world. This discussion can be an extremely effective way to understand things better. As such, it seemed to make sense that those discussions be shared with others and be public. [...] Therefore we set up cosmocoffee.info as an attempt to facilitate this. The forum "is intended for authorised arxiv authors and students" (http://cosmocoffee.info/faq.php#0, last visited March 1, 2017). Posts can be read by both registered and unregistered users, but posting is only for the formers. Within *Cosmocoffee*, the sub-forum "ArXiv papers" appears to have started with a post by UK cosmologist Antony Lewis on September 24, 2004. This section has received a total of 1031 posts on 260 topics ((http://cosmocoffee.info/index.php, last visited March 1, 2017), with present last post dated August 15, 2014 (as at March 1, 2017); other sections (especially "computers and software", "job vacancies", conferences and meetings") are still current. Cosmocoffee's administrators result to be Sarah Bridle (University of Manchester), Olivier Dore (JPL-CalTech), Antony Lewis (University of Sussex) and Mike Nolta (Canadian Institute for Theoretical Astrophysics) (http://cosmocoffee.info/faq.php#0, last visited March 1, 2017; http://cosmocoffee.info/ memberlist.php for (present) affiliations, last visited March 1, 2017). For as much as it results, *Cosmocoffee* has never been object of dedicated studies.

The present survey will focus on different-type resources which offer commenting features in the astrophysical and physical fields.

For presentation purposes, it seems possible to roughly divide them into three main categories:

a) resources or projects aimed at buiding new, open access and more interactive forms of the traditional scholarly journals. The model is that of the "overlay journal" or "epijournal" ([27], [28], [29]);
b) ArXiv-based commenting platforms. In some cases, they may be aimed at practical purposes such as selecting papers for "real life" scholarly discussions;
c) different tools which can very roughly be defined as variant forms of ArXiv – with whom they have no kind of affiliation or other apparent link. They will be conveyed firstly, due to their peculiar characteristics. The tools in this section generally have more limited web 2.0 capacities and are considerably different both from each other and partly from ArXiv, too. They witness a widespread effort to build upon the model, as well as ArXiv's totemic standing within the physics and astrophysics environments – e.g. in the names and graphic look of the first two tools in the list.

3.1. "Variant" Forms of ArXiv (3.c.)

The definition, as said before, is intentionally very broad and pragmatic, in order to group together online entities with a commenting feature appearing to be secondary compared to changes in some of ArXiv's main features (either improved search functions, or renewed visualization features of the original database, or adaptation to a different audience, or changes in authors' admission policy).

- **3.c.1.** *ViXra* (2009; http://vixra.org/, last visited March 17, 2017), created by the independent UK-based physicist Philip Gibbs, "has been founded by scientists who find they are unable to submit their articles to arXiv.org because of Cornell University's policy of endorsements and moderation [...]. ViXra is an open repository for new scientific articles. It does not endorse e-prints accepted on its website, neither does it review them against criteria such as correctness or author's credentials" (http://vixra.org/, visited March 17, 2017).. Its aim is "to enable anyone to distribute their works of science and mathematics irrespective of their status or affiliations" (http://vixra.org/faq, visited February 2, 2017). As at February 2, 2017 it contained 17632 outputs, 1093 of which in astrophysics (http://vixra.org/). Some information about *ViXra* is in [30] and [31]; the tool is listed in [1].

- **3.c.2.** *SnarXiv* (http://snarxiv.org/, last visited March 17, 2017) was born in 2010 by initiative of David Simmons-Duffin, by that time a PhD student in high-energy physics at Harvard University, and in a somehow bohemian spirit. "The snarXiv is a random high-energy theory paper generator" (http://davidsd.org/ 2010/03/the-snarxiv, visited February 2, 2017). It contains an interactive game: "arXiv vs. snarXiv" (http://snarxiv.org/vs-arxiv/), where players have to spot genuine ArXiv titles from *SnarXiv* ones.
- **3.c.3.** *Astrobites* (https://astrobites.org/, last visited March 20, 2017) is a successful project created in 2010 by and for undergraduate students in astrophysics (https://astrobites.org/about/, last visited March 20, 2017). Its "goal is to present one interesting paper per day in a brief format that is accessible to undergraduate students in the physical sciences who are interested in active research" (*ibid.*) – although in fact it's also a web portal for different –type information. Typically, the papers suggested are from ArXiv's astrophysics section "astro-ph". The resource is written by "a team of graduate students at universities around the world" (*ibid.*) – but apparently mainly based in the US and in the UK (https://astrobites.org/2017/01/03/astrobites-a-look-back-at-2016/, last visited March 20, 2017) – which reached thirty members as at January 3, 2017. Past web hosting was at Harvard University, with the help of James Guillochon (https://astrobites.org/about/); remarkably, "since 2016 Astrobites has been hosted and supported by the American Astronomical Society". Links to other 2.0 resources here included are provided (*VoxCharta*, *ArXiver*); Astrobites has an account on Twitter (@astrobites), and another one on Facebook (https://www.facebook.com/astrobites/) with more than 5500 and more than 3700 followers respectively.
- **3.c.4.** 2013 saw the debut of *PaperScape* (http://paperscape.org/, last visited March 17, 2017), "an interactive map that visualises the arXiv" according to number and typology of the citations received. As one of the authors explains, it "visualises the entire arXiv database as a map that can be explored by panning and zooming. The papers are sized according to their number of citations and positioned according to their references/citations. Different categories of the arXiv are assigned different colours, and newer papers are more brightly coloured. The original project complements this map by letting you draw graphs of the papers that interest you, with the papers as nodes and citations as links. It's possible to register a personal profile, with which you can tag relevant papers as well as save and share the graphs you make" [32]. Developers are young physicists Damien P. George, currently at the Department of Applied Mathematics of the University of Cambridge, and Robert J. Knegjens. [1] includes it.
- **3.c.5.** Late 2013 saw the debut of *arXiver* (initially https://arxiver.wordpress.com/, then http://arxiver.net/, last visited February 14, 2017), whose "original credit for the idea" is acknowledged to the young British

astrophysicist and web 2.0 activist Robert Simpson (http://www.arxiver.net/about/, visited February 14, 2017); the resource is (co-)maintained by Australian postgraduate student Vanessa Moss. The first post available is dated October 8, 2013, (https://arxiver.wordpress.com/2013/10/08/, visited February 14, 2017). While staying updated with the literature through ArXiv is said to be highly appreciated, it is also maintained that this database is presently "not very nice to look at (too much text!)" and "it would be nice to be able to glance at a visually-appealing summary of different papers to then go forth and read properly" (https://arxiver.wordpress.com/about/, visited February 14, 2017); this seems to basically consist in providing meaningful, selected pictures from the article by the side of the ArXiv abstract. In fact, readers can also assign "likes" to papers' posts, but for as much as it's possible to see this has scarcely ever happened. An interesting feature was the initial absence of author names in new papers' posts, in order to correct for any possible author bias (later on, authors' names have been included up to the first three authors; https://arxiver.wordpress.com/faq/, visited February 14, 2017). Since its debut, *ArXiver* was equipped with a Twitter feed, @arXiver, which has 727 followers as at February 14, 2017.

- **3.c.6.** *Cloudy Science* (https://cloudyscience.wordpress.com/, last visited February 14, 2017) was born presumably either in 2014 or shortly before, but "revived" in January 2015 "after a long period of stagnation" (https://cloudyscience.wordpress.com/updates/, last visited February 14, 2017). It is defined as a "partner site" by *ArXiver* (http://arxiver.net/, last visited February 21, 2017). "The goal of Cloudy Science is to present automatically generated wordclouds that give a researcher insight into the content of a paper, offering another way to quickly judge whether a paper might be […] relevant to them. It currently only focuses on arXiv's astro-ph" (https://cloudyscience. wordpress.com/about/, visited February 14, 2017). Readers can assign "likes" to single papers, but this feature appears to have been very scarcely used; also, papers can be shared to Twitter, Facebook and Google+. At the moment of writing, *Cloudy Science* is "brought to you" by Australian postgraduate student Vanessa Moss (Sydney Institute for Astronomy, http://sydney.edu.au/science/people/vanessa.moss.php, visited February 14, 2017).

ArXivist (http://arxivist.com/, last visited March 13, 2017) and *ArXiv Sanity Preserver* (http://arxiv-sanity.com/, last visited March 13, 2017) were both born in 2016; they also share the feature of using readers' preferences – as provided in a web 2.0 environment – for customizing ArXiv daily updates for users accordingly. Both developers (Anton Lukyanenko and Andrej Karpathy respectively) are US-based and are active in the mathematic field (the former) and in computer science (the latter), which suggests not to get into further details in the present context.

3.2. ArXiv-Based Overlay Journals (3.a.)

Mathematicians, computer scientists and physicists have notoriously shown a rather active attitude about the implementation of ArXiv-based overlay journals ([28], [29]; early examples in [33]). Meaningful samples of some computer scientists' views on the subject, supplemented by a specific project, can be read at the blogpost "Scientific journals in the e-publishing age", written by computer scientist Philip Thrift on his blog "Occupy publishing" on February 1, 2012 and widely commented (http://occupy publishing.blogspot.it/2012/02/scientific-journals-in-e-publishing-age.html, last visited March 14, 2017). New projects have recently enriched this scenario, such as mathematician Tim Gowers' *Discrete analysis* (http://discreteanalysisjournal.com/, 2015, last visited March 17, 2017; announcements on Gowers' blog, e.g. https://gowers. wordpress.com/2016/03/01/discrete-analysis-launched/, last visited March 17, 2017).

New achievements have been accomplished in physics, too and will be accounted for in more detail.

- **3.a.1.** Dutch platform *Scipost* (https://www.scipost.org/, last visited March 20, 2017), born in 2016, presently provides two ArXiv overlay journals, "SciPost Physics" and "SciPost Physics Lecture Notes", whose contents are published under the CC-BY 4.0 license (https://www.scipost.org/FAQ, last visited March 21, 2017) and equipped with DOIs. Commenting is possible for registered *SciPost* contributors. "Scipost Physics" publishes research articles in experimental, theoretical and computational physics, including cosmology and astroparticle physics (https://www.scipost.org/journals/scipost_physics/ about, last visited March 20, 2017); as at writing, three issues have been published for a total of twenty-five accepted articles. Outstandingly, *Scipost* is endorsed by the Netherlands Organization for Scientific Research (NWO) (https://www. scipost.org/, last visited March 20, 2017), which presently supports for operational costs (https://www.scipost.org/FAQ). *Scipost* relies upon a wide, international editorial college of almost fifty members (as at March 21, 2017); the advisory board includes eleven academics from the Netherlands as well as from Italy, France, Switzerland and Germany. This resource is listed in [1].
- **3.a.2.** *Quantum* (http://quantum-journal.org/, 2016, last visited March 17, 2017) is an ArXiv overlay journal for quantum physics and related fields: "all papers submitted to Quantum must be listed on (or cross-listed with) the arXiv section quant-ph. In case of acceptance, the final version must be uploaded to the arXiv before publication" (http://quantum-journal.org/about/faqpage/, last visited March 21, 2017). Little more than three months after submission opening (November 20, 2016, cfr. http://quantum-journal.org/quantum-opens-for-submissions/, last visited March 21, 2017), *Quantum* is reported to have received over 40 submissions (http://quantum-journal.org/40-submissions/, last visited March 21, 2017). In an interview to the blog "Scholastica", co-founder Christian Gogolin states that "we were strongly inspired by other arXiv overlay journals; perhaps *Quantum's* distinguishing feature is the strong emphasis on community involvement" (http://buff.ly/2k5yqUx, last visited March 21, 2017). The fourty-members editorial board is international, with a prevalence of European scientists. Accepted papers will be published under a CC BY 4.0 license and will receive a DOI through Crossref. "To provide a long term perspective for the journal, Quantum is backed by a democratic non-profit society", (the Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften based in Vienna; http://quantum-journal.org/ impressum/, last visited March 21, 2017). A subreddit has been provided for discussions, https://www.reddit.com/r/quantumjournal/ (last visited March 21,

2017); Twitter and Facebook accounts are active as well, with 983 and 793 followers respectively as at March 21, 2017. In the field of astrophysics, a single example of ArXiv-based overlay journal has seen the light up to the moment (3.a.4., *infra*), but previous, sometimes advanced efforts in this direction had been made before within this scholarly community.In a blog comment to the later experience of 3.a.4. (*infra*), Daniel Fischer witnessed that about 1997 some researchers attending a conference in Germany had already conceived the idea of creating a journal "ArXiv mated with open peer review" [...] the name that journal should be given: "Open Astronomy", but "the concept never saw the light [...]". It seems credible that the same consideration has arisen elsewhere too in the global astrophysical community; this is proved as at June 2005 among a group of young but very mindful British astrophysicists contributing to CosmoCoffee, which included Antony Lewis and Sarah Bridle (http://cosmocoffee.info/viewtopic.php?t=27 6, last visited February 8, 2017). Some years later, two relevant projects reached far more advanced, though different, stages of fulfilment and appear to be or have been very well-rooted within the astrophysical community.

- **3.a.3.** The first one was the impressing *RIOJA Project* (Repository Interface to Overlaid Journal Archives), who has been recognized as the first overlay project in astrophysics [34]. This initiative, born in 2007, was supported by prominent scholarly institutions both in the UK and in the USA: University of Cambridge, Imperial College London, University of Glasgow, UCL, Cornell University, and funded by JISC. It was preceeded by a careful examination of the side conditions inclusive of a wide survey among 683 researchers by Polydoratou & Moyle ([16], [34]), as well as by a feasibilty study [35]. A final report was also provided in 2008 [36]. Although a demonstrator implementation was achieved, as witnessed by the final report, it results that no overlay journal has subsequently been built on that technology as at September 2015. The *RIOJA Project* has been accounted for by relevant studies ([37], [28] and [29]).
- **3.a.4.** Five years later (2012), and still in a UK context, a new project was launched by cosmologist Peter Coles, *The Open Journal of Astrophysics* (http://astro.theoj.org/). The launch was made through the blogpost *A Modest Proposal – The Open Journal of Astrophysics* [38], published by Coles on his blog "In the dark" on July 17, 2012 – following previous discussions within and outside this blog. The proposal was expressed as follows: "[..] My suggestion is that we set up a quick-and-easy trial system to circumvent the traditional publishing route. The basic is that authors who submit papers to the arXiv can have their papers refereed by the community, outside the usual system of traditional journals. I'm thinking of a website on which authors would simply have to post their arXiv ID and a request for peer review. Once accepted, the author would be allowed to mark the arXiv posting as "refereed" and an electronic version would be made available for free on the website" (*ibid.*); the accepted articles are published under a CC-BY license and the reviewer comments can be disclosed "at the joint discretion of the authors and reviewers" (http://astro.theoj.org/about, last visited March 17, 2017). Coles' proposal raised interest within the community, as demonstrated by almost 70 qualified comments received by his blogpost from other scholars within the following fortnight (plus others successively). Interestingly, one of the

comments came from one of the researchers previously involved both in the mentioned lively discussion on CosmoCoffee in June 2005 and later in the Rioja project, who now is a member of OJA's editorial board (http://astro.theoj.org/about, retrieved February 8, 2017). Also, Robert Simpson (see 3.c.5. above) collaborated to the code development (https://telescoper.wordpress.com/tag/the-open-journal-for-astrophysics/, visited February 14, 2017). On 22 December 2015 it was announced that "The Open Journal is Open for Submissions" [39]; shortly after, *Nature* published an article about this initiative [40]. As of March 2017, three papers appear at http://astro.theoj.org/ as "accepted".

3.3. Other web 2.0 platforms and experiences with prevalent commenting features (3.b)

- **3.b.1.** A web 2.0 tool aimed at providing ArXiv with commenting features was *Naboj*, created in 2005 (http://www.naboj.com/, last visited February 8, 2017) and now apparently abandoned. Its name appears to be an anagram of its creator's Bojan Tunguz first name; Tunguz reports to have been "an international [physics] student and faculty at various US colleges and universities" (http://www.tunguz.com/About/, last visited March 1, 2017). The tool is described as "a dynamical website that lets you review online scientific articles. […] the [...] articles that are available for review are those that have been posted at Los Alamos ArXiv and PubMed Central". In fact, the papers commented come almost exclusively from ArXiv. The resource seems to have been used by a restricted number of people active in the physics domain: from 2005 to 2010, 23 comments were made, almost entirely on physics papers. More than 78% of the comments were made during the first two years of Naboj's existence. Rather interestingly, comments themselves could be voted as "useful" or "not useful". The last review available on *Naboj* results to have been made on February 18, 2010 (http://www.naboj.com/recent_reviews.php?s =0&np=5, visited February 8, 2017). Naboj was accounted for by [11] and [37], as well as mentioned by [8].
- **3.b.2.** *Scirate* (https://scirate.com/, last visited March 21, 2017) was originally created by US physicist and computer scientist Dave Bacon in January 2007 (http://scienceblogs.com/pontiff/2010/06/07/what-to-do-with-scirate/, dated June 7, 2010, and https://groups.google.com/forum/#!topic/scirate/WAHKx 8TAUo8, dated October 16, 2011, both last visited March 22, 2017). In early 2012 it was rewritten by Bill Rosgen (https://groups.google.com/forum/#! topic/scirate/wnjkKSZYZkI, dated April 24, 2012, last visited March 22, 2017); its code is on GitHub and user data are published under a Creative Commons license (https://scirate.com/about, last visited March 21, 2017). The information on its features appears to be synthetic on the website - at least for non-registered users ("Follow arXiv.org categories and see the highest ranked new papers; scite [i.e.: vote] papers and subscribe to categories, sign up to customize your view of the site" (*ibid.*), but the interface is rather self-explanatory. ArXiv's categories which are presently available result to be: astrophysics, condensed matter, nonlinear sciences, "more physics", mathematics, computer science, quantitative biology, quantitative finance and statistics. Users need to have registered. For a temptative assessment of its usage, the ArXiv papers which had been "scirated" at least twice during the

year from 6 April 2015 to 6 April 2016 were 53; 2 of them have been commented. The resource appears to be more widely used by mathematicians and physicists, which is probably related with what seems to be the predominant research interest within the *Scirate* working group ("the Scirate Collaboration", https://scirate.com/about, last visited March 21, 2017), i.e. quantum physics. *Scirate* is listed in [1].

- **3.b.3**. *VoxCharta* (http://harvard.voxcharta.org, 2009, last visited March 17, 2017) is somehow peculiar among the tools in this group, inasmuch it provides rating and commenting features for ArXiv papers for a practical aim: selecting papers for subsequent real-life scholarly discussions. Thus, *VoxCharta* seems somehow to bridge the gap between the two different ecosysytems of virtual and real-life scholarly communication. This might be the reason for its adoption also among research groups which have demostrated to appreciate more traditional means of internal communication such as conversations or email exchanges ([17], [18]). *VoxCharta* is self-defined as "a clone of arXiv used primarily for astronomy and astrophysics paper discussions. Users have the ability to vote for papers they would like to talk about at the next local discussion session. All papers that received votes since the previous discussion appear in an "agenda" at the top of the main page, sorted by the number of votes each paper receives […]. The basic idea is that everyone who is planning to go a department astro-ph discussion should use Vox Charta to "vote up" papers they find interesting. Additionally, each paper has a "comments" link that allows you to post things that people who are reading astro-ph may find interesting, or might be useful to look at when talking about the paper at a discussion section. Viewing the web page can be done anonymously, but voting and commenting on papers requires an account. As quite a few spammers try to sign up for accounts, each department that uses Vox Charta has a person designated as a "liaison" who approves all new accounts for that department." (http://harvard.voxcharta.org/about/about-this-website/, last visited March 7, 2017). *VoxCharta* was designed and is maintained by James Guillochon, (http://harvard.voxcharta.org/about/about-this-website/, last visited March 7, 2017), currently an Einstein Fellow at the ITC at Harvard-Smithsonian Center for Astrophysics (https://astrocrash.net/about-me/biography, last visited March 7, 2017). Thanks to the author's courtesy, we know that the first discussion took place on July 28, 2009 and that shortly after, due to other institutions' expressions of interest, the ability for the site to support multiple institutions simultaneously was added. The original number of ArXiv categories was gradually extended including, e.g., high energy physics. *VoxCharta* is listed in [1].
- **3.b.4.** Another prominent experience is *PaperRater* (http://www.paperrater.org/, last visited March 17, 2017), created by young German astrophysicist Peter Melchior in 2010 and listed in [1]. PaperRater's Getting Started Guide states the tool's fundamentals: "PaperRater.org reads the daily submission to any category of arXiv and searches for published papers onThe SAO/NASA Astrophysics Data System (ADS) […]. You can help PaperRater.org to help all of its users [...] by rating, tagging or commenting papers. You can rate every paper only once, but you can change the rating later at any time. Your rating is anonymous. The distribution of ratings will be shown once a sufficient number of ratings is reached. You can

add as many tags to each paper as you like, but three is often a good number. These tags can be updated at any time. [...] No other user can find out, which papers you rated or even what your rating was, nor what tags you chose. In contrast, comments are meant to be public. If you […] decide that you want to stay anonymous [...] you can choose to do so for any comment independently" (http://www.paperrater.org/help/getting-started.html, dated March 3, 2012, last visited February 28, 2017)."Sharing has been enabled with Twitter, Facebook, Google+, and LinkedIn (http://paperrater.org/blog/social-bookmarking-and-altmetrics.html, dated August 20, 2012, last visited February 28, 2017). PaperRater's interface looks user-friendly and the tool's mission is clearly stated in the first post of the dedicated blog (October 8, 2010): "The peer review process has a long-standing tradition in improving manuscript quality [...] However, it is not infallible [...] as students and researchers we all read papers daily, evaluate and judge them [...] this process is able to improve a paper's quality beyond what a single referee could achieve. If the joint wisdom of the community could be bundled. This is what PaperRater.org is all about: to augment and eventually replace the intransparent process of peer review as a lone quality measure for publications by a public one" (http://www.paperrater.org/blog/mission-statement.html, last visited March 17, 2017). In March 2016 the author's kindness made it possible to give some figures of users' response to *PaperRater* over time. Reads had increased significantly from 2010 (1467) to 2012 (2964), starting then to decrease (678 in 2013) until the last year available (363 in 2015). Ratings had reached a maximum during the first year (111), were 92 in 2012 and decreased markedly after 2013 (when they were 20). Registered users were 558 - as at March 20, 2016.

- **3.b.5.** The idea of *YouASTRO* (http://youastro.dyndns.org:43905/, last visited March 17, 2017) came during a post-conference international evening colloquium among astrophysicists – as kindly reported from project co-creator, Italian astrophysicist Fabrizio Bocchino (Italian National Institute for Astrophysics), who wrote the YouASTRO code. The other involved researchers were Javier Lòpez-Santiago, Juan F. Albacete-Colombo and Niccolò Bucciantini. The tool was operative in 2011, but some comments to an article published in May 2010 can date the tool back to the year before. The project was presented to the ADASS conference in November 2011 [41]. The definition on the website states that "YouASTRO is a web application which allows us to leave comments and give rating to refereed astrophysical papers. For now, the papers which can be commented are only the papers appearing on the SAO/NASA Astrophysics Data System [i.e. the widest database for the astrophysical literature worldwide] […] The YouASTRO Board of Editors think that the YouASTRO "leave a comment" feature can be of great benefit to the scientific community, if used widespreadly. It promotes the online scientific discussion focussed on papers, it is a way to pinpoint strong and weak points of papers, in the framework of a general and continuous improvement of the quality of scientific publications, and the overall advance of science" (http://youastro.dyndns.org/faqs.html, last visited March 17, 2017). Suitable consideration was had both for authors' sensibility and for statistical significance: "registered users can vote a paper, one vote per paper [...] rating goes from 1 (very poor) to 10 (excellent). Ratings are always anonymous [...] YouASTRO only shows average ratings [...] after more than 3 ratings have been received". The placement in the web 2.0 ecosystem occurs clearly, but the focus results to be on published articles rather than on preprints (among other clues: "comments to astroph papers will be automatically migrated to the refereed version (...) when it appears"). As at June 2016, *YouASTRO* had 434 registered users (were 100 on 20.12.2011, http://youastro.dyndns.org/news.html, visited July 4, 2016). Peaks of activity were achieved during the first years of operativity, as witnessed by the data from two public tables of YouASTRO top-ranked papers (http://youastro.dyndns.org:43905/#highest, last visited March 17, 2017). Among the top 10 most commented papers, 10% were published in 2009 and in 2012, 60% in 2010, 20% in 2011; 64.28% of the public comments they received were made in 2010, 21.42% in 2011. 70% of top 10 most recently commented papers were published in 2013, none afterwards; 85.70% of their public comments were equally divided between 2012 and 2013. Public comments result to be only 34.69% and, among them, anonymity is the standard (92,08%), as understandable in a small community of users.

- **3.b.6.** Presumably at the beginning of 2012, young mathematician Ralph Furmaniak, a PhD student at Stanford University by that time, created *ArXaliv*. When publicising his tool on a forum for colleagues on March 28, 2012, Furmaniak wrote "I have set up the reddit software to work with the arxiv database [..] Each day it will update the list with the latest papers and you can upvote, downvote, comment, save links of interest, search, post new links, or create your own communities/arxalivs to post in or have others post links or writings of interest to them. [...]" (http://publishing.mathforge.org/discussion/ 83/, last visited March 17, 2017). Exactly one year later, Furmaniak posted ArXaliv's codebase on GitHub in case "one day [...] there are other people interested" (https://github.com/rfurman/arxaliv, last visited March 17, 2017). In fact, the tool looked "defunct" to another mathematician on a blogpost dated November 12, 2013 and is presently no more available at the original website http://arxaliv.org/.Also to the mathematic field and to 2012 seems to have belonged the project of *arXiv Review* (no more available at http://arxiv-review.org/ as at March 2017). Apparently, it was intended as an ArXiv overlay journal with commenting and rating features. Related documentation can be found at http://occupypublishing.blogspot.it/2012/02/guidelines-for-arxiv-review.html and https://plus.google.com/113026609770667182181, last visited March 20, 2017.
- **3.b.7.** *Selected Papers* (https://selectedpapers.net/) was developed in 2013 by US computational biologist Christopher Lee (see the post https://johncarlosbaez.wordpress.com/2013/06/07/the-selected-papers-network-part-1/ on US mathemathical physicist John Carlos Baez' blog "Azimuth", last visited 27 February 2017; see also Lee's blogpost https://johncarlosbaez.wordpress.com/2013/07/12/the-selected-papers-network-part-3/ dated July 12, 2013, last visited 27 February 2017, and [42]). This tool – listed in [1] –, which enabled commenting on ArXiv papers, had distinctive features among which using Google+ authentication and seems to have raised interest among researchers. In March 2016, anyway, *Selected Papers* resulted to be unaccessible, which remains unchanged at the moment

of writing, although a detailed documentation about this project is still available at http://docs.selectedpapers.net/ (last visited March 17, 2017). Due to this situation, and to Lee's specific research area, a more in-depth account of this resource won't be provided.

- **3.b.8.** *Xiv* (https://www.reddit.com/r/Xiv/, last visited March 17, 2017) is "an interdisciplinary reddit for discussing papers submitted to arXiv (http://arxiv.org/), an open-access journal for e-Prints." It "aims to support arXiv by providing an open forum for papers and by calling attention to great papers" (https://www.reddit.com/r/Xiv/, last visited March 7, 2017). Registered users – who result to be 431 as at March 7, 2017 - can submit text posts or arXiv abstracts, and may receive comments from other registered users. Deductively, *Xiv* made its debut in 2014. As at March 7, 2017, posts – which can be upvoted – result to be 47, 41 of which were published in 2014, 2 in 2016, 4 in 2017; 53.19% of them received one or more comments. Thirty-eight posts (80.85%) have a tag and these are in many subfields of physics, included astrophysics, though the great majority are in quantum physics. There are two moderators, who appear to be active in quantum physics; only their nicknames are available and apparently they can't be contacted by non-members. It can be noted that Reddit hosts further relevant subreddits, e.g. in cosmology and in astronomy, but the discussions don't appear to be based upon ArXiv papers.
- **3.b.9.** *ArXiv Analytics* (http://arxitics.com/, last visited March 17, 2017) was developed in 2014 by Chinese graduate student on high energy physics Zan Pan (Institute of Theoretical Physics, Chinese Academy of Sciences of Beijing), who is also the resource's maintainer. Collaboration and feedbacks were gained also from other nations (https://github.com/arxitics/arxiv-analytics/network /members, last visited March 6, 2017). This resource is defined as "a web portal that offers more features and a better user interface for reading eprints provided by arXiv.org. You can search, subscribe, bookmark, review eprints, and interact with the community. The project is still under development." (http://arxitics.com/site/about, last visited March 1, 2017). ArXiv Analytics' main functions appear to be: "advanced search interface to find articles" (includes sorting by "reader counts" or by "rating score"); configure eprint subscriptions - by several parameters including keywords, tags, authors; manage one's preferences/activities in a personal account (e.g. bookmarks, reading, rates, votes); post reviews and make comments; openly upload one's original content that have not been published online (under CC BY-SA 4.0 license; all from http://arxitics.com/, visited March 1, 2017), thus gaining twenty "reputation points" for each document (http://arxitics.com/help /documents, visited March 1, 2017). The reputation system (http://arxitics.com /help/reputation, last visited March 1, 2017) shows some apparent oddity such as losing reputation points when rating an article or voting a review (-1 in each case, but +5 for publishing a new review); this might be due to a value system that encourages sharing significantly (+20 for sharing a document) rather than judging on a small scale. Thanks to Zan Pan's courtesy we get to know that there are 295 registered users at present, many of which are Chinese students; for them, *ArXiv Analytics* also provides a chat. The number of rated papers is presently "less than 100" (the feature is still experimental).

- **3.b.10.** Another tool which appears to have been tailored upon ArXiv in a web 2.0 environment was *ArQuiv* (http://arquiv.org/), which was presumably born in 2014. It was retrieved and visible on March 23, 2016 but is no more available as at writing (March 2017). Anyway, even more than it happens with other similar tools, the information supplied on the website was poor for those not registered, so that for example it was impossible to credit ArQuiv to its authors otherwise than "arQiv.org has no affiliation with arXiv.org or Cornell University" – and the homepage description was limited to: "arQiv.org: revolutionize scientific discussion by connecting readers and authors. To discuss any arXiv article, just change **"X"** to **"Q"** to visit arQiv". One of the ideas seemed to be to modify the typical url of an ArXiv paper in order to enable comment reading. *ArQuiv*, anyway, clearly belonged to the family of platforms aimed at supplementing ArXiv with web 2.0 features.
- **3.b.11.** In 2015 young physicist Florian Beutler and cosmologist Morag Scrimgeour created *Benty Fields* (http://www.benty-fields.com/, last visited March 17, 2017). The resource is described as "the academic network with daily papers and journal club organizer" (*ibid.*). In fact it's more than this as it "allows you to read the daily arxiv publications in a user friendly environment [...] You can organize papers in a library, including a reading list. If you are member of a journal club, you can directly vote for papers to put them on the agenda for the next journal club. If you want to point out a paper to a colleague you have the Recommend Paper option. You can create a journal club and invite your colleagues to join. The journal club agenda shows all the papers voted for by members of your journal club. Under My Profile you can let others know about your academic career and interests. [...] Benty-fields is organized like a social network, so you can follow your colleagues and they can follow you. The social network aspect is still under development, but already there are advantages when following others. For example you can easily contact them, send them messages or read their profile" (http://www.benty-fields.com/, as retrieved on April 14, 2016). A remarkable characteristic is the tool's social networking feature, which definitely locates this experience in the post-social networks era. Consistently, *Benty Fields* appears to be integrated with Facebook, Twitter, Google+ and LinkedIn. Registration is required. The interface is agreeable and the tool is sophisticated enough to provide a section about Terms and conditions as well as a privacy policy (http://www.benty-fields.com/tos#priv, last visited March 17, 2017).

4. Conclusions

The availability of an established and comprehensive database of open access literature in physics and astrophysics such as ArXiv is likely to have fostered the birth of a significant number of web 2.0 experiences in these research fields and may have shaped them as electively literature-based. This seems to have happened rather early in some cases and anyway independently from ArXiv's adoption of a web 2.0 setting.

In this respect, the vision of ArXiv as a founding ground for physicists' accreditation within their community results to be appropriate, not so much as the elegant socio-hermeneutical proposal of a database having a legitimizing role for itself among physics researchers [19] but rather, in addition to the sanctuary, like a function of early repository for the discipline's literature – as a lively catalyst for web 2.0 scholarly exchanges within astrophysics and physics.

On the basis of the 2016 users survey and analysis conducted, the ArXiv team appears now to be somehow mediating between the "conservative" and still prevailing attitude, focussed on keeping the platform "to the core mission", and an emerging 2.0 trend which favours innovations such as rating and commenting on top of it ([25], [26]). The ArXiv-Next Generation initiative, which has only just started ([26]), will perhaps mark a change in this respect, for as much as it's possible to understand at the moment.

As for the tools here surveyed, and again for as much as it has been possible to observe, the outcomes appear to have been often affected by the physical limits of the local circles involved. For example, it has been found repeatedly that researchers committed to a project didn't know about the existence of parallel efforts among other colleagues, or that the news about a project's development didn't circulate well enough among interested people outside the circles – as witnessed by blog comments. An apparently rare piece of research about extending ArXiv's features to open peer-review and publishing [43] doesn't mention any of the ArXiv-based commenting resources for scholars which were already in place by that time according to our findings. All this testifies that, although obviously internet-based, many of these experiences were in fact very local level-dependent, at least during the first years of their existence. All in all, actually, web 2.0 tools in astrophysics seem to have been strongly affected by local circumstances, both for the good (e.g. motivation) and for the bad – restricted scholarly communities can seldom provide the critical mass for a new tool to take off, especially when validation is involved.

For a significant part, the web 2.0 tools which have been accounted for above appear to have been created in a few astrophysical circles, mainly concentrated in the UK and in the USA; specially lively environments have proven to be the University of Sussex and Harvard University. Following the academic pathway of some of the creators of these tools, who sometimes were foreign students or researchers, might contribute to the history of web 2.0 commenting platforms in astrophysics. This anyway goes beyond the aim of the present study and is probably more appropriate for retrospective future research.

There are clues that this aspect, and the common local perspective, might be changing in the latest years – approximately starting around 2012, e.g. with a stronger presence of multi-national development teams. This might have to do with the diffusion of worldwide sharing platforms such as GitHub, although this is a simple hypothesis. 2012 also seems to be the peak of one of the time flows in which the experiences surveyed seem to have debuted – which is in line with Peter Melchior's remark as expressed in a comment to mathematician Philip Thrift's blogpost ("the internet seems to be bursting these days with ideas about how to improve/replace peer review and classical journal. This is a very exciting time. [...]", http://occupypublishing.blogspot.it/2012/02/scientific-journals-in-e-publishing-age.htm dated February 1, 2012; last visited March 20, 2017).

On the whole, the astrophysical and physical communities look globally far from unaware or idle with regard to new communication and validation tools in the 2.0 ecosystem.

Acknowledgements

The author would like to thank (in alphabetical order): Fabrizio Bocchino, Peter Coles, Christian Gogolin (in the name of the Executive Board of *Quantum*), James Guillochon, Antony Lewis, Peter Melchior, Martin Moyle, Zan Pan, Panayota Polydoratou for the information provided.

References

[1] B. Kramer and J. Bosman, 400+ Tools and innovations in scholarly communication (2015), available at: http://bit.ly/innoscholcomm-list.
[2] S. Gass, Transforming scientific communication for the 21st century, *Science and Technology Libraries* **19** (2001), 3–18.
[3] P. Ginsparg, After Dinner Remarks, 14 Oct '94. *APS meeting at LANL* (1994), available at: http://www.cs.cornell.edu/~ginsparg/physics/blurb/pg14Oct94.html .
[4] P. Ginsparg, First Steps Towards Electronic Research Communication, *Computers in Physics* **8** (1994), 390-396.
[5] M.A. Rodriguez, J. Bollen, H. Van de Sompel, The Convergence of Digital-Libraries and the Peer-Review Process, *Journal of Information Science* **32** (2006), 149–159.
[6] T. O'Reilly, What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software (2005), avalable at: http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html .
[7] P. Ginsparg, Can peer review be better focused? *Science & Technology Libraries* ***22*** *(2002)*, 5–17.
[8] D. Taraborelli, Soft peer review: social software and distributed scientific evaluation, in: *Proceedings of the 8th International Conference on the Design of Cooperative Systems (COOP '08), Carry-Le-Rouet, May 20-23, 2008,* P. Hassanaly, A. Ramrajsingh, D. Randall, P. Salembier and M. Tixier, eds, Institut d'Etudes Politiques d'Aix-en-Provence. Aix-en-Provence, pp. 99-110. Available at: http://discovery.ucl.ac.uk/8279/1/8279.pdf .
[9] A. Birukou, J. Rushton Wakeling, C. Bartolini, F. Casati, M. Marchese, K. Mirylenka *et al.,* Alternatives to Peer Review: Novel Approaches for Research Evaluation. *Frontiers in Computational Neuroscience* **5** (2011). Available at: http://doi.org/10.3389/fncom.2011.00056 .
[10] M. Nentwich, R. Koenig, *Cyberscience 2.0. Research in the Age of Digital Social Networks,* Campus Verlag, Frankfurt-am-Mein, 2012.
[11] P. Dall'Aglio, Peer-review and journal models (2006). Available at: arXiv:physics/0608307v1 .
[12] M. Baez, A. Birukou, R. Chenu, M. Medo, N. Osman, D. Ponte et al. *State of the Art in Scientific Knowledge Creation, Dissemination, Evaluation and Maintenance* (Tech. Rep. # DISI-09-067). University of Trento, Department of Information Engineering and Computer Science, Trento, Italy, (2009). Available at: http://eprints.biblio.unitn.it/1773/ .
[13] I. Rowlands, D. Nicholas, B. Russell, N. Canty and A. Watkinson, Social media use in the research workflow. *Learned Publishing* **24** (2011), 183–195.
[14] A. Fresco-Santalla and T. Hernández-Pérez, Current and evolving models of peer review, *The Serials Librarian* **67** (2014), 373–398.
[15] C. Sugimoto, S. Work, V. Larivière and S. Haustein, Scholarly use of social media and altmetrics: a review of the literature (2016). Available at: https://arxiv.org/abs/1608.08112 .
[16] P. Polydoratou and M. Moyle, Exploring aspects of scientific publishing in astrophysics and cosmology: the views of scientists, in: *Metadata and Semantics. Post-proceedings of the 2nd International Conference on Metadata and Semantics Research, MTSR 2007, Corfu Island in Greece, 1-2 October 2007,* M.-A. Sicilia & M. D. Lytras, eds, Springer, New York, 2007, pp. 179–190.
[17] D. Harley, S.K. Acord, S. Earl-Novell, S. Lawrence, and C.J. King, (2010). Chapter 3.: Astrophysics case study, in: *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines,* Center for Studies in Higher Education, Berkeley: UC Berkeley, 2010, pp. 138–206.
[18] M. Marra, Professional social networks among Italian astrophysicists. Prospective changes in validation and dissemination practices? *Information Services and Use* **35** (2015), 243–249.
[19] A. Delfanti, Beams of particles and papers: How digital preprint archives shape authorship and credit, *Social Studies of Science* **46** (2016), 629–645.

[20] S. Haustein, T.D. Bowman, K. Holmberg, I. Peters and V. Larivière, Astrophysicists on twitter: An in-depth analysis of tweeting and scientific publication behavior, *Aslib Journal of Information Management* **66** (2014), 279–296.
[21] S. Haustein, T.D. Bowman, B. Macaluso, C. Sugimoto and V. Larivière, Measuring Twitter activity of arXiv e-prints and published papers, in: *Altmetrics14 an ACM Web Science Conference,* 2014, pp. 138–206.Available at: http://dx.doi.org/10.6084/m9.figshare.1041514 .
[22] K. Holmberg, T.D. Bowman, S. Haustein and I. Peters, Astrophysicists' Conversational Connections on Twitter. *PLOSone* **9** (2014), e106086. Available at: http://doi.org/10.1371/journal.pone.0106086.
[23] S. Ritson, "Crackpots" and "active researchers": The controversy over links between arXiv and the scientific blogosphere. *Social Studies of Science* **46** (2016), 607–628.
[24] P. Ginsparg, ArXiv at 20, *Nature* **476** (2011), 145–147.
[25] O.Y. Rieger, G. Steinhart and D. Cooper, ArXiv@25: Key findings of a user survey (2016). Available at: http://arxiv.org/abs/1607.08212 .
[26] O.Y. Rieger, J. Entwood, C. McLaren, S. Payette and G. Steinhart, arXiv Update - January 2017. Available at: https://confluence.cornell.edu/display/arxivpub/arXiv+Update+-+January+2017 .
[27] J.W.T. Smith, The deconstructed journal: a new model for academic publishing, *Learned Publishing* **12** (1999), 79–91.
[28] J. Brown, An introduction to overlay journals (2010). Available at: http://discovery.ucl.ac.uk/19081/1/19081.pdf .
[29] J. Priem and B.H. Hemminger, Decoupling the Scholarly Journal. *Frontiers in Computational Neuroscience* **6** (2012).
[30] J. Cartwright, Fledgling site challenges arXiv server, *PhysicsWorld.* Available at: http://physicsworld.com/cws/article/news/2009/jul/15/fledgling-site-challenges-arxiv-server .
[31] P.E. Gibbs, A good year for ViXra. *Prespacetime Journal* **4** (2013), 87–90.
[32] R.J. Knegjens, Paperscape. Available at: http://robjk.net/paperscape .
[33] S.M. Hitchcock, *Perspectives in Electronic Publishing: Experiments with a New Electronic Journal Model*. PhD thesis in Engineering And Applied Science, University of Southampton, 2002. Available at: http://eprints.soton.ac.uk/258233/2/prelims.pdf .
[34] P. Polydoratou and M. Moyle, Exploring overlay journals: the RIOJA project [slides of the talk], in: *OAI5. CERN workshop on Innovations in Scholarly Communication*. Geneva, 2007 (online). Available at: http://indico.cern.ch/event/5710/contributions/1212287/attachments/988173/1405151/Presentation_OAI5_Polydoratou.pdf
[35] P. Polydoratou and M. Moyle, Scientific journals, overlays and repositories: a case study of costs and sustainability issues, in: *Digital Libraries: Universal and Ubiquitous Access to Information, 11th International Conference on Asian Digital Libraries, ICADL 2008,* G. Buchanan, M. Masoodian and S. J. Cunningham, eds, Springer, Berlin, 2012, pp.154-163.
[36] M. Moyle and A. Lewis, *RIOJA (Repository Interface to Journal Archives) Final Report*. 2008. Available at: http://discovery.ucl.ac.uk/12562/1/12562.pdf .
[37] M. Cassella and L. Calvi, New journal models and publishing perspectives in the evolving digital environment, *IFLA Journal* **36** (2010), 1–12.
[38] P. Coles, A Modest Proposal – The Open Journal of Astrophysics | In the Dark on WordPress.com, July 17, 2012, available at: http://telescoper.wordpress.com/2012/07/17/a-modest-proposal-the-open-journal-of-astrophysics/ .
[39] P. Coles, The Open Journal is Open for Submissions, | In the Dark on WordPress.com, December 22, 2015, available at: https://telescoper.wordpress.com/2015/12/22/the-open-journal-is-open-for-submissions/ .
[40] E. Gibney, Open journals that piggyback on arXiv gather momentum. *Nature* **530** (4 february 2016), 117–118.
[41] F. Bocchino, J. Lopez-Santiago, F. Albacete-Colombo and N. Bucciantini, YouASTRO: a web-based bibliography management system with distributed comments and rating features for astrophysical ADS papers, in: *Astronomical Data Analysis Software and Systems XXI***,** P. Ballester, D. Egret and N. Lorente, eds, Astronomical Society of the Pacific. San Francisco, 2012, pp. 889-892. Available at: http://www.aspbooks.org/a/volumes/article_details/?paper_id=34807.
[42] C. Lee, Open peer review by a selected-papers network, *Frontiers in Computational Neuroscience* **6** (2012). Available at: http://dx.doi.org/10.3389/fncom.2012.00001.
[43] A. Boldt, Extending ArXiv.org in order to achieve open peer review and publishing, *Journal of Scholarly Publishing* **42** (2011), 238–242.

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-118

The Transition to Open Access

The State of the Market, Offsetting Deals, and a Demonstrated Model for Fair Open Access with the Open Library of Humanities

Martin Paul EVE[a1], Saskia C.J. DE VRIES[b] and Johan ROORYCK[c]
a. Birkbeck, University of London and the Open Library of Humanities
b. Sampan – academia & publishing and Linguistics in Open Access
c. Leiden University and Editor in Chief of Glossa

Abstract. In this article, we explore the state of the OA market and the current situation with respect to offsetting deals in the Netherlands. We then offer a case study of the LingOA model for a transition to open access, backed by a consortial funding mechanism: the Open Library of Humanities (OLH). We also suggest how this approach can be extended into new disciplinary spaces (in particular, mathematics and psychology, where there is already some willingness from editors).

Keywords. Fair Open Access Publishing, APCs, Open Library of Humanities, Flipping existing subscription journals, Sustainable model for scholarly communication.

1. Introduction

In May 2016, the European Council of Ministers set an ambitious goal: to make open access (OA) to scientific publications the default by 2020.[1] This bold proposal to make peer-reviewed research available freely to read and re-use online, though, came with few indications as to how it would actually be achieved and, thus far, the road to OA has been winding and bumpy. Further, at least in this transition period, the total cost of academic publishing has been raised by hybrid journals and a lack of effective offsetting measures.[2,3] The concentrating effect of new business models, such as Article Processing Charges (APCs), have made it clear that any transition will also require careful thought about economic distribution.[4] A range of proposals and approaches are currently in simultaneous development in order to achieve what was

1 Corresponding author, Department of English and Humanities, Birkbeck, University of London, United Kingdom. Email: martin.eve@bbk.ac.uk.

thought to be a singular goal, from OA offsetting deals with large, existing publishers through to models for global "flips" to OA.[5,6]

In this article, we explore the state of the OA market and the current situation with respect to offsetting deals in the Netherlands. We then offer a case study of the LingOA model for a transition to open access, backed by a consortial funding mechanism: the Open Library of Humanities (OLH). We also suggest how this approach can be extended into new disciplinary spaces (in particular, mathematics and psychology, where there is already some willingness from editors).

2. The State of the Open Access Market

Scholarly communications environments have long been deemed strange when compared to other market environments. For instance, Peter Suber has noted that "[e]very scholarly journal is a natural mini-monopoly in the sense that no other journal publishes the same articles", which makes the possibility of substitute goods and effective price pressure mechanisms extremely difficult to introduce.[7] Further, Martin Paul Eve has shown how a symbolic economy of prestige maps onto the material library economy and leads to a lack of price sensitivity among researchers.[8] It is these features, combined with the mass expansion of higher education (and, therefore, research volume), that have led to the *circa* 300% rise above inflation in required serials expenditure since 1986 charted by the ARL.[9]

A recent report commissioned by OpenAIRE on behalf of the European Commission, 'Towards a Competitive and Sustainable OA Market in Europe', thus defines the scholarly publishing market as "an 'intermediated market'", worth approximately ten billion USD per year "with researchers acting as both producers and consumers of research, while the purchase of content is typically undertaken by academic libraries".[10] As before, this leads to the situation that "weakens the price sensitivity of consumers" on both author and reader fronts.[10] While the OpenAIRE report notes that, in 2015, annual revenue from English-language science, technology and medicine (STM) journals was estimated at between $7 and $10 billion, the broader STM information publishing market is worth around $26 billion of which roughly 55% comes from the US and 28% from Europe according to a report using data from Simba.[11] This economic environment, which draws mostly on university library budgets, is compounded by payments to national copyright agencies for reproduction rights.

The OpenAIRE report identifies four paths for open access to scientific research (where "scientific" is used in the European sense to span many more disciplines than the Anglo-American usage of the term), splitting the well-known green/gold distinction into a range of sub-components that specify business models:

- Green open-access archiving: usually an author's accepted manuscript, sometimes with an embargo, within an institutional or subject repository
- Hybrid gold open access: peer-reviewed articles inside subscription/toll-access journals are made immediately open access, by the publisher, often upon payment of an APC. This can be achieved either directly or through an offsetting arrangement (a kind of "big deal" for open access)

- Gold open access via APCs: fully open-access journals that require a payment from an author, institution, or funder
- Gold open access without APCs: fully open-access journals that require no payment and have alternative business models in place (sometimes also called: "platinum OA").

This fresh characterization of the modes is useful, since the initial terms were never supposed to specify the underlying business models.[12] Yet, it is also clear that the redistribution proposed by article-processing-charge based models comes with new challenges.

These distributional effects can be considered through an allegory.[13] Consider, for example, that there are 100 people in a room. They have $10 each. An academic speaker will give the audience a talk but the venue wants $50 to cover its costs (and any profit/surplus). There are 40 such talks per year. There is, finally, an indefinitely large group of people (let us call them "the general public") who might want to hear the talk but who can't afford to pay anything.

The subscription logic would be: each person pays $0.50 and gets access to the talk. If a person does not pay, s/he/they may not hear the talk. This logic is implemented to introduce a classical economic system. With the funding available, each person can choose to attend this talk or another. However, each of the 40 talks is different and doesn't cover the same material. The attendees do not really know whether a talk will be useful to them in advance. They can attend 50% of the talks. This model spreads costs but limits access; 50% of the talks could be attended by 100% of the attendees but nobody from the "general public" group gets to hear the talks. Further, it is unlikely that all 100 participants will attend the same 40 talks, so knowledge of the talks' contents is diffuse. Some believe this is the best way of ensuring the venue is compensated and remains open for talks because it incentivizes people to pay. The speaker doesn't necessarily get the largest possible audience from this model.

The logic of an APC would be: the speaker will pay the venue's cost of $50 and let anybody hear the talk for no charge. This makes sense to the academic as her only motivation is to be heard (she is one of the lucky ones who has an academic post). The problem is that she only has $10 herself. This model concentrates costs (sometimes impossibly so) but allows the theoretically widest access. In this particular case, though, an idealised logic led to no access since no single individual can afford the total cost. APCs have a problem of the current distribution of resources.

Finally, the logic of new consortial OA funding mechanisms such as Knowledge Unlatched and the Open Library of Humanities would be: 5 people attend each talk. They each spend their full allowance of $10 on that single talk. However, they let everybody else attend any talk for which they have paid, in expectation of reciprocity and for the public good. They record the talk and let others view this for no charge. This model spreads costs and allows broader access than the subscription model; 50% of the talks could be heard by not only 100% of the attendees but also by the group who can't afford to pay. This is the logical choice for those present but some are worried that they may pay while others might not return the favour.

There are also arguments that the $50 venue fee is extortionate, since it appears that 35% of it ($17.50) is pure profit for the venue organization, which is in fine financial health. Some point out that were this closer to 6% ($3.00) the organization would still be fine and could pay all its staff but each talk would only cost around $35. At that rate, it would be possible to host approximately 29 of the planned talks and, with the distribution in the different models, allow other groups to have access.

The current state of the open-access market, though, is a mix of these different types of logic, all adding costs on top of one another. The proliferation of author/institution-facing charges has led to a need to see reductions in the current level of subscription expenditure, even though not all players in the global ecosystem are moving to gold OA via APCs at the same rate. This has meant that a series of "offsetting deals" have come into play.

3. Open Access Offsetting Deals

In the hybrid gold environment described above, researchers and institutions are able to purchase open-access for specific articles, even while subscriptions continue to be charged for the journal. This has led to the accusation that publishers are "double-dipping"; that is, charging twice for the same material through both subscriptions and APCs.[14] Representatives of large publishers like Elsevier, such as Alicia Wise, have denied that they double-dip, though, stating in a kind of double-think that "there is no connection between subscriptions and APCs: they are 'decoupled'. She says the money coming in through a journal subscription is used to pay for a particular number of articles, and that open-access articles in hybrid journals are additional to that".[15,16]

Nonetheless, as APCs have grown to consume more and more of library resources, it has become necessary to find ways to "offset" subscription expenditure against gold payments (whether hybrid or pure).[2,17] The OpenAIRE report identifies four different types of offsetting arrangements that have been put into place:

- A local reduction from a subscriber's fees of the total amount of all APC revenue from the previous year
- A cap, whereby subscriptions are maintained, but subscribing organizations pay no extra to have all their own outputs made openly available
- An APC discount, sometimes of up to 95% of the standard APC, for authors at subscribing institutions
- A voucher system equivalent to a subscription spend to be used on APCs.[18]

One country that has been actively monitoring the effects of these deals, aside from the already-cited sources in the UK, and from which we can glean some knowledge is the Netherlands.

Agreements between Dutch university libraries and traditional academic publishers with open-access options have been actively monitored in the country since 2015.

There has also been an active effort to collate data on the costs incurred per-university, per-publisher using the Government Information (Public Access) Act.[19]

A September 2016 request to the VSNU (the association of Dutch universities) under this legislation asked for “provision of a copy of the open access licenses purchased by your institution in the past year from various publishers such as Elsevier, Springer, Wiley, Taylor & Francis, ACS, SageKarger, Thieme, Walter de Gruyter, RSC, Emerald and any comparable licenses, with the essential understanding that the institution shall pay a previously-established fee to the publisher, in exchange for which the publisher will publish accepted academic articles by authors affiliated with your institution open access in licensed journals”.[20] This request revealed a large variance in spending, with the largest share going to Wiley at €3,818,000, with Taylor and Francis a close second at €2,318,584.

Perhaps most interesting, though, was the deal struck in more recent days between Elsevier and Dutch Universities. The paragraph around open-access provision in the leaked details of this deal paint a picture of a large organization attempting to hinder progress towards OA, at least in some interpretations. As Sicco de Knecht has put it:

> “The agreement draws a disheartening picture of the so called ‘Golden deal’ reached by the Dutch universities with their major publisher: Elsevier. Hindered by severe restrictions only Dutch corresponding authors from the combined institutions are eligible to publish in a very select set of journals in the Elsevier collection.
>
> Simultaneously Elsevier raises its collective fees in 2017 and 2018, with 2.5% and 2.0% respectively from the level of €11,697,147.68 in 2016. The contract also states that Elsevier will not levy publication charges to authors. This is included in the price of the deal which has been raised by the publisher to cover the lost revenue. At the end of the contract period parties will decide whether the ‘experiment’ was worth their while.”[21]

While these models retain the distributional characteristics of the subscription environments, one of the primary concerns about such setups is that they also perpetuate lock-in. That is, because the libraries have already agreed to pay this group of publishers, it becomes difficult for new actors to mount any substantial market challenge. What is clear, though, is that these deals are becoming more prevalent; a fact that we attribute to a desire for the distributional regime of subscriptions within an open-access environment. For the final section of this paper, we turn to a project description of the Fair Open Access model; the LingOA project; and the Open Library of Humanities platform, that we believe retain the desired characteristics of subscription funding while accruing the benefits of open access, within a competitive price framework.[22]

4. The Fair Open Access Model

In the Linguistics in Open Access (LingOA) model, several international linguistics journals have recently moved from their traditional publisher to a new open access publisher, moving their entire editorial staff, authors, and peer reviewers from the traditional subscription model to a model that we call "Fair Open Access". The Open Access publisher has to comply with the following conditions, a.k.a. the Fair Open Access Principles:

1. The journal has a transparent ownership structure, and is controlled by and responsive to the scholarly community.[23]

2. Authors of articles in the journal retain copyright.[24]

3. All articles are published open access and an explicit open access license is used.[25]

4. Submission and publication is not conditional in any way on the payment of a fee from the author or its employing institution, or on membership of an institution or society.[26]

5. Any fees paid on behalf of the journal to publishers are low, transparent, and in proportion to the work carried out.[27]

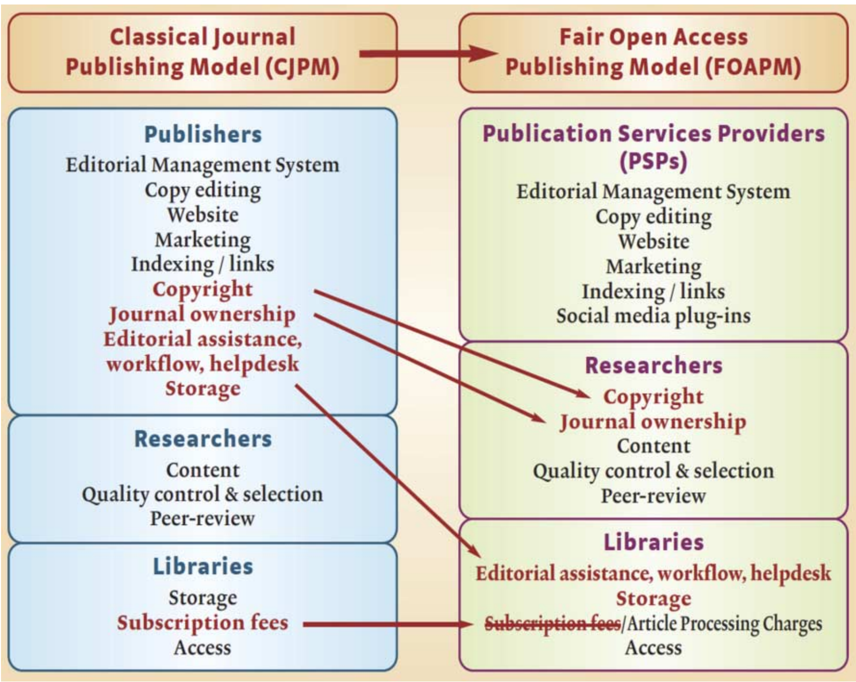
LingOA facilitates this radical move by paying for the article processing charges (APCs) of the articles published in these journals during the first five years as a transition measure. The journals *Glossa*, *Laboratory Phonology*, and the *Journal of Portuguese Linguistics* are now published by Ubiquity Press. The *Italian Journal of Linguistics* joined LingOA early 2017. The publisher of this last journal, Pacini in Italy, complied with the Fair Open Access conditions, so this journal stayed with its publisher. To ensure long-term sustainability, LingOA has partnered with the Open Library of Humanities (OLH). OLH will guarantee the continued publication of the journals associated with LingOA after the first five years through its consortial library funding model, provided that its Library Board votes in favour through its journal selection procedures. OLH is a charitable organization dedicated to publishing Open Access scholarship with no author-facing APCs. This will provide long-term sustainability for Fair Open Access journals, ensuring that no researcher will ever have to pay for APCs out of his or her own pocket.

Because the community of linguists is relatively small and close-knit, this created nearly ideal conditions for the project. Once the transition of the first journals – including all of their editors, editorial board, peer reviewers, and authors – had been completed, we immediately found that the linguistics community realized that the journals have essentially remained the same. Authors and readership was not affected, even in case of a name change. Citation-, ERIH- and H-indices – important quality metrics in this field – are therefore not expected to decline, or will at least recover quickly.

The editorial board at the Elsevier journal, *Lingua*, left the publisher to establish a new journal. In its first year, 2016, this new journal, called *Glossa*, has had 319 articles submitted, 51 published and 54 in production. In 2017, 80 articles have already been submitted and 27 published, with 55 in production. Thus, the editorial board saw no negative repercussions from their move.

After the successful transition of these journals, LingOA hopes to convince the editors of many other prominent linguistics journals to join them. As such, LingOA expects to become a model for the transition to Fair Open Access in other disciplines as well. Various editors of journals in for example Mathematics and Psychology have already expressed their interest in flipping their journal to fair open access. Such an approach will require up-front funding, both for the initial transition period, and to expand consortia like the Open Library of Humanities to ever-larger sizes to facilitate the wide distribution of costs.

This will result in a transition of the classical journal publishing model to a Fair Open Access publishing model, by redistribution of several labour functions of the publishing model back to academia as follows:



The journal flipping procedure consists of two distinct stages:

1. The transition of three years*:* The journal's editorial board seeks admission to the Open Library of Humanities. The editorial board then asks the publisher to comply with the conditions of Fair Open Access. If the publisher refuses to comply, the entire editorial board leaves the journals to set up a new journal with a publisher who does. APCs are paid for by a 3-year fund. For LingOA, the fund is financed

by the Netherlands Organization for Scientific Research NWO and the Association of Dutch Universities (VSNU). Radboud University Library has provided a journal manager for the 4 journals.

2. The final stage after three years*:* Journals that move have re-established their prestige metrics. APCs are paid by the consortium of librariesparticipating in the Open Library of Humanitiesensuring long-term sustainability.'

The Open Library of Humanities is a non-profit, academic-led open access publisher and open-access funder for the humanities and social sciences. OLH promotes flipping existing subscription journals to open access. It consists of a library consortium model: participating libraries pay an annual membership fee (currently approximately €500 - €1500) that pays for all APCs of OLH-associated journals. Libraries vote on which journals to admit to OLH after an initial screening procedure. By March 2017, over 220 libraries participated in the OLH, including Harvard, Princeton, Yale, Carnegie Mellon, UCL, and Cambridge. OLH recognises that there is labour in publishing that must be fairly remunerated but also subscribes to the Fair open access principles and is willing to work with and potentially fund any publishers who also do so. Thus, OLH provides a long-term sustainable solution for flipping existing journals from subscription to Fair Open Access, enabling libraries to redirect funds from subscriptions to APCs.

There are several features of this flipping model that make it a desirable alternative to the above big-deal and/or offsetting deals:

1. A discipline-based approach. Within each academic discipline, a foundation is set up that helps flipping established subscription journals to Fair Open Access. Existing networks within the discipline are exploited to influence editors to flip their journal to Fair Open Access. This allows for an understanding of specific disciplinary circumstances.

2. No APCs. The foundation pays for the APCs during the transition period. It also covers legal advice costs associated with flipping the journals. This avoids the cost-concentration effects of article processing charges.

3. Long-term sustainability. After the transition period, journals join a worldwide library consortium such as the one provided by the Open Library of Humanities. The worldwide library consortium durably pays for APCs. Thus, library funds are redirected from subscriptions to APCs while maintaining the distribution effects of a subscription environment.

The LingOA model provides a tested roadmap for flipping subscription journals to Fair Open Access; a model that increases downward price pressure and cost re-distribution. Investment in the funding for the transition period is temporary, because it is only necessary during the transition period. On the other hand, long-term return on investment is substantial across library bodies, since in the longer term the model facilitates the cancellation of subscriptions, the re-negotiation of big-deal bundles, and the costs of open-access publications become transparent and in proportion to the work carried out. The scheme also facilitates downward price pressure on APC models since the OLH is also an in-house publisher and so is able to generate a rationale for a specific price point. The library consortia of the model of the Open Library of Humanities enable library funds to be redirected from subscription to open access, although additional upfront philanthropic or government funding may be required to expand such consortia to the necessary size to support larger disciplines since building such a consortium is highly labour intensive. This transition to Fair Open Access is also driven by editors and authors. And last but not least: academics face no costs for publishing or accessing research results.

5. Conclusion

The road to open access will likely involve multiple business models. Although the APC route has gained favour with a number of organizations because it presents a mode in which a unit cost can be assigned per article, the concentrating effects of this model make it difficult to scale without economic redistribution. Subscription models retain the distributional characteristics that spread these costs among many actors but come with an access gap. Consortial models for open access retain the best of both of these models but only by changing to a non-classical economic environment. That is, there is the potential for free riders.

The LingOA model, underwritten in the long-term by the Open Library of Humanities, presents a case study for a transition to open access that includes offsetting using a non-classical economic model. While costs to libraries remain low, there is no evidence of the possibility of free-riders causing cancellations and the OLH has a 100% renewal rate. As the model scales, it may be necessary to fragment the offerings into different packages to avoid both "Big Deal"-style lock-ins and also to ensure that scaling does not lead to library drop-outs. Further, in targeting publication venues rather than individual authors, the social features of this model lead to open access without the need to convince every academic author of its merits.

Finally, there is enthusiasm to extend the LingOA model into other disciplinary spaces, specifically in mathematics and psychology. An expansion to these other disciplines will require up-front funding to cover the transition period and to cover the expansion of the OLH consortium.

References

[1] Council of the European Union, *The Transition towards an Open Science System - Council Conclusons*, 2016.

[2] Stephen Pinfield, Jennifer Salter, and Peter A. Bath, The "total Cost of Publication" in a Hybrid Open-Access Environment: Institutional Approaches to Funding Journal Article-Processing Charges in Combination with Subscriptions, *Journal of the Association for Information Science and Technology* **67.7** (2016), 1751–66 <https://doi.org/10.1002/asi.23446>.

[3] Stuart Lawson, Jonathan Gray, and Michele Mauri, Opening the Black Box of Scholarly Communication Funding: A Public Data Infrastructure for Financial Flows in Academic Publishing, *Open Library of Humanities* **2.1** (2016) <https://doi.org/10.16995/olh.72>.

[4] Martin Paul Eve, Co-Operating for Gold Open Access without APCs, *Insights* **28.1** (2015) <https://doi.org/10.1629/uksg.166>.

[5] Liam Earney, Offsetting and Its Discontents: Challenges and Opportunities of Open Access Offsetting Agreements, *Insights the UKSG Journal* **30.1** (2017), 11–24 <https://doi.org/10.1629/uksg.345>.

[6] Max Planck Digital Library, Roadmap, *Open Access 2020*, 2017 <https://oa2020.org/roadmap/>.

[7] Peter Suber, *Open Access, Essential Knowledge Series,* MIT Press, Cambridge, 2012, p. 39 <http://bit.ly/oa-book>.

[8] Martin Paul Eve, *Open Access and the Humanities: Contexts, Controversies and the Future,* (Cambridge University Press, Cambridge, 2014, chap. 2 <http://dx.doi.org/10.1017/CBO9781316161012>.

[9] Association of Research Libraries, *ARL Statistics 2009-2011*, 2014 <http://www.arl.org/storage/documents/expenditure-trends.pdf>.

[10] Research Consulting, *Towards a Competitive and Sustainable OA Market in Europe (OpenAIRE, February 2017)*, p. 17 <https://blogs.openaire.eu/wp-content/uploads/2017/03/OA-market-report-28Final-13-March-201729-1.pdf>.

[11] Deni Auclair, *Open Access 2015: Market Size, Forecast, and Trends (Outsell, 2015)*, p. 7 <http://www.casalini.it/retreat/2015_docs/auclair.pdf>.

[12] Peter Suber, *Open Access, Essential Knowledge Series,* MIT Press, Cambridge, 2012, p. 53 <http://bit.ly/oa-book>.

[13] First posted at Martin Paul Eve, *100 People in a Room: On the Distributional Effects of Different Open-Access Funding Models*, Martin Paul Eve <https://www.martineve.com/2017/04/03/100-people-in-a-room/>.

[14] David Prosser, *The Costs of Double Dipping*, Research Libraries UK, 2015 <http://www.rluk.ac.uk/about-us/blog/the-costs-of-double-dipping/>.

[15] Adam Smith, 'Interview with Alicia Wise', *Research Fortnight*, 2014, 6.

[16] For a critique, see Martin Paul Eve, *How Can Elsevier Claim That Its OA Revenue Stream Is Separate from Subscriptions in a Hybrid Environment?*, Martin Paul Eve, 2015 <https://www.martineve.com/2015/02/14/how-can-elsevier-claim-that-its-oa-revenue-stream-is-separate-from-subscriptions-in-a-hybrid-environment/>.

[17] Stuart Lawson, R*eport on Offset Agreements: Evaluating Current Jisc Collections Deals. Year 1 – Evaluating 2015 Deals (Jisc Collections, 24 October 2016)*<http://eprints.rclis.org/30139/>.

[18] Research Consulting, *Towards a Competitive and Sustainable OA Market in Europe (OpenAIRE, February 2017)*, p. 23 <https://blogs.openaire.eu/wp-content/uploads/2017/03/OA-market-report-28Final-13-March-201729-1.pdf>.

[19] *Publisher Deals*, Openaccess.nl, 2017 <http://openaccess.nl/en/in-the-netherlands/publisher-deals>.

[20] *Public Access Request*, VSNU, 2016 <http://www.vsnu.nl/en_GB/public-access-request>.

[21] Sicco de Knecht, *Leaked Elsevier Contract Reveals Pushback*, ScienceGuide, 2017 <http://www.scienceguide.nl/201703/leaked-elsevier-contract-reveals-pushback.aspx>.

[22] Parts of the description in the following section have been published, or are under submission, elsewhere.

[23] Clarification note on Fair Open Access principle 1: This could be ownership by an editorial board, or by a democratically controlled scholarly society, for example. Key points are that the controlling organization, not a commercial publisher, must own the journal title, so that a change of service provider can be achieved without changing the title, and so publishing companies simply compete to offer services to the journal. We strongly recommend that the ownership structure allow for democratic input by the community of readers, authors and referees, in addition to editors, and that procedures for making key decisions about the journal's future be formally (even legally) specified. We strongly recommend that the governing organization be fully nonprofit (for example, IRS 501 (c) (3) in USA). A for-profit company accountable only to shareholders is not compatible with Principle 1.

[24] Clarification note on Fair Open Access principle 2: The journals and their publishing house can still propose, among their services, to take care of possible legal issues pertaining to copyright on the author's behalf, under the author's oversight. We strongly recommend that reviewers also retain copyright of their reviews, and journals retain ownership of all correspondence and mailing lists compiled on the electronic submission system put at their disposal by the publisher.

[25] Clarification note on Fair Open Access principle 3: Any form of subscription paywall is unacceptable, including "hybrid OA". We strongly recommend that the industry standard CC-By licence be used. All content of the journal should be easily accessible from the journal website to anyone with a standard internet connection.

[26] Clarification note on Fair Open Access principle 4: The key idea is that journals be "free at the point of use" by authors and readers. Principle 3 deals with readers and Principle 4 with authors. Compulsory APCs (article publication charges) are not compatible with this principle. Journals should ideally be funded by general contributions from universities and research funders, with these contributions not tied to individual articles or groups of authors. Principle 4 is not compatible with "APC Big Deals", whereby institutions pay for APCs of their employees but do not contribute to a general fund. Also not compatible is the practice of charging APCs by default to the author's institution, with waivers for authors who do not have institutional funds. The principle does not preclude voluntary APCs, but requests for these must be unobtrusive and no barrier to publication. APCs must be "opt-in", never "opt-out".

[27] Clarification note on Fair Open Access principle 5. "Low" depends on the particulars of each journal, but we strongly recommend an absolute maximum of $1000 per article published or $50 per page for the total expense of any journal, and substantially lower fees in all possible cases. We recommend that

an itemized price structure be made public in order to ensure transparency and make the proportionality principle apparent.

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-129

Rethinking Openness: Challenges and New Approaches to Open Scholarly Journals

Andre Luiz APPEL[a,1], Maria Lucia MACIEL[a] and Sarita ALBAGLI[b]
[a] *Universidade Federal do Rio de Janeiro*
[b] *Instituto Brasileiro de Informação em Ciência e Tecnologia*

Abstract. This paper‘s main objective is to present and discuss some results of a research project in progress on issues currently in debate on open scholarly journals dedicated to the publication of research outputs on an open, democratic and transparent basis. It contemplates a brief review of the literature about challenges regarding openness of the current scholarly publication system and political-economic constraints to its democratization, to support the analysis of two case studies on open publication platforms - Research Ideas and Outcomes and Wellcome Open Research - based on information available on their websites. As results, we present an analysis of publication practices and policies in action on these platforms and their policies.

Keywords. scholarly journals, open scholarly journals, open access, open science.

1. Introduction

What should an open scholarly journal be like? To what extent and in what ways does it differ from the current journals format so far? What are the features which may distinguish a journal aimed to promote the concept of open in as many as possible aspects? To contribute to answers to these questions, the main objective of this paper is to present and discuss some results of a research project in progress on issues currently in debate on open scholarly journals dedicated to the publication of research outputs on an open and transparent basis.

The study was performed in two phases. The first phase was based on a review of the literature about challenges regarding openness of the current scholarly publication system and political-economic constraints to its opening. The second phase involved the development of two case studies on open publication platforms: Research Ideas and Outcomes (also known as RIO Journal or RIO)[2] and Wellcome Open Research (WOR)[3]. Given the evidences of practices highlighted in the first phase of the study, we developed an analysis of their publication practices and policies based on information available on their websites.

The development of new and more open publishing practices has evolved since the advent of the Open Access movement and is in tune with the emerging Open Science movement. According to Albagli [1] "the movement for Open Science must be

[1] Corresponding Author: Programa de Pós-Graduação em Ciência da Informação IBICT/UFRJ, Rua Lauro Müller, 455 / 4º andar – Botafogo, Rio de Janeiro, RJ, Brasil; E-mail: alappel@gmail.com.
[2] http://riojournal.com/.
[3] https://wellcomeopenresearch.org/.

considered within the context of the social movements that have emerged in the scenery of the changing conditions of production and circulation of information, knowledge and culture”. It is motivated by greater efficacy of data and information sharing as a basis for the cooperative, cumulative generation of eventually reliable additions to the stock of knowledge, to putting new findings in the public domain which permit data and information to be concurrently shared in use and re-used indefinitely [2].

These movements may be framed in a broader perspective which benefits the understanding of knowledge and information – and all the infrastructure and means necessary for their creation, storage/conservation and dissemination - as common shared entities. This approach involves circumventing the commercial for-profit exploitation of scientific knowledge and information infrastructures to create alternatives under the principle of the commons, driven by social norms or regulations [3,4].

2. Emerging Challenges for Journal Publishing Regarding Openness

The literature reveals that, after 350 years of publication of the first scholarly journals – *Journal des Sçavans* and the *Philosophical Transactions of the Royal Society* –, scientific journals are now facing challenges regarding, on the one hand, the expansion of electronic means of publication and, on the other, the pressure to be more open. This implies the need to enhance reproducibility, to incorporate outputs beyond the research results enclosed in text-only format, such as data sets and early stages of the research cycle, and to incorporate new approaches of peer review and post-publication evaluation. Cope and Kalantzis [5] point to an epistemic disruption in the scientific knowledge communication system with repercussions on academic journals. It has been driven by technological, cultural, economic and (geo)political factors, favoring the adoption of platforms and practices for more distributed knowledge production and circulation.

One of the main challenges to the establishment of open publications relates to finding a suitable business model that allows for long term and sustainable funding beyond commercial and monopolistic exploitation. Cope and Kalantzis [5] raise concerns about the unsustainable costs and inefficiencies of traditional commercial scientific publishing, which lead to the expensive costs of subscription journals. It is also important to emphasize that commercial publishers have a focus on high profits, and that the subscription fee or cost per article does not necessarily reflect the production costs but also the journal influence [6]. Thus, a large proportion of the community (researchers and the public) cannot access the published research [7]. The rise of open access (OA) journals is challenging the business models of scientific journals [5] and is also demanding the development of more sustainable publishing models. Open access journals need financial sustainability, just as commercial journals need to adapt if they are to continue making a profit in the open access model. Some OA journals charge the authors a publication fee, also known as Article Processing Charge (APC), once the article is accepted for publication. Others, known as hybrid journals, still published under a subscription model, charge the authors to publish their article in OA.

Another challenge is the need for sharing research data along with the published article. Considering the growing amount of shared data, Brown [8] argues that due to limited page space in an article to present data, it became necessary for scientists to organize, disseminate, and archive their research-related data digitally, and then link that data to the article. This practice is reinforced by Tenopir and King [9] who highlight that citation linking within and between articles and links from the article to external data sets represent some of the future trends of journals and article publication. Data sharing has been increasingly valued to enhance scientific knowledge credibility and certification without the intervention of discursive rhetoric of the authors when the research results presentation is limited to the article text [5]. Open access publishing does not necessarily reduce the closure in scholarly knowledge production and communication. It still "by and large perpetuate[s] the print analogue workflow of PDF, with all its intrinsic deficiencies as an open knowledge system" [5]. As with data, many findings are currently not published, such as small studies and software papers [7]. The open publication of diverse results throughout the research cycle may enhance collaboration besides enabling other researchers to replicate studies or to find new results without the need to re-collect data.

The standard peer review system is also being affected by new approaches to meet the demands for more open, transparent and rapid review processes and to increase the possibility of granting credit to all those involved in the process. Tracz and Lawrence [7] argue that the lapse of time since the article is submitted until the time it is published and the lack of transparency in the anonymous review process are some of the problems of the current system. Another problem highlighted by the authors is the waste of time involved in finding a journal that accepts the article, caused by inefficient reviewing processes. As an alternative, with the implementation of an open peer review system, articles are readily published if they meet the editorial standards and guidelines required by the journal and then become available for the referees to make public comments – an example of transparent review or the post-publication review as described by Ford [10]. This process should increase both the credit and accountability for peer reviewing [11] especially if the comments and reviews are published along with referees' Open Researcher and Contributor IDs (ORCID).

Journals dedicated to openness in the publication of research may also engage with alternative forms to assess articles' relevance and impact after publication. Regarding post–publication evaluation, Cope & Kalantzis [5] raise questions and concerns about the fact that this evaluation is centered primarily on citation or impact analysis, while many researchers advocate for the adoption of alternative metrics ("altmetrics") and article level metrics – such as article download counts or those collected from reference management tools and social media – as a complement to assessing article impact, and how it is being discussed, shared and used [12–14].

3. Political-Economic Aspects of Science Publishing

For over three centuries the commercial publishing industry has kept a monopoly of scholarly journals publishing, considering the high level of investments in fixed and circulating capital necessary for their printed versions. With the rise of new techniques for publication and dissemination of science, derived mainly from the advent and popularization of personal computing and the Web, these barriers to entry no longer made sense, since most researchers were then able to publish their findings by themselves.

By the beginning of the 21st century, with the spread of electronic publishing, Houghton accounted for an increase in competition in the publishing market, with a possible reduction of the monopolies, and a transfer of scholarly communication from the hands of commercial publishers into those of the creators [15]. Nonetheless, almost twenty years after this assertion, we keep facing the resilience of the traditional journal format and the prevalence of journals maintained by commercial publishers. According to Larivière et al. [16], only a few publishers – namely Reed-Elsevier, Wiley-Blackwell, Springer, and Taylor & Francis – are responsible for the publication of almost 50 percent of all papers.

Such resilience also tends to reproduce some flaws in the publication system of the print era. As stated before, papers' text and results are enclosed in PDF format, which represents a barrier to the processes of sharing and reuse of previous studies and data within the paper. Aligned with this, the transference of copyrights to publishers by authors also prevents the reuse of this paper in processes like Text and Data Mining (TDM) for knowledge-generation, automated screening for errors and automated literature searches that renew scientific discovery [17].

The transference of copyrights also led to the continuity of value exploitation by publishers. Throughout the print paradigm, publishers have invested in the commodification of scientific knowledge and information with their commercialization as marketable and tangible objects. Electronic publishing led to the dissipation of the exchange value of journals or papers as saleable goods, since "the publisher does not have to upload or produce an additional copy each time a paper is accessed on the server as it can be duplicated *ad infinitum*, which in turn reduces the marginal cost of additional subscriptions to 0" [16], leaving no parameters for the definition of subscription prices. Publishers thus have been operating towards pure rentier capitalism, by monopolizing a public resource then charging exorbitant fees to use it [18], taking advantage of the rights granted by authors. It is the consolidation of a regime of scarcity in which access to knowledge and information is controlled and limited, mainly by price, technical barriers and/or legal (copyright) constraints [3].

Some other persistent flaws are related to the minimal or inexistent possibility for the authors to manage or at least actively contribute to the review, editing and publication processes, which are still by and large mediated by commercial publishers. Scientists face the complete alienation from the dissemination of their creative work by giving up control and decisions regarding this process to publishers. This mediation is also achieved with intense exploitation of other scientists' labor for free, performing tasks such as peer reviewing, editing and editorial duties. Based on five studies addressing the economics of the scholarly journal system, King and Tenopir [19] concluded that researchers' time dominates the overall cost of scholarly journal communication, accounting for 79.5 percent. Those costs are not covered by publishers, the main profiteers of the system. They are paid by public investment, that is, by society.

The aim here would be to examine whether objective, technical, practical changes in ways of producing and distributing knowledge are being – or can be – combined efficiently into a changing culture of openness along the entire process of production while leaving behind the economic gridlocks of for-profit centered economic models.

4. Analysis and Discussion of the Selected Cases

Throughout this section, we describe the selected two case studies on open publications, pointing out their main editorial practices and the possible relations of these practices with the issues raised in this paper.

The selection of these two cases was based on the following criteria: (1) the possibility of publishing different types of research-related documents and outputs, beyond the regular article format; (2) the opportunity for authors to publish or register the full research cycle on a single platform (particularly in the case of RIO Journal), given the model of charging authors for publication and the variety of output formats; (3) open peer review as the default review system; (4) authors are responsible for the selection of reviewers or are responsible for conducting the peer-review process. These features are strongly related to a recovery of authors' control over the publication process.

RIO Journal (RIO) was launched in September 2015 with an innovative approach, creating a venue for researchers to publish the full process of their research cycle, from research ideas, proposals and methods to theses and research articles. RIO is maintained by Pensoft Publishers[4] along with many other open access journals sharing a common platform. Although RIO is a commercial for-profit operation, its founding editor declares it is not a profiteering one [20].

Wellcome Open Research (WOR), launched in November 2016, is oriented to the publication of articles and other types of documents that have at least one author who has been, or still is, a recipient of a Wellcome Trust Foundation[5] grant. It is maintained by Wellcome and operates over F1000Research - another open access venue - publishing platform.

In Table 1, we summarize some of the journals' main characteristics regarding publication, access and submission policies, which we discuss later based on the topics addressed in the literature review.

The ARPHA platform name is an acronym for Authoring, Reviewing, Publishing, Hosting, and Archiving, which emphasize its capacity to grasp the full publication cycle. ARPHA allows papers to be authored right up to the platform, with no requirement for external software, such as word processors or PDF makers. One of the advantages of this feature is that the reviewing process becomes faster, since there's no need for the reviewers to download the manuscript before and after evaluation. Another advantage relates to the processes of automated output generation in formats such as HTML, PDF and XML, as soon as the manuscript is approved by post submission editorial check. The F1000Research platform used by WOR does not offer writing manuscript functionality, and they can be submitted as Word (DOC or DOCX) or rich text format (RTF) files only. LaTeX users can alternatively submit via Overleaf[6], using journals' specific template. Since in both cases reviewing process is disclosed after publication, both platforms support article versioning. Regarding diversity of output formats, RIO presents a larger spectrum of formats (31) in line with its proposal to contemplate full research cycle publication. WOR allows the publication of only 11 different article types, but works with the concept of "living" articles, which allows authors to update their articles with novel relevant information to the findings.

---

[4] Pensoft Publishers are a publisher of scientific literature based in Sofia, Bulgaria. http://pensoft.net/.
[5] A medical research charity funding research into human and animal health based in United Kingdom. https://wellcome.ac.uk/funding.
[6] https://www.overleaf.com/.

Both publications work with an article-based (or continuous) publishing model and articles are made available as soon as they are approved by editorial/technical check or peer review. Both adopt open and public peer review. In WOR the authors led the peer review process openly, inviting reviewers after the article is made public. RIO Journal provides three stages of peer review: (1) author-organized, pre-submission; (2) community-sourced, post-publication; and (3) journal-organized, post-publication (optional). RIO also provides a pre-submission stage of review that may be conducted as an invisible college where the authors may invite colleagues, reviewers, linguistic and copy editors prior to submission for checking the manuscript.

Table 1. List of journal characteristics

Characteristics	RIO Journal	Wellcome Open Research
ISSN	2367-7163	2398-502X
Platform	ARPHA*	F1000Research**
Document versioning	Yes	Yes
# of diverse outputs	31	11
Peer review	Open (post-publication)	Open (pre-submission and post-publication)
Submission charges	-	-
Article Processing Charges	€ 50 - 550 (for single publications) € 430 - 4,250 (for “research cycle packages”)	US$ 135 - 900
License for articles	CC BY or CC 0	CC BY
License for data	Exclusively CC 0	CC 0
Copyright retention	Authors retain the copyright	Authors or their institution retain the copyright
Altmetrics	Yes	Yes

Notes:
(*) Pensoft Publishers, http://arphahub.com/about/platform.
(**) https://f1000research.com/.

Given these platforms’ characteristics, it is possible to highlight the importance of authors’ interaction with the process of publishing their production, from the beginning of submission to the final confirmation and/or complementation of peer reviewed versions. With authors becoming fully responsible for the authoring and the reviewing processes, the level of alienation in the publication system is strongly reduced, hence they achieve the status of real owners of their production. In addition, the open peer review process grants credit to reviewers’ work. Each reviewer contribution has its own DOI number being possibly cited or retrieved for other purposes.

Both RIO and WOR require authors to share the articles' supporting research data, either as supplementary material, under Creative Commons Zero licensing, or by deposit in a proper data repository. WOR presents extensive and detailed data preparation guidelines, and both platforms suggest specific repositories for deposit and ways to present or link the data in the document submitted to the journal. Such requirements enhance the replicability of published studies as well as allowing other researchers to share diverse interpretations of the same phenomena [5] or to disclose new analysis and conclusions over the same data.

Regarding licensing and copyright policies, RIO allows authors to choose between Creative Commons Attribution (CC BY) or Creative Commons Public Domain Dedication (CC-Zero) licenses, while WOR works exclusively with CC BY license. The authors remain as the copyright holders to their articles in both platforms.

Regarding business models, both platforms apply Article Processing Charges. WOR charges are based on word counts – with costs ranging from 135 to 900 USD – and since the authors are funded by Wellcome, the charges are covered with these funds. WOR charges represented 90 percent of the charges applied by its hosting platform, F1000Research. RIO, besides word counts, also charges according to publication type, ranging from 50 to 550 EUR. RIO also works with a charging mode called "Research Cycle Packages", which allows authors to publish a certain number of outputs along a research project. The most expensive package, intended for large collaborative projects, covers up to 15 publications and costs 4,250 EUR. This model resembles PeerJ "lifetime" memberships that allow, for example, five peer-reviewed publications per year. We could not find mentions to charges waiver policies in practice in either platform.

In addition to adopting business models that drop subscription charges and the enclosure of distribution and copyrights in favor of the application of article processing charges for publication with open nonrestrictive licenses, it is possible to highlight, regarding the selected cases, an effort to make the editorial processes more open and led by the academic community. The combination of practices such as open peer review led by authors, the adoption of open and flexible formats, standards for knowledge and information distribution, all beside the application of charges and types of publication that can meet the different interests of the authors, may contribute to the expansion and consolidation of what can be understood as an open publication.

5. Final Remarks

These case studies are good examples of how scientific journals can transform their editorial processes to incorporate more open and innovative practices regarding the publication of research outputs. It is becoming common sense that open goes beyond access, in order to also affect formats, evaluation, sharing, assessment, etc. As such, the concept of open journal is under development to accomplish new standards related to openness.

We highlight the need for the development of studies or initiatives favoring increased control of the scholarly communication system in the hands of scientists. The communication process is part of scientists' work, therefore it could only be fully concretized if its objectives are defined by those who perform it. Besides the creation and maintenance of open venues for dissemination of scholarly outputs, it is also relevant to guide the production and the workforce (including editorial staff, reviewers, developers, etc.) to these venues, in order to stimulate a culture of availability and openness which supports open infrastructures. In this regard, we must highlight a lack of studies questioning the low rate of adherence or the persistent constraints to a massive adherence of the scholarly community to open initiatives, such as by publishing and reviewing exclusively in/for open access journals.

References

[1] Albagli S. Open science in question. In: S. Albagli MLM, Abdo AH, editors. Open Science, open issues [Internet]. Rio de Janeiro, Brasil: IBICT/UNIRIO; 2015. p. 9–25. Available from: http://livroaberto.ibict.br/handle/1/1061
[2] David PA. The Historical Origins of "Open Science": An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution. Capitalism and Society [Internet]. 2008;3(2). Available from: http://dx.doi.org/10.2202/1932-0213.1040
[3] Kuhlen R. Knowledge is the water of the mind: how to structure rights in the "immaterial commons." In: Bollier D, Helfrich S, editors. The wealth of commons: a world beyond market and state [Internet]. The Commons Strategy Group; 2012. Available from: http://wealthofthecommons.org/
[4] Meretz S. The structural communality of the commons. In: Bollier D, Helfrich S, editors. The wealth of commons: a world beyond market and state [Internet]. The Commons Strategy Group; 2012. Available from: http://wealthofthecommons.org/
[5] Cope B, Kalantzis M. Sign of epistemic disruption: Transformation in the knowledge system of the academic journal. First Monday [Internet]. 2009;14(4):1–26. Available from: http://firstmonday.org/ojs/index.php/fm/article/view/2309/2163
[6] van Noorden R. The true cost of science publishing. Nature [Internet]. 2013;495(7442):426–9. Available from: http://dx.doi.org/10.1038/495426a
[7] Tracz V, Lawrence R. Towards an open science publishing platform. F1000Research [Internet]. 2016;5(130):130. Available from: http://dx.doi.org/10.12688/f1000research.7968.1
[8] Brown C. Communication in the sciences. Annual Review of Information Science and Technology [Internet]. 2010;44(1):285–316. Available from: http://dx.doi.org/10.1002/aris.2010.1440440114
[9] Tenopir C, King DW. The growth of journals publishing. In: Cope B, Phillips A, editors. The Future of the Academic Journal [Internet]. 2nd ed. Oxford, England: Chandos Publishing; Elsevier; 2014. p. 105–23. Available from: http://dx.doi.org/10.1533/9781780634647.159
[10] Ford E. Defining and Characterizing Open Peer Review: A Review of the Literature. Journal of Scholarly Publishing [Internet]. 2013;44(4):311–26. Available from: http://dx.doi.org/10.3138/jsp.44-4-001
[11] van Rooyen S, Godlee F, Evans S, Black N, Smith R. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. British Medical Journal [Internet]. 1999 Jan 2;318(7175):23–7. Available from: http://dx.doi.org/10.1136/bmj.318.7175.23
[12] Priem J, Taraborelli D, Groth P, Neylon C. Altmetrics: a manifesto [Internet]. 2010. Available from: http://altmetrics.org/manifesto
[13] Neylon C, Wu S. Article-level metrics and the evolution of scientific impact. PLoS Biology [Internet]. 2009 Nov;7(11):e1000242. Available from: http://dx.doi.org/10.1371/journal.pbio.1000242
[14] Tananbaum G. Article-Level Metrics: A SPARC Primer [Internet]. Washington, DC: Scholarly Publishing and Academic Resources Coalition; 2013 p. 14. Available from: http://sparc.arl.org/resource/sparc-article-level-metrics-primer
[15] Houghton JW. Crisis and transition: the economics of scholarly communication. Learn Publ [Internet]. 2001;14(3):167–76. Available from: http://dx.doi.org/10.1087/095315101750240412
[16] Larivière V, Haustein S, Mongeon P. The Oligopoly of Academic Publishers in the Digital Era. PLoS One [Internet]. 2015 Jun 10;10(6):e0127502. Available from: http://dx.doi.org/10.1371/journal.pone.0127502
[17] Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink CHJ. The academic, economic and societal impacts of Open Access: an evidence-based review. F1000Res [Internet]. 2016;5:632. Available from: http://dx.doi.org/10.12688/f1000research.8460.3
[18] Monbiot G. Academic publishers make Murdoch look like a socialist. The Guardian [Internet]. 2011 Aug; Available from: https://www.theguardian.com/commentisfree/2011/aug/29/academic-publishers-murdoch-socialist

[19] King DW, Tenopir C. Some economic aspects of the scholarly journal system. Annual Review of Information Science and Technology [Internet]. 2011;45(1):295–366. Available from: http://dx.doi.org/10.1002/aris.2011.1440450114
[20] Rabesandratana T. A new journal wants to publish your research ideas [Internet]. Science | AAAS. 2015 [cited 2017 Mar 23]. Available from: http://dx.doi.org/10.1126/science.aad1696

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-138

Alternative Metrics for the Evaluation of Scholarly Activities: An Analysis of Articles Authored by Greek Researchers

Aspasia TOGIA[a,1], Eleftheria KOSEOGLOU[b] and Sofia ZAPOUNIDOU[b]
[a] *Department of Library Science & Information Systems, ATEI of Thessaloniki*
[b] *Library & Information Centre, Aristotle University of Thessaloniki*

Abstract. Recently, altmetrics have emerged as alternative means of measuring scholarly impact, aiming at improving and complementing both traditional and web-based metrics. The aim of the present study is to contribute to the altmetrics literature by providing an overview of the coverage of altmetrics sources for the Aristotle University of Thessaloniki (AUTh) publications. We used Scopus to collect all research articles stating AUTh as the affiliation of at least one author and published from 2010 to 2016. The altmetric data originated from Altmetric Explorer, a service provided by Altmetric.com. Only 17% of all publications retrieved from Scopus had some kind of mentions, while there was a clear increasing trend over the years. The presence of altmetrics was different from each Altmetric.com attention source. Around 81% of all mentions came from Twitter. Facebook was a distant second, followed by news outlets. All other sources had very low or negligible coverage. The overwhelming majority of tweets had been posted by members of the public, who do not link to scholarly literature. Medical Sciences had by far the highest number of publications with altmetric scores, followed, in a distance by Sciences. However, Arts, Humanities and Social Sciences publications exhibited a significant altmetric activity. More research is needed in order to get a better insight into the altmetric landscape in Greece and develop an understanding about the kind of influence altmetrics measure, and the relationship, if any, between altmetric indicators and scientific impact.

Keywords. Bibliometrics, altmetrics, social media metrics, research impact, Greece

1. Introduction

The problem of measuring the impact of scientific publications is of high importance to scholars, research teams and academic institutions. The impact of research is being taken into account in decisions about tenure, promotion, and fund allocation. “As the demand for greater accountability in all areas of public expenditure is constantly growing, the topic of research assessment becomes very relevant” [1]. A common

[1] Corresponding Author. Address: A.T.E.I. of Thessaloniki, P.O BOX 141, GR - 574 00 Thessaloniki, GREECE. Tel.: +302310013559, +302310320858. E-mail: aspatogi@libd.teithe.gr

approach to scholarly impact that prevails today is using bibliometric indicators based on citation analysis. Citations are being used to measure the impact of articles [e.g., 2,3], journals [4], researchers [5–7], and scientific fields [8,9].

However, several limitations question the validity and reliability of these traditional measurements. Citations take time to accumulate [10] and measure only one type of research product, peer-reviewed articles [11]. There is also evidence that authors cite only a fraction of their influences, and they do not always cite important works in a positive manner [12]. In addition, there are variations in citation practices across different countries, time periods, disciplines, and even specialties within the same discipline, while some areas of research are not frequently cited [13–15]. Another problem with impact indicators based on citation counts is that they assess the extent to which research is noticed by other researchers, and are not informative of the impact on stakeholders outside the academy, like students, practitioners, policy makers, and even the general public [16].

In order to alleviate some of these problems, scientometricians have created more diverse measures of research impact [17], while the World Wide Web provided new mechanisms of measuring access to information, and made possible the development of webometrics or cybermetrics, a modern branch of bibliometrics, which embraces the quantitative analysis of activity on the Web [18]. But even these new approaches have limitations and are incapable of capturing the real impact of scientific endeavors. Webometrics, for instance, "are affected by the distributed, diverse and dynamical nature of the Web and by the deficiencies of search engines" [19, p. 78]. Recently, altmetrics have emerged as alternative means of measuring scholarly impact, aiming at improving and complementing both traditional and web-based metrics [20]. Altmetrics "measure Webdriven scholarly interactions, such as how research is tweeted, blogged about, or bookmarked" [21]. Altmetrics cover a wide range of research products, since they not only measure article-level metrics, measure impact in a broader sense by looking at more than citations, give an insight into impact on diverse audiences, and an indication of societal impact of research, and provide a better understanding of how a scholarly product is being used [22,23].

While altmetrics is a growing research area and have the potential to meet many of the challenges faced by traditional bibliometric indicators, there is a need to further investigate their reliability, validity and context [24,25]. Therefore, the present study seeks to contribute to the altmetrics literature by providing an overview of the coverage of altmetrics sources for the Aristotle University of Thessaloniki (AUTh) publications. In particular, the study focuses on the following research questions:

1. How much and what kind of altmetrics data exist for the documents authored by AUTh faculty members and published between 2010 and 2016?
2. What is the presence of altmetrics for AUTh publications across different subject fields?
3. What is the demographic breakdown of the mentions for the top mentioned publications?

2. Related Work

Several studies have been conducted to investigate the usefulness of altmetrics as sources of impact assessment. These studies examine the extent to which scholarly journal articles are represented on various social media platforms, the attention they receive, and their correlations with citations. Basic measures, such as coverage, density and intensity of altmetrics have been reported, and composite indicators, such as the the Altmetric Score, have been proposed [26]. "Coverage is defined as the percentage of papers with at least one social media event or citation. Density is the average number of social media counts or citations per paper (…) while intensity indicates the average number of social media or citation counts for all documents with at least one event (non-zero counts)" [27, p.5].

Coverage of research articles on social networking sites, such as Facebook and Twitter, has been found to be rather low. Twitter seems to be the source that provides more scores, while the values for other social media counts, such as Facebook and blogs, are very small, with coverage values usually below 10% [17,27–31]. Lower rates (just over 8%) have been found for publication from Latin-American countries [28], while a cross-disciplinary study of altmetrics found that only 1.6% of the sampled papers published between 2005 and 2011 had at least one tweet. [30].

On the other hand, online reference managers have significantly higher coverage of documents. Mendeley, in particular, seems to dominate, as it has been shown to have a much greater number of readers per document than other online reference managers. In a number of studies, Mendeley has emerged as the most exhaustive altmetrics data source [17,28,30,32–36].

There is evidence, however, that social media coverage is increasing over time, and varies by discipline and specialty. In a study of tweets in biomedical literature, the proportion of papers having at least one tweet increased from 2.4% of the papers published in 2010 to 20.4% of the papers published in 2012 [32]. According to [31] the percentage of publications that received some altmetric score increased from around 11% in 2011 to over 25% in 2013. Furthermore, in a recent investigation of the coverage of altmetrics in Singapore, a significant increase from 7% in 2009 to 28% in 2013 was observed [37].

In a number of studies, Medical and life sciences had a comparatively high share of publications with altmetric scores [30,33,34,40]. Articles from the social sciences and humanities also exhibit a high altmetric activity, while their altmetric density is similar to their citation density [30,34,39]. On the contrary, Mathematics, Physics, Computer Science and Engineering seem to receive lower number of altmetric scores [27,33,34,39], although in one case the most papers mentioned on Twitter related to Physics [38].

Many researchers have examined the correlation between altmetrics and citation counts, often with contradictory and inconclusive results. There is evidence that publications cited in Wikipedia and blog posts are mostly in high impact journals [39,40]. A study of eleven altmetrics showed that six of them (tweets, Facebook wall posts, research highlights, blog mentions, mainstream media mentions and forum posts) associated with citation counts, but no evidence about the strength of the correlation was provided [41]. Among social media metrics, citations seem to correlate the most with Mendeley, with correlations ranging from moderate to high [17, 30, 35–37, 40, 43–45]. Strong correlations have been found between citations from Google Scholar and tweetations [46], and between Twitter mentions, arXiv downloads, and article citations [38]. Other studies, however, found weak correlations between altmetric indicators and citations. [27], [33] and [44] identified low correlation between the number of citations and tweets per document, a finding in accordance with [31] and

[28], who suggested that correlations, where existed, were positive but low, and [37], who identified small to medium correlations between citation counts and altmetrics.

3. Methodology

The aim of the present study is to give an overview of the coverage of altmetrics sources for Aristotle University of Thessaloniki (AUTh) publications. Therefore, we used Scopus to collect all research articles stating AUTh as the affiliation of at least one author and published from 2010 to 2016. Scopus was selected because it seems to have more thorough coverage than the Web of Science, at least for more recent articles [47]. The altmetric data used in the study originated from Altmetric Explorer for Institutions, a service provided by Altmetric.com. Altmetric Explorer gathers article-level metrics from a range of sources, including policy documents, social networks, online reference managers, mainstream media and blogs, post-publication peer review forums, and other online sources, such as Wikipedia and multimedia platforms. Once the Altmetric data have been retrieved, they are displayed on the Altmetric details page, along with the Altmetric Attention Score and donut, which demonstrate how much and what kind of attention a research output has received [48]. Altmetric Explorer was chosen because it has been reported as the most comprehensive source of altmetric data associated with scholarly articles [27].

A prerequisite for retrieving data from Altmetric Explorer is that the specific item has a unique identifier, such as a PubMedID, or a Digital Object Identifier (DOI). Thus, altmetric data were gathered only for publications having a DOI. All data were collected during the last week of January 2017. Mendeley was excluded from the analysis.

4. Findings

4.1 Presence of Altmetrics for AUTh Publications

Table 1 presents the number of AUTh publications indexed in Scopus for the period 2010-2016, the number of publications for which altmetric data were gathered via Altmetric Explorer, and the number of publications with altmetrics. As it can be seen, only 17% of all publications retrieved from Scopus had some kind of mentions. There is a clear increasing trend over the years, with coverage ranging from about 5% for items published in 2010 to above 25% for documents published in 2016.

Table 1. Altmetrics coverage for AUTh publications

Year	Publications in Scopus	Publications with DOI entered in Altmetric Explorer	Publications with altmetrics
2010	2681	2102	127 (4.7%)
2011	2962	2308	229 (7.7%)
2012	3054	2379	499 (16.3%)
2013	2861	1907	512 (17.9%)

2014	3174	2597	649 (20.4%)
2015	3091	2526	739 (23.9%)
2016	3190	2623	811 (25.4)
Total	**21013**	**16442**	**3566 (17.0%)**

The presence of altmetrics was different from each source. Around 81% of all mentions came from Twitter. Facebook was a distant second (7.5%), followed by news outlets. All other sources had very low or negligible coverage (Table 2).

Table 2. Mentions by source

Source	**Mentions**
Twitter	20828 (81.5%)
Facebook	1912 (7.5%)
News	1212 (4.7%)
Blogs	494 (1.9%)
Wikipedia articles	435 (1.7%)
Google +	305 (1.2%)
Policy docs	143 (0.6%)
Other	225 (0.9%)
Total	**25554**

4.2 Presence of Altmetrics Across Subject Fields

In order to examine the presence of altmetrics across different subject fields, we grouped the 26 subject categories of Scopus in four broad domains: Sciences, Arts, Humanities & Social Sciences, Engineering, and Medical Sciences, following the organisation of faculties and departments in AUTh. Table 3 presents the distribution of altmetrics across fields. Medical Sciences have by far the highest number of publications with altmetric scores, followed, in a distance by Sciences.

Table 3. Mentions across subject fields

Subject field	**Scopus subject category**	**Mentions**	**Publications with altmetrics**
Sciences	1. Computer science 2. Physics 3. Biochemistry	7048 (25.5%)	1469 (18.4)

	4.	Agricultural & biological sciences		
	5.	Chemistry		
	6.	Environmental		
	7.	Earth & planetary		
	8.	Mathematics		
Arts, Humanities & Social Sciences	9.	Social sciences	1445 (5.2%)	352 (21.2%)
	10.	Arts & humanities		
	11.	Psychology		
	12.	Business management		
	13.	Economics		
	14.	Decision sciences		
Engineering	15.	Engineering	1066 (3.9%)	352 (7.9%)
	16.	Materials science		
	17.	Chemical Engineering		
	18.	Energy		
Medical Sciences	19.	Medicine	18085 (65.4%)	2266 (34.4%)
	20.	Veterinary		
	21.	Dentistry		
	22.	Nursing		
	23.	Health professions		
	24.	Neuroscience		
	25.	Pharmacology		
	26.	Immunology		

As depicted in Table 4, medical publications received the highest number of mentions in all data sources. The major source for altmetrics data in all subject fields was Twitter. Publications in Medicine and Sciences received a considerable amount of attention in news outlets, and have been mentioned in several Wikipedia articles, while documents in all subject fields have been mentioned in policy documents.

Table 4. Mentions by source and subject field

	Sciences	**Arts, Humanities & SS**	**Engineering**	**Medical Sciences**
Twitter	5495	1152	866	14763
Facebook	626	197	75	1480
News	363	38	26	801
Blogs	215	18	13	273
Wikipedia articles	108	8	11	324
Google +	118	12	21	219
Policy docs	56	10	16	92
Other	67	11	38	133

4.3 Characteristics of the Top Mentioned Documents

The fifty publications with the highest attention score are presented in Table 5. The Altmetric Attention Score is automatically calculated, and is based on three main factors: the number of people who mention an item, the sources of mentions, and the authors of mentions. Instead of representing the raw number of mentions, the attention score is a weighted count of online attention a research output has received [48]. All fifty articles have been published in highly prestigious journals, mostly in the field of medicine, among which stand out the Lancet and Nature. There seems to be no obvious relationship between Altmetric Attention Score and number of citations, as papers with high score have very few citations and vice-versa. Twenty-two percent of the papers belong to open-access journals.

Table 5. The fifty papers with the highest Altmetric Attention Score

Altmetric Attention Score	Title	Journal	Year	Scopus citations	Open-access
2152	Global, Regional, And National Incidence, Prevalence, And Years Lived With Disability For 301 Acute And Chronic Diseases And Injuries In 188 Countries, 1990–2013: A Systematic Analysis For The Global Burden Of Disease Study 2013	The Lancet	2015	677	
1280	Global, Regional, And National Age-Sex Specific All-Cause And Cause-Specific Mortality For 240 Causes Of Death, 1990-2013: A Systematic Analysis For The Global Burden Of Disease Study 2013	The Lancet	2015	983	
1266	Global, Regional, And National Comparative Risk Assessment Of 79 Behavioural, Environmental And Occupational, And Metabolic Risks Or Clusters Of Risks In 188 Countries, 1990–2013: A Systematic Analysis For The Global Burden Of Disease Study 2013	The Lancet	2015	258	
591	Global, Regional, And National Disability-Adjusted Life Years (Dalys) For 306 Diseases And Injuries And Healthy Life Expectancy (Hale) For 188 Countries, 1990-2013: Quantifying The Epidemiological Transition.	The Lancet	2015	219	
503	Widespread Exploitation Of The Honeybee By Early Neolithic Farmers	Nature	2015	7	
486	Meta-Analysis Of 74,046 Individuals Identifies 11 New Susceptibility Loci For Alzheimer's Disease	Nature Genetics	2013	639	
355	Early Farmers From Across Europe Directly Descended From Neolithic Aegeans	Proceedings of the National Academy of Sciences of the United States of America	2016	15	
291	Effect Of Increased Gravitational Acceleration In Potato Deep-Fat Frying	Food Research International	2014	0	
250	Exome Sequencing And The Management Of Neurometabolic Disorders	New England Journal of	2016	19	

Altmetric Attention Score	Title	Journal	Year	Scopus citations	Open-access
		Medicine			
210	Evaluation Of Excess Significance Bias In Animal Studies Of Neurological Diseases	PLoS Biology	2013	91	Y
192	Probiotic Microbes Sustain Youthful Serum Testosterone Levels And Testicular Size In Aging Mice	PLoS ONE	2014	19	Y
179	Plasma Proteins Predict Conversion To Dementia From Prodromal Disease	Alzheimer's & Dementia: the Journal of the Alzheimer's Association	2014	60	
158	Prevalence Of Cerebral Amyloid Pathology In Persons Without Dementia: A Meta-Analysis.	JAMA: Journal of the American Medical Association	2015	118	
157	3d In Vitro Model Of A Functional Epidermal Permeability Barrier From Human Embryonic Stem Cells And Induced Pluripotent Stem Cells	Stem Cell Reports	2014	40	Y
157	Multiple Independent Variants At The Tert Locus Are Associated With Telomere Length And Risks Of Breast And Ovarian Cancer	Nature Genetics	2013	215	
149	Dynamics Of Extinction Debt Across Five Taxonomic Groups	Nature Communications	2016	1	Y
134	Genome-Wide Association Analysis Identifies Txnrd2, Atxn2 And Foxc1 As Susceptibility Loci For Primary Open-Angle Glaucoma	Nature Genetics	2016	19	
126	Management Of Hyperglycemia In Type 2 Diabetes: A Patient-Centered Approach Position Statement Of The American Diabetes Association (Ada) And The European Association For The Study Of Diabetes (Easd)	Diabetes Care	2012	3	
125	Crimean-Congo Hemorrhagic Fever: Epidemiological Trends And Controversies In Treatment	BMC Medicine	2011	13	Y
124	Meeting Report: First International Conference On Crimean-Congo Hemorrhagic Fever	Antiviral Research	2015	6	
123	Tt-Seq Maps The Human Transient Transcriptome	Science	2016	6	
121	Reconstructing The Population History Of European Romani From Genome-Wide Data	Current Biology	2012	28	
114	Management Of Hyperglycemia In Type 2 Diabetes, 2015: A Patient-Centered Approach: Update To A Position Statement Of The American Dabetes Association And The European Association For The Study Of Diabetes	Diabetes Care	2015	731	

Altmetric Attention Score	Title	Journal	Year	Scopus citations	Open-access
114	Queen Mary: Nobody Expects The Spanish Inquisition	The Lancet	2012	2	
112	Tocopherols And Tocotrienols Plasma Levels Are Associated With Cognitive Impairment	Neurobiolog y of Aging	2012	40	
108	The Protagoras Study To Evaluate The Performance Of The Endurant Stent Graft For Patients With Pararenal Pathologic Processes Treated By The Chimney/Snorkel Endovascular Technique	Journal of Vascular Surgery	2016	8	
97	Dysfunction Of Lipid Sensor Gpr120 Leads To Obesity In Both Mouse And Human	Nature	2012	2035	
96	Pharmacologic Interventions For Painful Diabetic Neuropathy: An Umbrella Systematic Review And Comparative Effectiveness Network Meta-Analysis	Annals of Internal Medicine	2014	24	
95	Fastkd2 Is Associated With Memory And Hippocampal Structure In Older Adults	Molecular Psychiatry	2015	8	
93	Early Patterns Of Blood Pressure Change And Future Coronary Atherosclerosis	JAMA: Journal of the American Medical Association	2014	4	
91	Kinetic Trapping Through Coalescence And The Formation Of Patterned Ag–Cu Nanoparticles	Nanoscale	2016	2	
90	Treatment Of Non-Small Cell Lung Cancer (Nsclc).	Journal of Thoracic Disease	2013	51	Y
84	Inherited Mutations In 17 Breast Cancer Susceptibility Genes Among A Large Triple-Negative Breast Cancer Cohort Unselected For Family History Of Breast Cancer	Journal of Clinical Oncology	2015	101	
82	Economic Crisis, Restrictive Policies, And The Population's Health And Health Care: The Greek Case	American Journal of Public Health	2013	60	
80	The Bite Of The Honeybee: 2-Heptanone Secreted From Honeybee Mandibles During A Bite Acts As A Local Anaesthetic In Insects And Mammals	PLoS ONE	2012	9	Y
80	Synonymization Of Key Pest Species Within The	Systematic Entomology	2015	44	
80	Astromap European Astrobiology Roadmap	Astrobiology	2016	3	
79	Actn3 R577x And Ace I/D Gene Variants Influence Performance In Elite Sprinters: A Multi-Cohort Study	BMC Genomics	2016	5	Y
79	Reinforcement Learning Agents Providing Advice In Complex Video Games	Connection Science	2014	10	
78	Microbial Symbionts Accelerate Wound Healing Via The Neuropeptide Hormone Oxytocin.	PLoS ONE	2013	37	Y

Altmetric Attention Score	Title	Journal	Year	Scopus citations	Open-access
73	The Role Of Human-Related Risk In Breeding Site Selection By Wolves	Biological Conservation	2016	0	
69	Microbial Reprogramming Inhibits Western Diet-Associated Obesity	PLoS ONE	2013	47	Y
68	The Effect Of High Vs. Low Carbohydrate Diets On Distances Covered In Soccer.	Journal of Strength & Conditioning Research	2013	8	
64	Pneumoscrotum After Tracheal Intubation	Acta Anaesthesiol ogica Taiwanica	2015	1	
63	Prevalence Of Refractive Error In Europe: The European Eye Epidemiology (E3) Consortium	European Journal of Epidemiolog y	2015	35	
63	Low Vitamin C Values Are Linked With Decreased Physical Performance And Increased Oxidative Stress: Reversal By Vitamin C Supplementation.	European Journal of Nutrition	2016	4	
63	Environmental Radioactivity Measurements In Greece Following The Fukushima Daichi Nuclear Accident	Radiation Protection Dosimetry	2012	14	
62	Musical Expertise Is Related To Altered Functional Connectivity During Audiovisual Integration	Proceedings of the National Academy of Sciences of the United States of America	2015	6	
61	The Professional Status Of European Chemists And Chemical Engineers.	Chemistry - A European Journal	2015	1	Y
60	Plasma Protein Biomarkers Of Alzheimer's Disease Endophenotypes In Asymptomatic Older Twins: Early Cognitive Decline And Regional Brain Volumes.	Translational Psychiatry	2015	2	

As Twitter was the main source of altmetrics, we examined the demographics collected from the profiles of tweeters who shared the highly mentioned papers. The overwhelming majority of tweets (over 70%) have been posted by laypersons, members of the public who do not link to scholarly literature. Approximately one quarter of the mentions came from members of the scientific community (researchers or clinicians), while around 5% came from journalists, bloggers or journal editors [48]. The mean values of mentions from each category are presented in Figure 1.

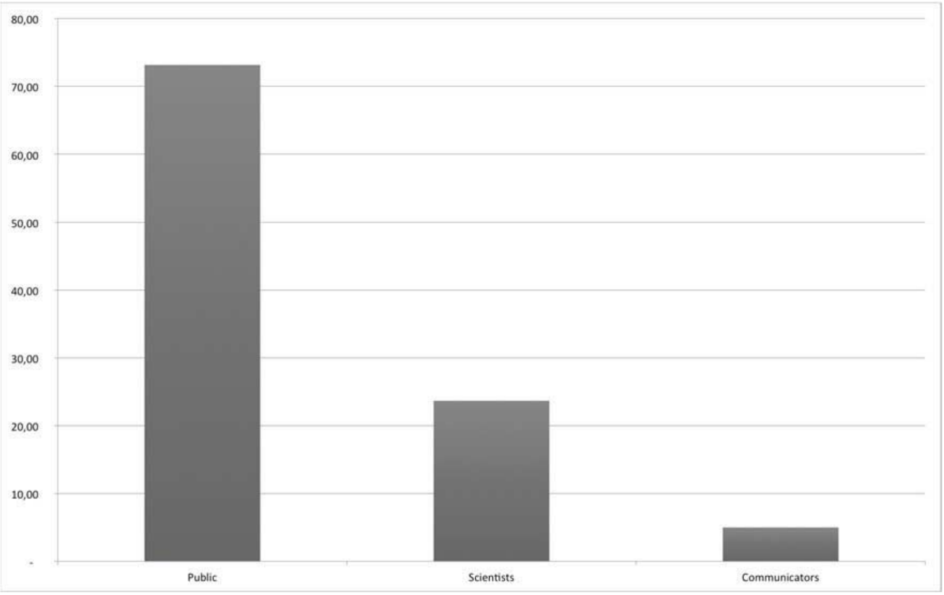


Figure 1. Twitter demographics for the fifty papers with the highest Altmetric Attention Score

5. Discussion and Conclusions

The purpose of the present paper was to present the altmetric landscape of Greece, taking as a case study the Aristotle University of Thessaloniki (AUTh) publications. Although this is a small-scale study, it seems to confirm the general patterns in the presence of altmetrics identified in previous studies. The coverage of altmetric indicators observed in this study is rather low (17%), and similar to that reported earlier in the literature. For instance, [31] found that around 15% of the publications they examined had any altmetric measures, and [37] found a coverage of 18% for the years 2009-2013. We also observed higher mentions for more recent publications, a finding in accordance with [32] and [37] who found a significant increase in altmetric coverage over the years.

Twitter emerged as the most prevalent source, accounted for over 80% of total mentions. In earlier studies Twitter was ranked second, after Mendeley, in social media activity associated with scholarly articles [27,36]. Twitter demographics revealed that overwhelmingly more attention comes from the general public, a finding that contributes to the idea that altmetrics are different from citations, as they trace a different kind of post-publication reception of research [27,31,44].

Altmetric mentions were more frequent in Medical Sciences, a pattern observed by other researchers as well, who reported that the highest share of publications with altmetric scores were the Biomedical and Health Sciences [27,30,32,37]. Moreover, although the absolute value of Arts, Humanities and Social Sciences publications mentioned in social media is relatively small, their percentage suggests a significant altmetric activity, a finding in line with that of other studies [31,36]. The present study seems to support the argument that "altmetrics scores could have an interesting added value for the analysis of humanities and social sciences, fields that traditionally are not well represented by traditional citation analysis" [31, p.20].

An interesting finding was the mentions of publications in mainstream media and policy documents. This finding indicates a broader impact of AUTh's research on both peers and the general public, difficult to be captured through other means. In recent years, societal impact of research is gaining increasing importance, as funding bodies, evaluators, and national assessment systems are interested in understanding the diffusion and use of research outputs beyond the academic audiences [49–51]. Societal impact is perceived by evaluators as an "outcome" that brings change or makes a difference in people's lives [52]. According to Wilsdon et al. [53] "research has a societal impact when auditable or recorded influence is achieved upon non-academic organisation(s) or actor(s) in a sector outside the university sector itself—for instance, by being used by one or more business corporations, government bodies, civil society organisations, media or specialist/professional media organisations or in public debate" (p.6). Non-academic organisations or stakeholders outside academia are usually involved in writing policy documents [54], thus mentions of scholarly papers in these documents signify that research preformed in AUTh influences policy formulation and policy-making process, and has tangible effects on larger society. This kind of information can help faculty and institution administrators to monitor and assess their outreach endeavours.

This is an exploratory study confined to publications authored by researchers of a single institution. As such, its findings are difficult to generalise beyond AUTh and should be interpreted with caution. A replication of the study with larger and more diverse sample of publications would be desirable, in order to get a better insight into the altmetrics landscape in Greece. Moreover, the aim of the study was to give a general overview of the presence of altmetrics for AUTh publications, and it did not go into much depth on the correlation between altmetrics and traditional citations, or about understanding how scientific publications are mentioned in social media, who publishes citations to scholarly articles in social web and why they publish them [29,41]. Large-scale quantitative analysis should be complemented with qualitative research and content analysis in order to reach safe conclusions about the kind of influence altmetrics measure, and the relationship, if any, between altmetric indicators and scientific impact [34,44].

There is evidence that an active online presence and visibility on social media networks is likely to have an impact on the attention that researchers get via altmetrics [55]. Yet, there is no relevant research regarding Greek researchers' online presence, or their attitudes in relation to scholarly communication. The only study available is that of "101 innovations of scholarly communication" [56]. Aristotle University of Thessaloniki Library & Information Centre participated in this survey with 217 respondents, but due to the small sample size no solid conclusions can be drawn. Therefore, a study focusing on Greek scholars' attitudes towards Open Science and new scholarly communication tools, as well as their online presence would reveal trends, habits and practices.

Finally, it should be noted that altmetrics may include more and different metrics than those provided by the Altmetric.com. Altmetric Explorer is not but a tool for detecting the activity around research products in online environments, and should be carefully distinguished from altmetrics as a concept. If we only see what the specific tool enables us to see, that could be a serious limitation on how we view and comprehend alternative indicators, and their potential to capture the impact of research and the multidimensionality of scholarly discourse.

Acknowledgements

The authors would like to thank Altmetric.com for providing them access to Altmetric Explorer for Institutions for the purposes of the present research.

References

[1] Togia A, Tsigilis N. Impact factor and education journals: A critical examination and analysis. Int J Educ Res. 2006;45(6):362–79.
[2] Harnad S, Brody T. Comparing the impact of open access (OA) vs. non-OA articles in the same journals. D-Lib Mag [Internet]. 2004 [cited 2016 Nov 30]; 10(6). Available from: http://dlib.org/dlib/june04/harnad/06harnad.html
[3] Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: Toward an objective measure of scientific impact. Proc Natl Acad Sci U S A. 2008; 105(45):17268–72.
[4] Brown LD, Gardner JC. Using citation analysis to assess the impact of journals and articles on Contemporary Accounting Research (CAR). J Account Res. 1985;23(1):84.
[5] Oppenheim C. Using theh-index to rank influential British researchers in information science and librarianship. J Am Soc Inf Sci Technol. 2007;58(2):297–301.
[6] Meho LI, Rogers Y. Citation counting, citation ranking, and h -index of human-computer interaction researchers: A comparison of Scopus and Web of Science. J Am Soc Inf Sci Technol. 2008;59(11):1711–26.
[7] Costas R, Bordons M, van Leeuwen TN, van Raan AFJ. Scaling rules in the science system: Influence of field-specific citation characteristics on the impact of individual researchers. J Am Soc Inf Sci Technol. 2009;60(4):740–53.
[8] Nederhof AJ. Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A review. Scientometrics. 2006;66(1):81–100.
[9] Baumgartner H, Pieters R. The structural influence of marketing journals: A citation analysis of the discipline and its subareas over time. J Mark. 2003;67(2):123–39.
[10] Brody T, Harnad S, Carr L. Earlier Web usage statistics as predictors of later citation impact. J Am Soc Inf Sci Technol. 2006;57(8):1060–72.
[11] Palmer CL, Weber NM, Cragin MH. The analytic potential of scientific data: Understanding re-use value. Proc Am Soc Inf Sci Technol. 2011;48(1):1–10.
[12] MacRoberts MH, MacRoberts BR. Problems of citation analysis. Scientometrics. 1996;36(3):435–44.
[13] Sancho R. Misjudgments and shortcomings in the measurement of scientific activities in less developed countries. Scientometrics. 1992 Jan;23(1):221–33.
[14] MacRoberts M., MacRoberts B. Problems of citation analysis: A critical review. J Am Soc Inf Sci. 1989;40(5):342–9.
[15] Peritz BC. Opinion paper on the objectives of citation analysis: Problems of theory and method. J Am Soc Inf Sci Jul. 1986;43(6).
[16] Aguinis H, Shapiro DL, Antonacopoulou EP, Cummings TG. Scholarly impact: A pluralist conceptualization. Acad Manag Learn Educ. 2014;13(4):623–39.
[17] Priem J, Piwowar HA, Hemminger BM. Altmetrics in the wild: Using social media to explore scholarly impact. arXiv Prepr arXiv12034745 [Internet]. 2012 Mar 20 [cited 2016 Nov 30]; Available from: http://arxiv.org/abs/1203.4745
[18] Björneborn L, Ingwersen P. Toward a basic framework for webometrics. J Am Soc Inf Sci Technol. 2004;55(14):1216–27.
[19] Björneborn L, Ingwersen P. Perspectives of webometrics. Scientometrics. 2001;50(1):65–82.
[20] Galligan F, Dyas-Correia S. Altmetrics: Rethinking the way we measure. Ser Rev. 2013;39(1):56–61.
[21] Howard J. Scholars Seek better ways to track impact online. Chron High Educ [Internet]. 2012; Available from: http://www.chronicle.com/article/As-Scholarship-Goes-Digital/130482/?sid=wc&utm_source=wc&utm_medium=en
[22] altmetrics: a manifesto – altmetrics.org [Internet]. Altmetrics.org. 2017 [cited 30 April 2017]. Available from: http://altmetrics.org/manifesto/

[23] Piwowar H. Introduction altmetrics: What, why and where? Bull Am Soc Inf Sci Technol. 2013;39(4):8–9.
[24] Wouters P, Costas R. Users, Narcissism and control—tracking the impact of scholarly publications in the 21 st century. In: Proceedings of the 17th International Conference on Science and Technology Indicators, Repro-UQAM [Internet]. Montreal, Canada; 2012 [cited 2016 Dec 5]. p. 847–57. Available from: http://sticonference.org/Proceedings/vol2/Wouters_Users_847.pdf
[25] Haustein S, Peters I, Sugimoto CR, Thelwall M, Larivière V. Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. J Assoc Inf Sci Technol. 2014;65(4):656–69.
[26] Sugimoto CR, Work S, Larivière V, Haustein S. Scholarly use of social media and altmetrics: a review of the literature. [Internet]. arXiv preprint arXiv:1608.08112. 2016. Available from: https://arxiv.org/pdf/1608.08112.pdf
[27] Haustein S, Costas R, Larivière V. Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. PLoS One. 2015 Mar 17;10(3):e0120495.
[28] Alperin JP. Geographic variation in social media metrics: an analysis of Latin American journal articles. Aslib J Inf Manag. 2015;67(3):289–304.
[29] Hammarfelt B. Using altmetrics for assessing research impact in the humanities. Scientometrics. 2014;101(2):1419–30.
[30] Zahedi Z, Costas R, Wouters P. How well developed are altmetrics? A cross-disciplinary analysis of the presence of ‘alternative metrics’ in scientific publications. Scientometrics. 2014;101(2):1491–513.
[31] Costas R, Zahedi Z, Wouters P. Do ‘altmetrics’ correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. J Assoc Inf Sci Technol. 2015;66(10):2003–19.
[32] Haustein S, Peters I, Sugimoto CR, Thelwall M, Larivière V. Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. J Assoc Inf Sci Technol. 2014;65(4):656–69.
[33] Bar-Ilan J, Haustein S, Peters I, Priem J, Shema H, Terliesner J. Beyond citations: Scholars’ visibility on the social Web. In: Proceedings of the 17th International Conference on Science and Technology Indicators, Montreal, Quebec [Internet]. 2012 [cited 2016 Dec 4]. Available from: http://arxiv.org/abs/1205.5611
[34] Li X, Thelwall M, Giustini D. Validating online reference managers for scholarly impact measurement. Scientometrics. 2012;91(2):461–71.
[35] Bar-Ilan J. JASIST 2001-2010. Bull Am Soc Inf Sci Technol. 2012;38(6):24–8.
[36] Costas R, Zahedi Z, Wouters P. The thematic orientation of publications mentioned on social media: Large-scale disciplinary comparison of social media metrics with citations. Aslib J Inf Manag. 2015;67(3):260–88.
[37] Erdt M, Aw AS, Aung HH, Mohammadi E, Theng Y-L. Investigating Singapore’s altmetric landscape. Proc Assoc Inf Sci Technol. 2016;53(1):1–9.
[38] Shuai X, Pepe A, Bollen J, Eysenbach G, Brody T, Harnad S, et al. How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations. PLoS One. 2012;7(11):e47523.
[39] Nielsen F. Scientific citations in Wikipedia. First Monday. 2007;12(8).
[40] Groth P, Gurney T. Studying scientific discourse on the web using bibliometrics: A chemistry blogging case study. Web Science Conf [Internet]. 2010 [cited 30 April 2017]. [8 p.]. Available from: http://journal.webscience.org/308/2/websci10_submission_48.pdf
[41] Thelwall M, Haustein S, Larivière V, Sugimoto CR, Seglen P, Wu SNC, et al. Do altmetrics work? Twitter and ten other social web services. PLoS One. 2013;8(5):e64841.
[42] Thelwall M, Wilson P. Mendeley readership altmetrics for medical articles: An analysis of 45 fields. ournal Assoc Inf Sci Technol. 2016;67(8):1962.
[43] Haustein S, Peters I, Bar-Ilan J, Priem J, Shema H, Terliesner J. Coverage and adoption of altmetrics sources in the bibliometric community. Scientometrics. 2014;101(2):1145–63.
[44] Haustein S, Larivière V, Thelwall M, Amyot D, Peters I. Tweets vs. Mendeley readers: How do these two social media metrics differ? it - Inf Technol. 2014;56(5):207–15.
[45] Mohammadi E, Thelwall M. Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. J Assoc Inf Sci Technol. 2014;65(8):1627–38.
[46] Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res. 2011;13(4):e123.
[47] Maflahi N, Thelwall M. When are readers as good as citers for bibliometrics? Scopus vs. Mendeley for LIS journals. J Assoc Inf Sci Technol. 2015;67(1):191-199.
[48] Altmetric - Who’s talking about your research? [Internet]. [cited 2017 Mar 31]. Available from: https://www.altmetric.com/

[49] Bornmann L, Haunschild R, Marx W. Policy documents as sources for measuring societal impact: how often is climate change research mentioned in policy-related documents? Scientometrics.;109(3):1477–95.
[50] Dinsmore A, Allen L, Dolby K. Alternative perspectives on impact: The potential of ALMs and altmetrics to inform funders about research impact. PLOS Biol. 2014;12(11):e1002003.
[51] Thelwall M, Kousha K. Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web. El Prof la Inf. 2015;24(5):587.
[52] Samuel GN, Derrick GE. Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014. Res Eval. 2015;24(3):229–41.
[53] Wilsdon J, Allen L, Belfiore E, Campbell P, Curry S, Hill S. The metric tide: Report of the independent review of the role of metrics in research assessment and management. Bristol: Higher Education Funding Council for England (HEFCE); 2015.
[54] Haunschild R, Bornmann L. How many scientific papers are mentioned in policy-related documents? An empirical investigation using Web of Science and Altmetric data. Scientometrics. 2017;110(3):1209–16.
[55] Bik HM, Goldstein MC. An introduction to social media for scientists. PLOS Biol. 2013;11(4):e1001535.
[56] Bosman, J., Kramer, B. 101 innovations in scholarly communication - the changing research workflow [Internet]. figshare. 2017 [cited 29 April 2017]. Available from: https://figshare.com/articles/101_Innovations_in_Scholarly_Communication_the_Changing_Research_Workflow/1286826

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-153

Benefits of Open Science: An Analytical Framework Illustrated with Case Study Evidence from Argentina

Valeria ARZA[a,1] and Mariano FRESSOLI[a]
[a] *CENIT/ UNTREF/ CONICET*

Abstract. Doing open science is to collaborate with others in a scientific endeavor and to share the outcomes of the scientific process. However, there are many dimensions of openness. Thus when analyzing concrete open science initiatives one finds a full lot of hybrid forms of openness. We identify and discuss different elements of open science and their benefits, under the contention that benefits are related to how openness is achieved. We propose a bi-dimensional framework to characterize openness along research stages, which allows anticipating expected benefits. The first dimension accounts for the characteristics of the collaboration, while the second for aspects of access to shared outputs. We illustrate our framework by discussing four Argentinean open science initiatives.

Keywords. open science, Argentina, analytical framework, benefits, case-study

1. Introduction

In modern scientific tradition, collaboration among scientists and the production of scientific public goods have been the engine for scientific production and the justification for public investment in science [1]. Scientists have been expected to collaborate across disciplines and over generations so as to contribute to a stock of interconnected knowledge needed for scientific advance. This knowledge would be publicly shared and disseminated through publications [2]. However, in practice, scientific knowledge production has been much more closed, fragmented and isolated from social problems than the idealist conception of modern science expected, as a result of three phenomena:

Firstly, scientific practice has become locked in the pursuit of personal/individual success. Scientists compete to reach priority and much of their knowledge is not transmitted. This is due to fear of competition, criticism, convention in a given field or the intrinsic characteristics of the tacit knowledge involved. Thus, although scientists publish their results, some of the relevant information to be able to construct knowledge cumulatively is not published [3]. Notoriously, negative results of experiments are not generally published. As a result, scientific production has been much less collaborative than it could have been and also less transparent. Resources become misused affecting negatively research productivity and reproducibility (and therefore reliability).

1 Corresponding author email address: varza@fund-cenit.org.ar

Secondly, assessment schemes have been increasingly influenced by marketing strategies of academic publishers, which push for the use of quantitative indicators based on citations as proxy for research quality. Thus, researchers worldwide are motivated to guide their research to areas, topics and methods that would be widely cited worldwide [4], which does not need to coincide with societal needs [5].

Thirdly, scientific policies oriented to the commercialization of scientific knowledge have increasingly locked up scientific knowledge. Political pressures in the developed world have urged scientific production to demonstrate its social and economic utility [6][7]. In turn, intellectual property mechanisms implied the protection of scientific knowledge that previously remained in the public domain [8][9], so as to motivate private sector to invest in scientific production. These practices accelerated the processes of occlusion of science; knowledge become protected and could only be used with the owner authorization, with two different effects. Firstly, scientific incentives drove scientific production away from the idea that knowledge is a public good. Secondly, the virtue of learning collaboratively and the collective creation of cumulative knowledge stocks as platforms for future knowledge production became seriously endangered, affecting the rate of invention.

In parallel to these developments, the emergence and wide diffusion of ICTs created ever increasing opportunities for sharing and collaboration, which shortened geographic, disciplinary and expertise distances. There exist various technologies, tools and infrastructure that facilitate collaborative production processes in various social spheres, and scientific production is not and exception.

These new opportunities extended the boundaries of what is feasible to share and how to do it, enlarging the potential scale and scope of collaboration and openness in science [3] [10]. For example, other resources besides publication can now be shared; such as data, lab notes, infrastructure, etc. ICTs also broadened the range of actors and expanded the possible time for collaboration; the contributions can be brief and there are tools to improve accessibility to facilitate the collaboration of actors with different backgrounds. Similarly, ICTs also broadened the range of actors and extended the possible time of collaboration, and the contributions can be brief and there are tools that facilitate the collaboration of actors with different capacities and expertise. In addition, new technologies such as big data, machine learning, massive use of sensors, drones and greater availability of low-cost scientific tools are changing the way knowledge is produced.

The experience of open source software created an important precedent for the open science movements in terms of know-how and visions. Open software become a community of practice where open access to knowledge and wider collaboration overcome the old prejudices that only competition allocates resources efficiently. Open source software demonstrated for several years now that massive and open collaboration works, and that it could even become mainstream practice in fields where information is a key input [11]. In fact, current open-minded movement, in science and other fields, is inspired by free software and open source activists.[2] They probed that sharing sums up.

[2] Efforts to apply open source ideas to science can be traced back to the late 90s and early 2000s. These include several declarations in favor of open access including the Budapest, Bethesda and Berlin declarations. The role of new creative commons licenses was also important in order to allow scientist to manage their publications. Finally, there were direct efforts from people closed to Creative Commons to create initiatives around scientific commons [14].

Open science is rapidly changing how science is being produced and used. However, as with other buzzwords and fashion terms, there is no single definition of open science. There are different understandings, motivations and potential benefits from open science [12] [13].

However, they all aim at (i) producing public goods: publications, data, infrastructure, and tools available to all; (ii) encouraging greater collaboration among scientists from different disciplines and academic fields; and (iii) broadening the diversity of science-producing actors. By these means, efficiency in scientific production is enhanced, scientific knowledge is democratized and science becomes better connected with societal needs. These potential benefits work as motivational goals for the different meanings and practices of open science. However, there is yet little understanding on mechanisms and conditions that link open science practices with potential benefits. There is no guarantee that opening up some scientific practices or outputs in some way would univocally trigger knowledge democratization, research efficiency, and social responsiveness.

This paper aims at disentangle different meanings for open science and organize them so as to relate them with claims on benefits as referred in the literature. We argue that the wide array of open science practices could be displayed in a two dimensional space, with one dimension being features of collaboration in processes and the other being the characteristics of access to outcomes. The specific location in this space anticipates different types of expected benefits. Our contention is that this analytical framework could be used as a toolbox to assess different experiences of open science around the world against their proposed goals.

Next section describes benefits as informed by the literature. Section 3 presents the conceptual framework that relates dimensions of openness and benefits. The methodology to empirically illuminate the framework is explained in Section 4. Section 5 describes the cases and Section 6 uses them to illustrate our conceptual framework. Section 7 concludes.

2. Benefits Association with Open Science as Claimed by the Literature

Different strands of the scatter literature analyzing open science practices claim they trigger several benefits, which we organized in three groups:

i. Improving scientific efficiency

One of strong argument for supporting open science practices is that they increase efficiency [15].[3] This is the result two mechanisms: a) wider availability of knowledge resources that makes research cheaper and research success more likely and b) more fluent collaboration among heterogeneous knowledge actors that amplifies collective intelligence and creativity.

[3] To increase efficiency in scientific production means to be able to achieve more or better scientific outputs (i.e. findings, publications, trained scientists) using the same amount of scientific inputs (i.e. resources). This relates to costs advantages or to learning advantages of openness and collaboration. In turn, we may refer also to dynamic efficiency when there is an increase in the likelihood of improving efficiency in the future given current state of the art.

Open access to scientific final or intermediate outcomes, increases the pool of knowledge in common use. This increases efficiency because unnecessary duplication can be more easily avoided and because researchers can explore new questions and solutions to problems by standing on the shoulders of a *taller* giant. Sharing promotes beneficial spillovers among research programs and makes the most of investment in science. [16] Moreover, open access also increases efficiency because it enables the use of computing power machine that interconnecting everything that is already known, reusing online available data to arrive to new findings. This new capacity has been sometimes name as *data driven intelligence* [15], and depends on open access to use automated tools to mine the literature. In turn, open data allows reproducibility of key research findings (and also experimental methods) that could push science ahead [17].

However, it is not just availability of publication and data that helps. Digital tools have also opened up opportunities for a greater quantity of actors from a wider community, not just professional scientists, to participate directly in scientific production overcoming restrictions imposed by physical and cognitive distance [18]. Sometimes they participate in data collection (see Galaxy Zoo, Foldit and Great Sunflower Project) proving the scientific endeavor with new cognitive and manpower resources [3] [15].

In turn, collaboration and interaction with the community improves efficiency also by boosting creativity. Open science practices sometimes involve communities participating in analytical or design research stages (not just data collection). In those cases, non-academic actors or scientists from different disciplines could contribute by drawing knowledge resources and cognitive tools from their own experience, which throw new light to research problems. Social studies of science claimed that major innovation in different fields tend to be put forward by scientists trained in different disciplines, mainly because they are not bound by professional traditions [19] . A similar phenomenon has been observed in studies about innovation [20]. Jeppensen and Lakhani (2010) [21] claim that it is not just technical marginality but also social-political marginality which may contribute with novel ideas, for similar reasons, these actors are more prone to thinking unconventionally and therefore more creative.[4] Wider participation and interaction among diverse set of actors enable the mechanism known as 'the wisdom of the crowds' [15] [22], which basically states that a group could better solve a problem than any single individual from the same group.

Finally, collaboration among scientists in the same field triggers a different mechanism to improve efficiency. When they are able to interact fluently, collective intelligence is amplified by the mere fact of being able to share, validate and quickly rule out different ideas, assumptions, hypotheses or avenues of inquiry [15].[5] This consequence of collaboration is greater when using web technologies because it gets across once unconceivable distances of time and space and ideas could quickly go back and forward feeding from the interaction, augmenting the capacity to solve problems (see, for example, the Polymath project).

ii. Improving democratization of scientific knowledge

---

4 However, greater collaboration with non-scientific actors will probably require a lot of boundary work to translate scientific information to a wider public (see [23]).

5 Nielsen, 2012 [15] argues that such amplification of collective intelligence probably works better when interactive actors share at least some cultures of practice or when they are focused on the same problem-solving strategy.

There are three complementary mechanisms through which open science practices democratize scientific knowledge: by improving access to scientific resources; by enabling the participation of a wider community in the research process; and by making science better understandable for a wider population.

Open access movements emerged as a reaction to the closure of scientific knowledge imposed behind paywall to access scientific publications. While the rate of scientific production has been always increasing, the distribution in the possibility of using such knowledge has remained unequal [24]. Aronson (2004) [25] estimated that 56% of institutions in lowest income countries have no subscriptions to international journals in medical research. Open access is potentially democratising because it reduces the costs of using and reusing the worldwide accumulation of knowledge.

Open access increases the pool of information available to anyone not just scientists. Nurses, patients, teachers, students may get to interested to learn about latest treatment of certain diseases; small businesses may get to know about relevant techniques in several application fields; etc. A recent survey to Latin American users of open access portals show that 25,2% of articles were downloaded for non-academic use; either to satisfy personal interests (10.5%) or for professional practice unrelated to scientific production (non for profit: 4.2%, private: 3.8%, public 6.7%) [6] [26].

The same could happen with open data; when properly curated and easily available, it could be used by different actors including scientist from different disciplines but also the non-scientific actors such as NGOs, firms, and just citizens (see for instance [27]).

Scientific publications and data are an outcome of research which is largely funded with public investment (see [28] [29]). Thus, it is just fair that everyone could access to the outcome of the efforts of everyone. This idea is so powerful, that open access to data and publications as a way to improve the actual use of scientific knowledge, has become the focus of several public policies initiatives promoting open science. [7]

Open access contributes to a better informed society and fosters new processes of learning [30] [31] [32], which drives us to the second claim on open science as a democratizing force. Some open science practices promote wider participation of the society in the production of scientific knowledge. One example is citizen science projects, in which non-academic actors contribute to the production of scientific knowledge in disciplines like ornithology, astronomy and environmental conservation [33]. The emergence of new digital tools and web based protocols for gathering data is widening the scope of people that can participate of scientific research beyond "a privileged few" [34]. Furthermore, participation in the production of scientific data

[6] The reported data was for Scielo based on 58957 downloads. For Redalyc, based on 22910 downloads, 16% for non-academic use, split into personal interests (7.9%) and professional non for profit: 2.9%, private: 1.9%, public 3.4%),) percentages are the following:

[7] This includes, for example, the implementation of norms that commit scientists to make their publications and data freely available; changes in the form of the evaluation acknowledging and incentivizing the publication of the datasets [35] [36]; the creation of open digital repositories; the promotion of learning in management and data analysis [37]; the creation of incentives and mechanisms of acknowledging the support of the development of an open (software and tools) infrastructure [38] [36]; and the generation of new forms of publicly communicating science [37]. In Latin America Argentina and Perú are pioneer countries to get specific legislation to guarantee open access to publicly funded scientific outputs. In Argentina open access policies are institutionalized by the enactment of The National Law for the Creation of Digital, Institutional and Open Access Repositories (approved in 20137 and fully in force since 2016). After Argentina and Peru, other countries in the region started to move along similar paths.

allow learning processes leading to the construction of new questions and skills and, eventually, the development of forms of "science by the people" (see [39]). In cases such as biohackers and do-it-yourself data recollection projects, this has challenged the hierarchies and traditional orientation of science (see [40]).

However, there are still costs associating to training potential users so they become able to enjoy all functions of shared outputs and make the most of open access. These costs are inversely related to the investment in knowledge translation and communication efforts, and as Catlin-Groves (2012) [33] suggested, more complex data involvement from non-scientific actors will demand more training. This point links to the third motivation for open science projects associated to democratization: to make science understandable for a wider public [12] by fostering scientific education [18] or by designing tools and exploring new channels to disseminate scientific information (see [41]).

There is a multiplicity of approaches to the dissemination of science [42]. Traditionally, the focus was on closing the information gap regarding scientific knowledge. In the mid-1980s, public understanding of science emerged, seeking to raise the level of scientific knowledge in the public to reverse the growing distrust of scientific expertise. In the same vein, more recently, new outreach trends have emerged, based on the use of interactive techniques (games, videos, experiments, etc.) to encourage learning during practice rather than passive information consumption [43]. According to Wiggins and Crowston (2011) [18], several open science projects can be considered as educational projects that offer formal and informal learning services. There are also other initiatives promoting scientific education directly, such as online forums and online training courses such (tutorials, massive online courses, etc.) (see for instance [44]). Some open science initiatives are starting to introduce open science tools in students' curricula as a way to improve learning and research capabilities [45].

iii. Improving research capacity to attend societal needs

There are three mechanisms claimed by the literature on how open science practices improve the research capacity to solve societal needs.

Firstly, wider access helps visibility. Open science practices could help local problems to become visible and better communicated [36]. When using digital tools and social networks the dissemination of open access information allows that problems affecting powerless actors to become better known [31]. Marginalized groups could become better endowed with knowledge resources and political support to engage negotiation with other actors like authorities, the press or other potential supporters that could contribute to solving their problems [24].

Secondly, by promoting community actors to participate in the scientific endeavor the research agenda could be better guided towards solving problems affecting that group [37] [36]. Moreover, when the community gets involved in research, people could grab from their own informed experience to offer inputs for developing solutions, improving therefore the final outcome.

Finally, the open availability of scientific resources deters private appropriation of such resources. This could contribute to find cheaper solutions to societal problems. Open access and open licenses such as creative commons avoid the creation of barriers that hamper the process of turning scientific knowledge into concrete solutions to local problems. The societal impact of scientific research depends, in turn, on the potential for promoting a wide appropriation of research outcomes, through open access and open licenses [46]. This reacts against the phenomenon known as the "tragedy of the anti-commons", which turns out when there is such an accumulation of patents on small fractions of knowledge that makes it cumbersome and highly costly to combine all of those separate elements to produce useful solutions [47]. In contrast, open science practices are then seen as an alternative business model that could solve the anti-commons problem relying on open access, worldwide collaboration and open licenses. An interesting area where there is experimentation is open source drug discovery. These projects are creating open knowledge resources that could be freely used (e.g. Open Source Malaria [48]; Open Source Drug Discovery [49][50]; Malaria Box [51][52]; among others). Most of them, and not by chance, are oriented to produce drugs for tropical disease, where the economic rewards are low and not enough for large companies to get into business.

3. Conceptual Framework to Organize Open Science Practices

This section aims at organizing the different meanings of open science in an attempt to better relate practices with potential benefits. The intention is not to create an ideal type of open science, but rather to visualize some common aspects and, at the same time, to highlight that there are different paths to improve efficiency, democratization, and societal responsiveness of scientific practice. Open science practices have been previously classified according to: i) what is shared (e.g. publication, data tools etc.); ii) how it is shared and, iii) with whom to share [38]. We build on this classification to create a bi-dimensional framework for open science practices.

We use Benkler's twofold characterization of open and collaborative knowledge production [53][54]. A first dimension characterizes how actors **collaborate** among each other to produce knowledge, and a second one characterizes **access** to shared outcomes. Thus, while the first dimension characterizes social exchange of ideas to produce knowledge, the second one refers to existing institutions that regulate the capacity of social actors to use knowledge resources.

There are different aspects of **collaboration** that matter to achieve beneficial outcomes. We claim the *scale* of participation is important to activate mechanisms such as 'the wisdoms of the crowds', or the 'collective intelligence', or to reduce the costs of producing research as in the collection of data in citizen science practices. We also argue that not just scale matters but also the level of *interaction* among participants. Process of collective intelligence, for example, will not occur iF participants do not have the chance to rapidly rule out or validate their ideas [15]. In addition, learning is always an interactive process [55] and learning is key for democratization. Moreover, also *diversity* or a participation of a wider community in the scientific endeavor matters for the democratization of science, and for other mechanisms related to efficiency such as 'the wisdoms of crowds'. Finally, another aspect related to collaboration that matters especially for societal responsiveness but also for democratization is the degree of *participation and commitment* [56].

The second dimension aims, in turn, to take into consideration aspects of **access** to shared resources. This is related to the common based characteristics of shared resources. As in open source, the backbone principle of open science practices is that scientific resources should be used and re-used by everyone. However, there are formal and informal restrictions that make this principle work to different extent in practice.

For example, open access could be restricted by different types of paywalls (e.g. subscriptions to journals or licenses to used patented knowledge) or other formal restrictions to use, distribute, reproduce, etc. [57]. There may be also informal restrictions to use and re-use knowledge resources related to the specific skills, capabilities or capital resources needed for using shared scientific outputs.

When relating this dimension of open science with potential benefits, we could realize that some specific aspects of access matters relatively more in some cases than others. *Unrestricted open access* to publications and data matters for mechanisms affecting efficiency, such as 'data-driven efficiency'. For achieving these benefits, it would be enough to guarantee open access to academic actors. However, for democratization open access is needed also for a wider community. In turn, what really matters for democratization is to improve the *accessibility* to scientific knowledge to guarantee that a larger quantity and wider variety of actors become endowed with knowledge resources. Improving the communication of science could help in this case. Similarly, for solving societal needs accessibility is needed but what becomes crucial is to augment the *visibility* of societal needs and achievements [58]. For that aim, no just communication techniques but also a diversification of channels of communication could help.

These bi-dimensional characteristics of openness and collaboration could be drawn in a Cartesian diagram such as that in Figure 1. We also include in the Figure the different mechanisms and the associated potential benefits as have been discussed above. The actual location of benefits in the Figure is speculative. It was done by imagining that each of the different aspects of collaboration and access pulls towards the vertical or the horizontal end respectively. Thus, for efficiency, we venture that the collaboration dimension is particularly important, while both of them are important for societal responsiveness but especially for the democratization of science.

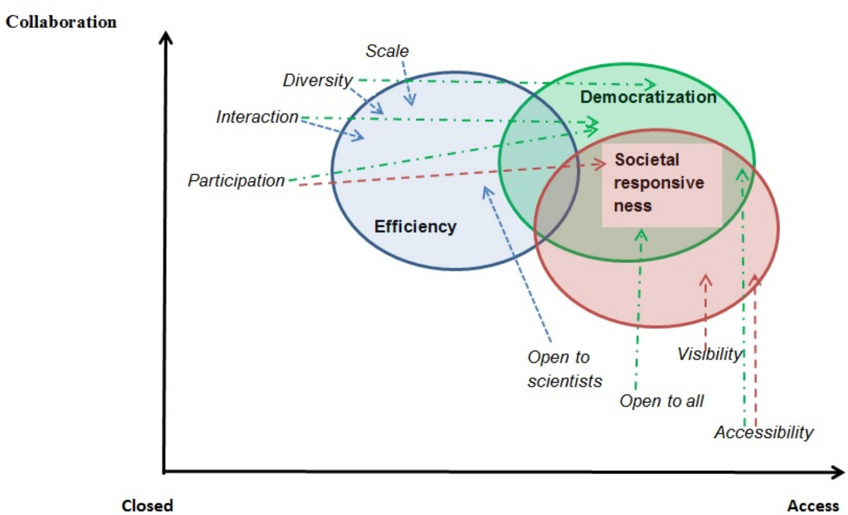


Figure 1. Two dimensions of open science

4. Analytical Methods in Case Study Research

We aim at characterizing open science initiatives in terms of the bi-dimensional framework presented in Figure 1. We will them relate location in the diagram with evidence-based benefits achieved by each experience.

Four case studies were selected from a group of open science experience identified through a national survey[8] - later enriched by online inquiries and discussions and interviews with key informants.[9] Case-study selection for this research took into account the need to cover the widest possible diversity of situations of openness processes to explore the heterogeneous spaces in which open science is being implemented in the country [59]. Among factors of heterogeneity we considered: research disciplines; socio-political contexts in which research was carried out (i.e. more or less subject to political disputes); processes of knowledge production (i.e. uni-disciplinary or transdisciplinary); techniques of participation (i.e. citizen science techniques, participatory action research, workshops, etc.); type of infrastructure (e.g. open databases; use of remote sensors, mobile applications, etc.).

The selected projects were: New Argentinean Virtual Observatory - NOVA (astronomy); Argentinean Project of Monitoring and Prospecting the Aquatic Environment - PAMPA2 (limnology); e-Bird Argentina (ornithology); and Integral Management of the Territory – IT (geography-chemistry)-.

In 2016 we carried out structured interviewed performed to one leader of each of the above-mentioned open science initiatives, to calculate the specific location of each initiative in terms of Figure 1. Closed questions were designed to assess levels of openness in a 4-points Likert scale in terms of participation, interaction, diversity of participants, access and accessibility along six different research stages: 1. Research design; 2. Collection of data; 3. Analysis; 4. Documentation and Publication; 5. Public/Social communication & engagement; and 6 Infrastructure.[10] In addition to their responses, we assess (in a 4-point Likert scale) aspects of *scale* of collaboration and *visibility* of research outcomes, based on additional data requested to interviewees (e.g. quantity of downloads, visits to their websites, followers in social networks, etc) and other secondary evidence we collected online (e.g. communication outcomes, characteristics of their website, etc.).

Case studies narratives were developed using semi-structured interviews to three referents for each project. These interviews were carried out in 2015 and they covered aspects of benefits and motivations, collaboration activities, infrastructure, financing, etc. We completed their accounts using secondary sources such as project reports, media stories and other material available primarily on the projects' website during 2016. This information is the basis for our empirical account on projects' benefits. Counterfactual information does not exist and our assessment is not based on project

[8] Survey was conducted in May 2015 using an online form to researcher form the Public Scientific Systems, largely those employed by the National Council for Scientific and Technical Research (CONICET) whose emails were available online 1463 researchers responded the survey. This implied a response rate of just 8%. The questionnaire was sent just once by email to every.

[9] We interviewed four key informants: one representative from the area of digital repositories of the Ministry of Science and Technology; one from a public-private organization specialized in R&D in ICTs (Fundación Sadosky); an advocate of open access; and a representative of a National University liaison office.

[10] The identification of research stages was inspired in RIN/NESTA, 2010[38], which includes seven different stages of the research cycle: Conceptualizing and networking, Proposal writing and design, Conducting and presenting, Documenting and sharing, Publishing and reporting, Engaging and translating, and Infrastructure.

impact systematic evaluation. Rather, it is partial and largely anecdotal based mostly on perceived benefits as expressed during interviews and other secondary evidence we got access to. More specifically, efficiency was assessed based on publications and citations to representatives of each of the initiatives[11] and other achievements such as participation in follow-up projects, quantity of data collected, etc. We do not have data to assess processes of collective intelligence, data driven intelligence, or amounts invested for each initiative. For democratization and social responsiveness our empirical data entirely relies on representatives' perceptions based on their comments during interviews.

5. The Four Case Studies

5.1 NOVA - New Argentinean Virtual Observatory

NOVA was launched in 2009. It aims at centralizing astronomical data and making them available to all users. It was created by researchers from various institutions in the country as a digital platform that aims to store and share already processed astronomical data. It facilitates collaboration of local and international astronomical community, through documentation, digitization and open access to data.

As a virtual observatory, NOVA has not required large investments in terms of infrastructure. The development of the site uses existing open source software developed by the Virtual German Observatory (GADO). In addition, an open software application to automatically upload and validate new pictures was developed locally. NOVA also developed digital manuals and organized training sessions for astronomers to encourage the use of the NOVA site.

The astronomical information stored in the database is open access and can be used by astronomers, researchers from other fields, students and the general public. However, it requires certain level of expertise to use specific software for image visualization.

The experience of NOVA and the aim of its founders to use it an educational tool triggered the conception of a related Project called Galaxy Conqueror. This is a game that motivates citizen to mark possible galaxies surfing on sky image as if it were Google map. It offers a brief tutorial that teaches basic characteristics of galaxies. Galaxies identified by users are then checked by volunteers from NOVA. Since the creation of the game in 2015, 50 new galaxies were identified. The game is part of a Citizen Science platform called Cientópolis, managed by some of the organisations that participate in NOVA.

5.2 PAMPA2 - Argentinean Project of Monitoring and Prospecting the Aquatic Environment

The Argentine Monitoring Project and Exploration of aquatic environments, better known as PAMPA2, started in 2011. It is an initiative that seeks to understand the reaction and behavior of water from lakes and ponds to certain natural and human

[11] We searched for publications authored by project leaders indexed in SCOPUS and we compared them annually before and after the beginning of each of the project.

events, to improve the design of management plans that may prevent deterioration and to preserve the population health.

PAMPA2 is an interdisciplinary network of scientists from seven different research laboratories. Lagoons are regarded by these scientists as early warning systems; thus, by analyzing them the project could contribute to detect changes that would eventually affect the whole region. This, in turn, could help to design technical and financially more viable resource management, mitigation or adaptation plans that take better care of the environment and the health of the population located in the nearby. To monitor the lagoons properly, diverse type of data are needed. So an interdisciplinary team of oceanographers, meteorologists, biologists, zoologists and engineers was formed to monitor thirteen lagoons distributed in the Pampa region during five years. Laboratory information from samples collected monthly or every six months from the lagoons is produced by participating teams.

In addition, in five of these lagoons buoys equipped with automatic sensors capable of measuring temperature, pressure, wind, rainfall, humidity, oxygen, chlorophyll and depth they have been installed. These devices are connected to a processor that stores information and then transmits it in real time to the laboratories responsible for its operation. Information can be openly accessed for free in a website but only for the present month, given restrictions in their infrastructure. Historical data generated by the sensors as well as other information generated by the project can be requested to the teams.

Originally, buoys were not designed following an open source approach; but the team is currently working in a new design based on open source software for more ambitious monitoring projects (i.e. buoys that can support more extreme environments, such as those in open seas).

Only those teams that originally formed the network participate in the design, collection and analytical phases. Actually, the project was designed predominantly by one of the networked organizations. There are no formal instances for interaction by all members: just one workshop held every year.

In terms of accessibility, one of the goals of the project was to disseminate results to a wider audience, especially the population living close to the lagoons. However, these activities were not performed so far because the team does not have the required expertise for doing public communication nor can they get the necessary resources to hire these services. Another shortcoming in terms of diffusion is that the website has not been designed so as to be easily used by outsiders.

Moreover, there is no written a protocol to allow users to work properly with the data the project produces. However, researchers do receive frequent requests from people that look at available data, for example for recreational or productive purposes.

PAMPA2 enabled increased interaction with other similar research projects around the world. It became integrated to the GLEON Network (Global Lake Ecological Observatory Network), an umbrella for organizations around the world that monitor lakes continuously through instrumented buoys. Similarly, some of the participants of PAMPA2 are also involved in the SAFER Project (*Sensing the Americas′ Freshwater Ecosystem Risk from Climate Change*), an initiative that integrates scientists from different disciplines from Argentina, USA, Canada, Chile, Uruguay and Colombia, in an attempt to define management and mitigation strategies which are both technically and economically feasible as well as culturally acceptable. This project includes several components to engage with civil society.

5.3 Integral Management of the Territory - IT

After the tragic floods in 2013 left the city of La Plata under water and caused nearly a hundred deaths, an interdisciplinary group of researchers designed an action-research project for integrated land management seeking to relieve the needs of two particularly affected areas. Thus, they expect to identify environmental consequences of this phenomenon to start thinking and developing appropriate technologies to help to reverse them. The project started in 2014.

The research group is formed by geographers, historians and environmental chemists. The project worked on two vulnerable areas that have been particularly affected by the flood events and it means to achieve an orderly, planned and sustainable land management. Two stages were involved: diagnosis and implementation of proposed solutions. At the time we did the case study they were half way through the first stage.

The neighbors participated in two ways during the first stage: in the so-called Catalyse method, by collectively designing the survey so that their views and needs were included from the beginning in the questionnaire, and in the sampling of rainwater, which measure their level Ph (to detect the acidity or alkalinity of water). These samples were then delivered to investigators.

The analysis of all collected data was performed by researchers (without the participation of the neighbors). And the obtained data have not been made public yet.

5.4 e-Bird Argentina

eBird is a citizen science project that receives bird sightings from anybody in any part of the world. The online platform was developed in the United States in 2002 by the Ornithology Laboratory at Cornell University and the National Audubon Society. In Argentina the portal started in 2013.

The platform is open access and it aims at managing and sharing online data of bird sightings made by amateur and professional watchers. eBird makes use of free software tools and online collaboration to efficiently gather, archive, and distribute information about birds to a much wider audience. eBird's regional portals are customizable in response to the need to meet the demands of local users. Each portal is integrated into the application infrastructure, and the whole database is saved in servers located in the United States.

The large amount of data collected by eBird, which contribute information about the spatial distribution of species and allow population trends to be followed, can help in the identification of important areas and sites for the conservation of birds and contribute, in this way, to the design of better plans for managing or recovering threatened species or those in danger of extinction.

Bird watchers who use eBird to report their sightings must follow a standardised protocol to load the information to guarantee the uniformity and quality of the registers. This protocol is quite dynamic and has improved with time, successively adding different characteristics that allow the watchers' data to be classified in a more precise way. Automatic control filters detect "unusual" registers. These are resent, also automatically, to the user who created them to check the data that has been flagged. If the data is confirmed to be correct, the list will then be passed to a regional expert, called an "inspector", for evaluation. If the register is rejected it will not form part of the eBird database, although it will be saved in the user's personal register. Interaction with the watchers is crucial for improving the quality of the controls, especially in regions where there is only one inspector for a very extensive area. In Argentina there are currently 20 experts who work as inspectors on a voluntary basis. Beyond the voluntary work of experts, other personnel dedicated locally to the project is minimal (four people), and as such it is entirely a citizen science project, depending on the voluntary participation of an amateur public.

The site appeals to amateur bird watchers who traditionally made their own lists of birds. One of the attractions of eBird for them is the ability to track their personal bird listings, share their data with other users, receive alerts about rare birds, upload their old sightings lists, explore information about when and where to find birds (which could be useful, for example, in planning a field trip), and play games that appeal to the competitive spirit. The site also gives users recognition for their sightings.

6. Openness and Benefits in a Bi-Dimensional Space

Table 1 assesses the different aspects of collaboration and access using information from the case-studies. This information is used in Table 2 to build indicators of expected benefits by calculating mean values on relevant aspects for each benefit as depicted in Figure 1. Figure 2 plot values from Table 2.

The four initiatives are heterogeneous in terms of dimensions of openness. Some aim at increasing collaboration while others are mostly based on improving access, accessibility and visibility of scientific outcomes. As a result, potential benefits also differ. Our goal in this section is to contrast expected benefits with actual empirical data on efficiency, democratization and social responsiveness.

Table 1. Degree of openness in different dimensions along the research cycle, 1-4 Likert Scale

		IT	**Pampa2**	**Nova**	**e-Bird**
i	Scale	1,6	1,6	2,0	3,7
ii	Diversity	2,8	2,4	1,3	3,3
iii	Interaction	2,0	2,8	2,0	2,5
iv	Participation	3,2	2,0	2,3	2,7
v	Visibility	2,0	1,0	2,0	2,5
vi	Accessibility	2,0	3,0	2,5	3,0
vii	Access by scientists	2,0	2,5	4,0	4,0
viii	Access by everyone	1,5	2,5	3,5	4,0

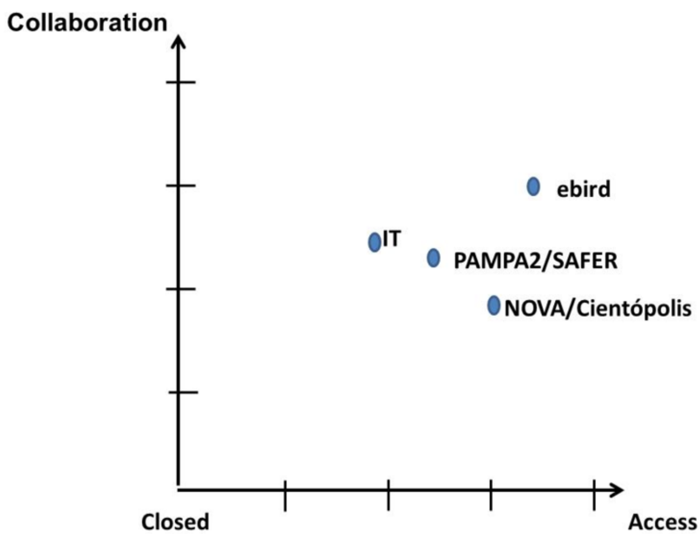
Source: Own elaboration based on responses to structured and semi-structured interviews (rows i, ii, iii, iv, vi, vii and viii) and completed with secondary information (rows i and v)

Table 2: Indicators built as mean values of different dimensions from Table 1 as informed by Figure 1

Indicators					
		IT	**Pampa2**	**Nova**	**e-Bird**
i to iv	**Collaboration**	2,40	2,19	1,92	3,04
v to viii	**Access**	1,88	2,25	3,00	3,38
i+ii+iii+vii	**Potential Efficiency**	2,10	2,31	2,33	3,38
ii+iii+iv+vi+viii	**Potential Democratization**	2,30	2,53	2,33	3,10
iv+v+viii	**Potential Responsiveness**	2,40	2,00	2,28	2,72

Source: Own elaboration based on Table 1

Figure 2. Four open science initiatives located in the two-dimension space of open science



Source: Own elaboration based on Table 2

eBird is the case that ranks the highest in most dimensions as can be seen in Table 1, which drives to high levels of expected efficiency, democratization and social responsiveness in Table 2.

In fact, the platform allowed the generation of a large database, updated daily, which can be used for the identification of areas that are critical for conservation of birds.[12] Since 2013, approximately 95% of bird's species in Argentina have been detected. Moreover, the platform enabled the interaction among professionals and

[12] In 2016 eBird international informed that more than 1/3 million eBirders have submitted 370 million bird sightings, representing 10,313 species (see http://ebird.org/content/ebird/news/2016review/ , Accessed 16th January, 2017)

birdwatchers around the country that improved the quantity and quality of stock of shared resources. This has been highlighted by representatives of e-Bird Argentina:

> "It is likely that without the volunteer work of birdwatchers and the collection infrastructure, it would not have been possible to gather this gigantic amount of data, globally"
>
> "Interaction with birdwatchers is crucial for improving the quality of controls, as experts can act as guides for inexperienced observers to improve their observational skills and to incorporate good data quality into the system."

Moreover, publications by Argentinean eBird representatives have doubled since the beginning of the project, while the annual citations to their work have more than tripled. We do not have data on the extent to which data from eBird Argentina were actually used in scientific projects, but there is anecdotal evidence that this was the case for the e-Bird international [27]. In fact, Argentinean representatives particularly value the potential use of their data for science and policy purposes.

> "For us, the project usefulness are the data. […]. These maps [of distribution of the species] have changed completely… for example, by overlapping a map of the distribution of the species made in 1975, the eBird data show which species expand their distribution or which have reduced or no longer exists."

Something similar can be said about the potential for democratization and social responsiveness. The project ranks the highest for those indicators in Table 2 because data is open access; the platform is very user-friendly, they advertise vastly their initiative (party relying on international efforts in this regard) and the infrastructure is open source.

In fact, there is evidence that the initiative had some effect on capability building. eBird familiarizes participants with the use of standardised techniques of data collection, sometimes using national contests. It increases their knowledge about birds, habitat, ecology, etc. through the interactive visualization tools, and it improves their ability to watch through interaction with regional experts. In sum, it leads to building amateur bird watchers' expertise.

Although, eBird stands out in all expected benefits and there is evidence that in fact this initiative showed great achievements, our conceptual framework is more useful to identify benefits within specific experiences than to compare outcomes across them. Every initiative may have very different goals, history, resources, etc. for comparability to make sense.

Under this light, we may say that eBird stands out in efficiency, NOVA in efficiency and democratization, PAMPA2 in democratization and IT in social responsiveness. Let's analyse whether the evidence accompanies these expected benefits.

NOVA has been very beneficial in terms of data sharing and data re-use among astronomers. The project has done a great effort to take astronomic data and images out from individual computers and to share them openly with everyone. This was recognized by the project representatives:

> "The most relevant, I think, was the VVV Survey because we uploaded 400 million positions in space with astronomical data and it was a challenge, in terms of data magnitude and the idea is that they continue to upload a lot more."

This has improved the quantity of information that is available for common use. Since the initiative started in 2009, there has been 125.075 data downloads. In 2016 there were 4171 downloads per month and in total 9400 monthly visits to the data repository. As in the case of eBird, the project leader has dramatically increased the number of annual publications and their annual citations (128% and 332% increase, respectively).

This information then agrees with efficiency as being one on the main expected benefits of this initiative. The other important expected benefit in Table 2 is democratization. Although our evidence suggests that NOVA ran a bit short in terms of amplifying its impact beyond the scientific community (e.g. their platform is not very accessible for the wide public), this has been changing lately with the creation of sister project which uses citizen science practices (Galaxy conqueror). This has improved the diffusion of astronomy among the wide public, and it might also contribute to capacity building and democratization of science, as has been observed in similar cases such as Galaxy Zoo.

> "People play but they do not forget they are in the real world with a certain purpose and, that makes it more fun" Galaxy Conqueror programmer."

Something similar happens with PAMPA2. Expected benefits (Table 2) seem to be primarily related to democratization. Evidence suggests that this is very much related to its international spin off project (SAFER), which is trend-setting in the use of community based strategies to produce knowledge and to manage natural resources. The diffusion of results to a wider audience is contemplated among the goals outlined by SAFER. For instance, this implies plans to spread the results of the project among the populations in the vicinity of the lagoons.

Not all participants of PAMPA2 participate in SAFER. Evidence based on PAMPA2 exclusively pushes us to conclude that they could do much better in terms of democratization. PAMPA2 project lacks a friendly website. The one they have, where they share buoys data, is not designed to receive inquiries from the public. Yet, researchers receive regular inquiries from people who consult the data available, such as for recreational and/or for productive purposes. As the process of opening of PAMPA2 advances, new challenges arose in diffusion of data, which in turn require better infrastructure and some precautions around the use of this data.

> "People who know that it exists and that is getting access to data that has not existed before… To those the project has helped… they could find the data useful. The only weather station from Monte Hermoso, or Pehuen-có is our station, so they enter our station to know what data are available. (..) But we also have to be cautious: it is something that we do and we release freely available but these are research stations, they are not official stations of weather forecast established by an authorized body."-PAMPA2 and SAFER Representative

PAMPA2 does seem to be doing quite well in terms of scientific performance. The group managed to create an interdisciplinary network of scientists who collaborate locally and internationally. Actually, open access to data has opened up opportunities to participate in new international projects widening local scientists' networks. The evolution of annual publications and citations has increased in 218% and 144% respectively since the beginning of the project. As a matter of fact, our interviews referred directly to the possibility of improving publications as one of the benefits they associated to the project.

> "We have already produced a special issue in a good quality high impact indexed Journal. It has data produced by our project and also previous data of the region."

> "We have co-authored several articles (…) good outcomes came out from our network and workshops, etc. We've presented our data in many congresses, seminars, conferences."

Finally, although the topic being investigated is central for communities, social responsiveness does not seem to be one of the promises of PAMPA2 (Table 2). It does not experiment with citizen science tools for data collection and it does not have a community capability building component. SAFER does, and so we could expect social responsiveness to improve as the new project develops. SAFER has an educational component and works with students from a middle school. Students collect data with the help of the IADO research team, and perform measurements of pH, water temperature, turbidity and they also take pictures. In 2014, this information was used in the school science fair. At the time of the interviews, the research team was putting together a basic kit with measuring instruments to perform periodic monitoring and if the experience were successfully concluded, they pretended to extend it to other areas.

In turn, the main expected benefit for IT is social responsiveness. The project was an ongoing project at the time of our case study, so we cannot really assess its benefits. The local community that participated in the project has increased their knowledge about territorial planning and they have also collected some data that could back their claims in the future. Thus, it does seem to be some evidence that the project is oriented towards achieving this expected benefit.

> "We propose a work methodology that brings people closer to the University. … to return the value back to people ... we, as scientists, get closer to the communities so that policy could be designed using more elements of judgment, from science, knowledge and with social support."-IT Representative

> "Then, when we go back to neighborhoods with the processed information … people become aware of what they had built … it contributes to a better knowledge balance."-IT Representative

Democratization also ranks high in Table 2, but in this case our evidence suggests that this achievement was somehow hindered by the political context in which the project emerged. Researchers said that it was puzzling to work with local communities in the context of political disputes (with local authorities), because they (the researchers) did not want to create false expectations on the outcomes of the project, while at the same time they needed to motivate the community to be part and committed to it. One specific and important problem faced by the project at the time of the interviews was political barriers to enable open access to data. Local authorities retained the right to decide when it was a reasonable (political) time to show certain results and to define what and when solutions would be carried out. They said:

> "It is not that data will not be known by people, on the contrary. But there should be some kind of mediation, so that it does not generate tensions, because data are very sensitive. The idea, of course, is always to democratize all the information that emerges from the investigation ... at different time stages, and with the needed care, so that instead of generating tensions, it could generate agreements. An untidy diffusion, generates the opposite one wants to ... that is, to get positions closer to each other."

Efficiency, in turn, does not seem to be one of the main promises in terms of how the imitative was designed and in fact, our interviews showed multidisciplinarity somehow risked the likelihood of obtaining publishable outcomes, partly because specialized journals normally belonged to certain disciplines and also because the final outcomes depended on the commitment of other researchers in a context where quality could not be cross-checked due to lack of specific skills. Annual publications and citations have increased annually since 2013, but much more moderately in comparison to the other mentioned initiatives (56% and 33%).

In sum, our conceptual framework helped us to identify the main expected benefits of each initiative which were largely validated by evidence we collected for each of the case studies: e-Bird stand out in all outcomes, but specially efficiency; NOVA in efficiency and democratization (thanks to its spin-off project), PAMPA in Democratization (thanks to its spin-off project) and efficiency and IT in social responsiveness.

7. Conclusions

This paper organizes different elements of openness in order to relate them to specific benefits claimed by the open science literature. We argue that benefits are related to the specific characteristics of the opening process. We built an analytical framework based on eight aspects (Scale, Diversity, Interaction, Participation, Visibility, Accessibility, Access by scientists and Access by everyone) of two key dimensions of open science: collaboration and access.

Using data from four case studies of open science initiatives from Argentina, we related the specific features of openness and collaboration with three reported benefits of open science as discussed in the literature: efficiency, democratization and social responsiveness. Our point is that there are several directions of openness and they could lead to different types of benefits.

The implications of these finding are that there is no need to commit to total openness to enjoy benefits of open science. There is no one single pathway to opening up; there are diverse dimensions scientists could explore, depending on their goals. Actually, in line with Whyte and Pryor (2011) [60] our findings show that researchers do not normally commit to total openness but rather attempt to open-up pragmatically, responding to specific requirements by funders or taking advantage of specific opportunities. Interestingly in our cases, once scientists start opening up some part of the research project, they later usually become interested in further the opening up other dimensions and stages of the research process, sometimes through spin-offs projects.

We believe our analytical framework could be informative for researchers, policy makers and practitioners as a guide for characterizing open science experiences and also, helping to identify specific aspects of open science practices that could be opened-up further for specific targeted outcomes.

Acknowledgements

The funding for this work has been provided through the Open & Collaborative Science in Development Network (OCSDNet) research project, supported by Canada's International Development Research Centre and the UK Government's Department for International Development. Find out more at www.ocsdnet.org and by the Interdisciplinary Centre for Studies on Science, Technology and Innovation (CIECTI) from the Ministry of Science, Technology and Productive Innovation of Argentina. In addition we would like to thank all our interviewees for their time and ideas shared with us in several opportunities.

References

[1] David, P. A., 2009. Towards a Cyberinfrastructure for Enhanced Scientific Collaboration: Providing Its "Soft" Foundations May Be the Hardest Part. . Social Science Research Network Working Paper Series 4 pp. 1-24.

[2] Merton, R. K., 1957. Priorities in Scientific Discovery. A Chapter in Sociology of Science. American Sociological Review 22 (6), pp. 635 - 59.

[3] Franzoni, C. and Sauermann, H., 2014. Crowd Science: The Organization of Scientific Research in Open Collaborative Projects. Research policy 43 (1), pp. 1-20.

[4] Ioannidis, J. P. A., 2016. Evidence-Based Medicine Has Been Hijacked: A Report to David Sackett. Journal of Clinical Epidemiology 73 pp. 82 - 86.

[5] Kreimer, P., 1998. Ciencia Y Periferia: Una Lectura Sociológica, in: M. M. (Ed.) (Eds.), La Historia De La Ciencia En El Siglo Xx Manantial, Buenos Aires.

[6] Mowery, D. C., 1995. The Practice of Technology Policy, in: P. Stoneman (Eds.), Handbook of the Economics of Innovation and Technological Change. Blackwell, Oxford.

[7] Nelson, R. R., 2004. The Market Economy, and the Scientific Commons. Research policy 33 (3), pp. 455-71.

[8] Dasgupta, P. and David, P. a., . , 1994. Toward a New Economics of Science. Research policy 23 (5), pp. 487 - 521.

[9] Mazzoleni, R. and Nelson, R. R., 2007. Public Research Institutions and Economic Catch-Up. Research Policy 36 (10), pp. 1512-28.

[10] Gagliardi, D.; Cox, D. and Li, Y., 2015. Institutional Inertia and Barriers to the Adoption of Open Science, in: E. Reale and E. Primeri (Eds.), The Transformation of University Institutional and Organizational Boundaries. Sense Publishers,

[11] Weber, S., 2004. The Success of Open Source. Harvard University Press,

[12] Fecher, B. and Friesike, S., 2014. Open Science: One Term, Five Schools of Thought, in: (Eds.), Opening Science. Springer,

[13] Grubb, A. M. and Easterbrook, S. M., 2011. On the Lack of Consensus over the Meaning of Openness: An Empirical Study. Plos One 6 (8),

[14] Schweick, C. M., 2011. Free/Open Source Software as a Framework for Establishing Commons in Science. , in: C. H. E. Ostrom (Eds.), Understanding Knowledge as a Commons. The MIT Press., Boston, Massachusetts.

[15] Nielsen, M., 2012. Reinventing Discovery: The New Era of Networked Science. Princeton University Press, New Jersey.

[16] David, P. A., 2003. The Economic Logic of Open Science and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer, in: J. M. Esanu and P. F. Uhlir (Eds.), The Role of the Public Domain in Scientific and Technical Data and Information. National Academies Press,

[17] J. K. Hartshorne and A. Schachner, 2012. Tracking Replicability as a Method of Postpublication Open Evaluation. Frontiers in Computational Neuroscience

[18] Wiggins, A. and Crowston, K., 2011. From Conservation to Crowdsourcing: A Typology of Citizen Science. Hawai.

[19] Ben-David, J., 1960. Roles and Innovations in Medicine. American Journal of Sociology pp. 557-68.

[20] Bijker, W., 1997. Of Bicycles, Bakelites, and Bulbs. Toward a Theory of Sociotechnical Change Cambrige. The MIT Press, MA.

[21] Jeppensen, L. B. and Lakhani, K., 2010. Marginality and Problem-Solving Effectiveness in Broadcast Search. Organization Science 21 (5), pp. 1016-33.

[22] Surowiecki, J., 2004. The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations Little. Anchor Books, New York.

[23] Mansell, R., 2013. Employing Digital Crowdsourced Information Resources: Managing the Emerging Information Commons. International Journal of the Commons 7 (2),

[24] Cribb, J. and Sari, T., 2010. Open Science: Sharing Knowledge in the Global Century. CSIRO Publishing, Collingwood.

[25] Aronson, B., 2004. Improving Online Access to Medical Information for Low-Income Contruies. New England Jorunal of Medicine 350 (10), pp. 966-68.
[26] Alperin, J. P., 2015. The Public Impact of Latin American's Approach to Open Access. Stanford University,
[27] Lagoze, C., 2014. Ebird: Curating Citizen Science Data for Use by Diverse Communities. International Journal of Digital Curation 9 (1), pp. 71 - 82
[28] Suber, P., 2003. The Taxpayer Argument for Open Access. SPARC Open Access Newsletter
[29] OECD, 2004. Science, Technology, and Innovation for the 21st Century: Meeting of the Oecd Committee for Scientific and Technological Policy at Ministerial Level.
[30] Gregson, J.; Brownlee, J. M.; Playforth, R. and Bimbe, N., 2015. The Future of Knowledge Sharing in a Digital Age: Exploring Impacts and Policy Implications for Development. the Institute of Development Studies, Sussex, Brighton.
[31] UN Independent Expert Advisory Group Secretary, 2014. A World That Counts: Mobilising the Data Revolution for Sustainable Development, United Nations Independent Expert Advisory Group on a Data Revolution for Sustainable Development. United Nations,
[32] World Bank, 2015. Open Data for Sustainable Development. World Bank,
[33] Catlin-Groves, C. L., 2012. The Citizen Science Landscape: From Volunteers to Citizen Sensors and Beyond. International Journal of Zoology 2012 pp. 1–14.
[34] Silvertown, J., 2009. A New Dawn for Citizen Science. Trends in ecology & evolution 24 (9), pp. 467-71
[35] OECD, 2015 Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, Paris: OECD Publishing, n.25.
[36] Stodden, V., 2010. Open Science: Policy Implications for the Evolving Phenomenon of User-Led Scientific Innovation. Journal of Science Communication 9 (1), pp. 1-8..
[37] European Commission, E., 2016. Open Innovation, Open Science, Open to the World. European Comission, Brussels.
[38] RIN/NESTA, 2010. Open to All? Case Studies of Openness in Research. Research Information Network (RIN) and National Endowment for Science, Technology and the Arts (NESTA), London.
[39] Martin, B., 2005. Strategies for Alternative Science, in: S. Fricklel and K. Moore (Eds.), University of Wisconsin Press, Madison.
[40] Delfanti, a., 2010. Users and Peers. From Citizen Science to P2p Science. Journal of Science Communication 9 (1), pp. 1-5.
[41] Burns, T. W., O'Connor, D. J., & Stocklmayer, S. M., 2003. Science Communication: A Contemporary Definition. Public Understanding of Science 12 (2), pp. 183–202.
[42] Bauer, M. W., 2009. The Evolution of Public Understanding of Science—Discourse and Comparative Evidence. Science Technology & Society 14 (2), pp. 221-40.
[43] Franco-Avellaneda, M., 2013. Museos, Artefactos Y Sociedad:¿ Cómo Se Configura Su Dimensión Educativa? . Universitas Humanística 76 (201307), pp. 97 - 123.
[44] Molloy, J., 2014. Open Training for Open Science. available at: http://science.okfn.org/2014/12/21/open-training-for-open-science/.
[45] Baden, T.; Chagas, A. M.; Gage, G. and Marzullo, T., 2015. Open Labware : 3-D Printing Your Own Lab Equipment. PLoS Biology pp. 1-12.
[46] Masum, H. and Harris, R., 2011 Open Source for Negleted Diseases: Challenges and Opportunities. Center for global health R&D Policy Assessment,
[47] Hu, M.; Schultz, K.; Sheu, J. and Tschopp, D., 2007. The Innovation Gap in Pharmaceutical Drug Discovery & New Models for R&D Success. Kellogg School of Management
[48] Robertson, M. N.; Ylioja, P. M.; Williamson, A. E.; Woelfle, M.; Robins, M.; Badiola, K. A.; Willis, P.; Olliaro, P.; Wells, T. N. C. and Todd, M. H., 2014. Open Source Drug Discovery - Limited Tutorial. Parasitology 141 (01), pp. 148-57.
[49] Årdal, C. and Røttingen, J.-A., 2012. Open Source Drug Discovery in Practice: A Case Study. PLoS Neglected Tropical Diseases 6 (9), pp. e1827.
[50] Bhardwaj, A.; Scaria, V.; Raghava, G. P. S.; Lynn, A. M.; Chandra, N.; Banerjee, S.; Raghunandanan, M. V.; Pandey, V.; Taneja, B.; Yadav, J. et al., 2011. Open Source Drug Discovery– a New Paradigm of Collaborative Research in Tuberculosis Drug Development. Tuberculosis 91 (5), pp. 479-86.
[51] Stadelmann, B.; Rufener, R.; Aeschbacher, D.; Spiliotis, M.; Gottstein, B. and Hemphill, A., 2016. Screening of the Open Source Malaria Box Reveals an Early Lead Compound for the Treatment of Alveolar Echinococcosis. PLoS Neglected Tropical Diseases 10 (3),
[52] Wells, T.; Van Voorhis, W. C.; Adams, J. H.; Adelfio, R.; Ahyong, V.; Akabas, M. H.; Alano, P.; Alday, A.; Alemán Resto, Y.; Alsibaee, A. et al., 2016. Open Source Drug Discovery with the Malaria Box Compound Collection for Neglected Diseases and Beyond. PLoS Pathogens 12 (7),

[53] Benkler, Yochai, (2006). The wealth of networks: How social production transforms markets and freedom, Yale University Press.

[54] Benkler, Y., y Nissenbaum, H. (2006). Commons-based peer production and virtue. Journal of Political Philosophy, 14(4), 394–419. http://doi.org/10.1111/j.1467-9760.2006.00235.x

[55] Lundvall, B.-Å., 1992. National Systems of Innovation : Toward a Theory of Innovation and Interactive Learning. Pinter Publishers; Distributed exclusively in the USA and Canada by St. Martin's Press, London ; New York.

[56] Arnstein, S. R., 1969. A Ladder of Citizen Participation. Journal of the American Institute of planners 35 (4), pp. 216-24.

[57] Molloy, J. C., 2011. The Open Knowledge Foundation: Open Data Means Better Science. PLoS Biology 9 (12), pp. p.e1001195.

[58] Felt, U.; Barben, D.; Irwin, A.; Joly, P.-B.; Rip, A.; Stirling, A. and Stöckelová, T., 2013. Science in Society: Caring for Our Futures in Turbulent Times. Policy briefing 50

[59] Yin, R. K., 2014. Case Study Research. Design and Methods. Fifth Edition. SAGE Publications, Thousands Oaks.

[60] Whyte, A. and Pryor, G., 2011. Open Science in Practice: Researcher Perspectives and Participation. International Journal of Digital Curation 6 (1), pp. 199-213.

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-174

Inventory of Research Data Management Services in France

Violaine REBOUILLAT[1]
Dicen-IDF, Conservatoire National des Arts et Métiers

Abstract. Data has become more and more ubiquitous in the research context. As a result, a growing number of services are created to analyze, store and share research data. This has induced the Research Data Working Group of the Digital Scientific Library (BSN10) to launch an inventory of French research data management services, funded by the Ministry of Higher Education and Research. The inventory covers all services that are managed by French institutions and infrastructures and dedicated to public research teams from all fields. Sixty services, provided by forty-five structures, have already been identified and analyzed. The paper describes the methodology used to carry out the inventory and analyzes these first results by service type, scope and research field. It also emphasizes the heterogeneous and emergent nature of the inventoried services.

Keywords. research data, data analysis, data sharing, data archiving, data discovery

1. Introduction

The paper provides an overview of research data management services developed in France. It is part of an ongoing study, funded by the Ministry of Higher Education and Research and led by the Research Data Working Group of the Digital Scientific Library[2] (BSN10). The ubiquity of data and the computing capacities to generate, mine and distribute this data increasingly influences research activities. For scientific and economic reasons, shared services are created to make data management easier for research teams. In this context, the ongoing BSN10 study consists in an inventory of French research data management services. By "research data management services", we mean the providing of human and/or technical resources for digital data management in one or more data lifecycle phases (Figure 1). A service is supplied by what we will call a "structure", i.e. by an entity in a research institution or by independent research infrastructure.

The inventory covers all services that are managed by French institutions and infrastructures and dedicated to public research teams from all fields. The aim is to:

- Better know which kind of research data services exist in France and how they are managed;
- Help researchers and librarians to identify structures most able to provide them the appropriate data management support;

[1] Corresponding Author, Dicen-IDF, Conservatoire National des Arts et Métiers, 292 Rue Saint-Martin, 75003 Paris, France; E-mail: violaine.rebouillat.auditeur@lecnam.net.
[2] Bibliothèque Scientifique Numérique (BSN) : http://www.bibliothequescientifiquenumerique.fr/

- Identify potential shortcomings and inform political stakeholders about where resources investment is needed.

The paper will first present similarly initiatives of research data services inventories, in France and abroad. It will then describe the methodology used to carry out the inventory and will report the first results. We will conclude by discussing two issues: the sustainability of data management services and their use.

2. State of the Art

Open Science and Open Data movements have increased the interest in research data management services. In 2012, the Royal Society published a report [1], in which research institutions were encouraged to implement data policy and to provide human and technical resources for research data management. Studies were also conducted with the aim to get an overview on existing data management services: Tenopir et al. [2], [3], [4] considered what types of research data services were offered by European and North American academic research libraries; they conducted surveys and showed that libraries offer more commonly informational and consultative services than technical services, such as preparing data for deposit into a repository. Delay-Artous [5] focused on research data services in the humanities and social sciences; she drew a graphical representation of initiatives and stakeholders, while emphasizing how quickly this representation would be obsolete. This point may explain why mappings of research data services also take the form of regularly updated catalogues. The Registry of Research Data Repositories (Re3data) is one of the best internationally known [6], [7]. It focuses on data dissemination and preservation infrastructures. In the Netherlands, the Leiden University has created a catalogue for data management facilities for researchers: the Leiden Research Data Information Sheets [8]. Its scope is larger than the Re3data's, since the catalogue includes not only research data repositories, but also research data archives or tools for data management plan assistance.

3. Methodology

The inventory is based on a four-step methodology.

The first step consisted of identifying structures which provide research data management services. For that purpose, we used different approaches: we investigated conferences on research data topic; we consulted associations of scientific and technical information professionals (i.e. EPRIST [3], Couperin [4] and the URFIST network[5]), because their members are often involved in data management services in their own institutions; sometimes, during meetings, we were also informed by services managers about the existence of other services.

---

[3] EPRIST is the association of scientific and technical information staffs from the French public research organizations: http://www.eprist.fr/.

[4] Couperin is a consortium of higher education and research institutions for access to digital publications: http://www.couperin.org/.

[5] The URFIST network is composed of seven regional scientific and technical information training units: http://urfistinfo.hypotheses.org/.

The second step focused on establishing a typology in which the already identified services were sorted by function. We built on the research data lifecycle (Figure 1).

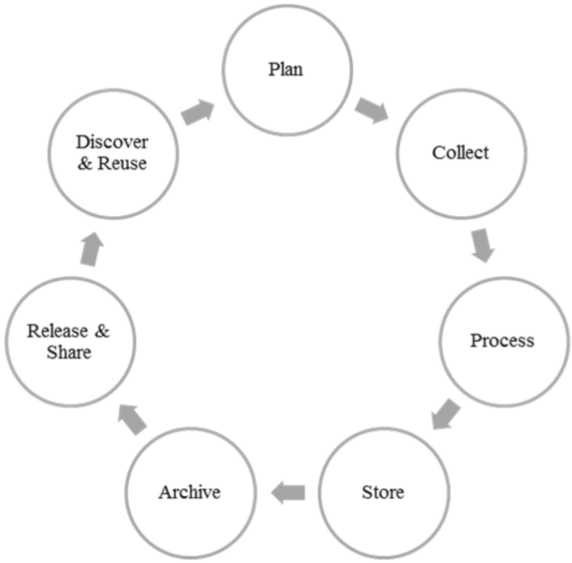


Figure 1. Research data lifecycle.

The typology of services consists of the following categories:

- Information: all websites that aggregate information about news, services, tools or good practices on research data topic;
- Training: face-to-face or distance training services in one or more research data management aspects;
- Support: human resources with IT, documentary, archival and/or legal expertise, which offer to research teams personalized assistance in research data management;
- Data management tool: a tool enabling data traceability such as data management plans (DMP) or persistent identifiers;
- Acquisition platform: an infrastructure providing human and technical resources to support research teams in data collection;
- Computing center: an infrastructure that provides to research teams high-performance computing resources for simulation, modeling and analysis;
- Data registry: an online database that describes scientific datasets;
- Data repository: an online platform that enables users to release or to discover scientific datasets;
- Archiving platform: a platform dedicated to long-term archiving of digital research data.

In the third step, we devised an analysis matrix for each service type, in order to collect information about its identity (name, start date, contact address, etc.), its management (supervisory institutions, human resources, sustainability, etc.), its functional features (compliance with standards and other technical aspects specific to each service type) and its use (targeted discipline and audience, access conditions, economic model, frequency of use, etc.).

The fourth step of our methodology involved the analysis of the identified services. For each one, we first gathered online documentation (on the website, in papers, reports or communications); then we contacted the service managers for an interview, in order to get additional information. The collected material was recorded in the matrix.

4. First Outcomes

The results reported in this paper should be considered as a photograph of the French data services landscape. They do not match the complete landscape, but only the amount of services inventoried between November 2015 and March 2017. Within that period, 60 services, provided by 45 structures, were actually identified and analyzed. These results are sorted by type in Figure 2.

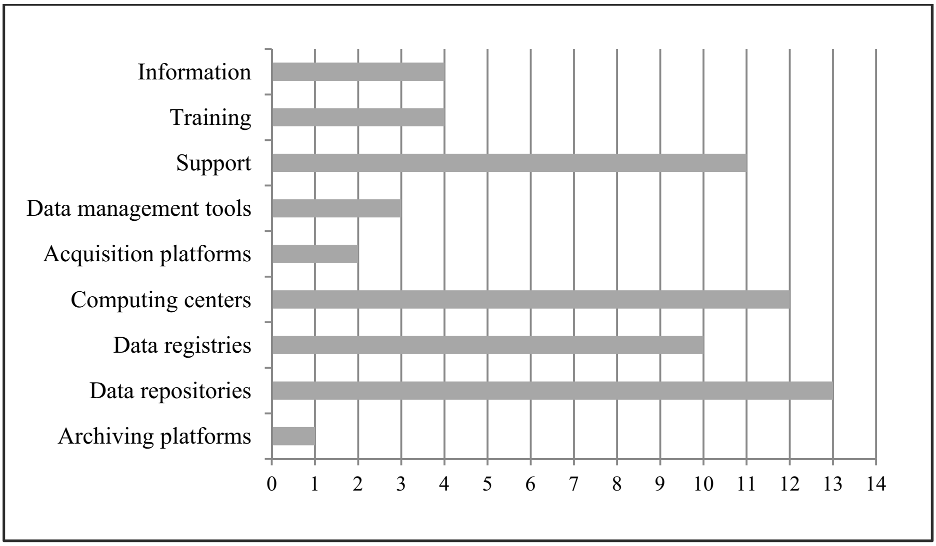


Figure 2. Types of the inventoried services.

Table 1 provides a more detailed insight into results distribution: for each service type, results are sorted both by disciplinary and by geographical scope. Most services (93%) are intended for an institutional or a national audience. Only four services have an international scope: the registry of published astronomical catalogues and tables VizieR and the astronomical objects repository SIMBAD, the sea data repository SEANOE and the archeological data repository ArkeoGIS. Furthermore, institutional services are the most numerous: they represent 63% of the total inventoried services. From a disciplinary perspective and considering the current state of the inventory, French data services seem to cover equally the different research fields: the services in

Human and Social Sciences (HSS) actually number 16; the services in Life and Physical Sciences (LPS) 21; the multidisciplinary services 23.

Table 1. Scopes and research fields of the inventoried services

		Institutional Scope	National Scope	Total
Information	HSS[6]	1	0	1
	LPS[7]	1	2	3
	Multidisciplinary	0	0	0
Training	HSS	2	0	2
	LPS	0	0	0
	Multidisciplinary	1	1	2
Support	HSS	4	0	4
	LPS	2	0	2
	Multidisciplinary	5	0	5
Data Management Tools	HSS	0	0	0
	LPS	1	0	1
	Multidisciplinary	1	1	2
Acquisition Platforms	HSS	0	1	1
	LPS	0	1	1
	Multidisciplinary	0	0	0
Computing Centers	HSS	0	0	0
	LPS	2	0	2
	Multidisciplinary	7	3	10
Data Registries	HSS	2	0	2
	LPS	4	2	6
	Multidisciplinary	1	0	1
Data Repositories	HSS	0	5	5
	LPS	3	0	3
	Multidisciplinary	1	1	2
Archiving Platforms	HSS	0	0	0
	LPS	0	0	0
	Multidisciplinary	0	1	1
	Total	**38**	**18**	**56**

More generally, we can draw two conclusions about today's French landscape of research data management services. The first conclusion is that it is an emergent landscape: 36% of the services we analyzed have been created after 2014; 30% between 2010 and 2014. Second conclusion is the heterogeneous nature of the landscape: variety and diversity prevail, which reflects the proximity of the services with research communities and institutions. This heterogeneity is visible through the

[6] HSS: Human & Social Sciences
[7] LPS: Life & Physical Sciences

different types and scopes of services, as described above. It is also visible through the various career profiles of services managers: these may be librarians, archivists, IT staffs or researchers. We noticed that the closer the service is to research teams, the more often this service is managed by researchers or research engineers.

Three examples among the inventoried services may illustrate the observations made above.

The first example is ECOSCOPE Metadata Portal [9], a non-profit data registry of biodiversity research observatories, launched in 2016. ECOSCOPE is a French observation data infrastructure for biodiversity research, maintained by the Foundation for Biodiversity Research (FRB). The ECOSCOPE Metadata Portal was created to make known the data produced by the 200 French observation observatories in the field of biodiversity. The aim is to foster data sharing between these observatories, which currently tend to work as closed circuits, using few external data. The portal also complies with the INSPIRE Directive. It will be harvested by national and international catalogs, such as Géocatalogue[8], the EU BON Portal[9] and GBIF[10]. In March 2017, ECOSCOPE described 52 datasets. The metadata schema is compliant with the Ecological Metadata Language (EML). ECOSCOPE ensures the quality of the metadata recorded by observatories, before editing them. The main challenge of ECOSCOPE is fostering uptake by the French biodiversity community. At this time, only eleven observatories described datasets in the registry. Regarding the consultation rate, ECOSCOPE does not have any tool yet to measure it.

The second example is the CINES archiving platform [10]. The National Computing Center for Higher Education (CINES) provides resources for long-term archiving of digital data to the French research community. It reports directly to the Ministry of Higher Education and Research. Its staffs have an archival and IT expertise. The CINES selects sustainable file formats, applies the persistent identifier system ARK and is compliant with the Dublin Core metadata schema. The archiving quality is certified by the Data Seal of Approval. The main challenge is that the CINES is the only stakeholder in France to provide archiving resources for digital scientific data. As a result, small research projects cannot access the CINES services, as these are subject to charges and give priority to large data volumes.

The last example is PUDC [11], a platform for human and social sciences data, which we classified in the category “support” of the services types. Created in 2011, the platform is based at the University of Caen Basse Normandie and is supported by the research infrastructure PROGEDO. It is managed by a lecturer in sociology and a research engineer specialized in statistics. Their role consists in offering personalized assistance to PhD students and researchers in human and social sciences from the University of Caen Basse Normandie. They help them to reuse quantitative data, especially these from databanks of national and international social surveys. When data do not exist, they provide them methodological advices to generate, analyze and interpret their own data. In 2016, about sixty persons used the PUDC services. The managers however report that researchers do not turn spontaneously to them; extensive communication efforts are necessary. Yet there is a need, especially for support in data analysis. The core challenge of PUDC is thus to keep attracting users, in order to ensure its sustainability.

[8] Géocatalogue: http://www.geocatalogue.fr/
[9] EU BON European Biodiversity Portal: http://biodiversity.eubon.eu/
[10] Global Biodiversity Information Facility: http://www.gbif.org/

5. Conclusion

We report here two issues that occurred in the course of the inventory.

The first issue is about the sustainability of data management services. We wondered what financial and human resources services had and, consequently, if these resources were sufficient to guarantee data preservation. We noticed that funding was a major concern. Services commonly have a long-term mission; yet, they often depend on the research funding system, which consists of short-term grants. They are compelled to constantly seek new sources of funding. We tried to determine how many services were sustainable. We considered a service as sustainable, when funded over the long-term and managed by a structure which has a legal entity. It appeared that 37% of the services we inventoried were not sustainable, i.e. more than a third. This situation has an impact on human resources: in most cases, actually, services managers are small teams of two or three staffs. Technical resources can also be restricted, which may impact data preservation. For instance, due to limited funding, the data repository ORTOLANG is only able to transfer one part of its datasets to CINES for archiving.

The second issue concerns the use of the services. Are the services actually used by researchers? Are researchers aware of their existence? How do managers foster the use of their services? During the analysis phase of the inventory, we had difficulty collecting quantitative data on the use of the services. Indeed, the services managers did not necessarily have usage statistics. Most of the time, only those who requested from researchers the creation of a user account to access the service were able to provide us figures. Moreover, research teams are generally not aware of the existence of data management services. A shift is still visible between data processing services and data releasing services: acquisition platforms, computing centers and archiving platforms are relatively well integrated in research practices, whereas services related to open science, such as data repositories, are less used, probably because they are not currently at the heart of researchers' concerns. Journal articles and other forms of publication are more valuable than research data in nowadays evaluation system.

Both issues are decisive for the future of research data management services. If governments and research funders keep fostering open science, financial resources will be invested in research data services and research teams will be encouraged to use these services. But without political incentives, it may be more difficult for services to gain legitimacy.

Acknowledgements

The study was enabled by a grant from the French Ministry of Higher Education and Research. We would like to thank the Ministry, as well as the Research Data Working Group of the Digital Scientific Library for supporting our work. We would also like to thank all services managers, who took time to answer our questions.

References

[1] The Royal Society. *Science as an Open Enterprise*. The Royal Society Science Policy Centre, 2012. Available from: https://royalsociety.org/~/media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

[2] C. Tenopir, B. Birch, S. Allard. *Academic Libraries and Research Data Services: Current Practices and Plans for the Future*, Association of College & Research Libraries, 2012. Available from: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

[3] C. Tenopir, R. Sandusky, S. Allard, B. Birch. Research Data Management Services in Academic Research Libraries and Perceptions of Librarians, *Library & Information Science Research* **36**, Issue 2 (2014), 84–90. Available from: http://www.sciencedirect.com/science/article/pii/S0740818814000255

[4] C. Tenopir, S. Talja, W. Horstmann, E. Late, D. Hughes, D. Pollock, B. Schmidt, L. Baird, R. Sandusky, S. Allard, Research Data Services in European Academic Research Libraries, *LIBER Quarterly* **27**, Issue 1 (2017). Available from: http://www.liberquarterly.eu/articles/10.18352/lq.10180/

[5] C. Delay-Artous, Où sont les données de la recherche ? Essai de cartographie, *Systèmes d'organisation des connaissances et humanités numériques*. ISTE Editions, London, 2017. Available from: https://halshs.archives-ouvertes.fr/halshs-01369745

[6] Registry of Research Data Repositories (re3data.org). Available from: http://www.re3data.org/

[7] M. Kindling, H. Pampel, S. Sandt, J. Rücknagel, P. Vierkant, G Kloska, M. Witt, P. Schirmbacher, R. Bertelmann, F. Scholze, The Landscape of Research Data Repositories in 2015: A re3data Analysis, *D-Lib Magazine*.**23**, Issue 3/4 (2017). Available from: https://doi.org/10.1045/march2017-kindling

[8] Leiden Research Data Information Sheets. Available from: https://vre.leidenuniv.nl/vre/lrd/Pages/information-Sheets.aspx

[9] ECOSCOPE Metadata Portal of Biodiversity Research Observatories. Available from: http://ecoscope.fondationbiodiversite.fr/ecoscope-portal

[10] CINES Archiving Platform. Available from: https://www.cines.fr/archivage/

[11] Plateforme Universitaire de Données de Caen (PUDC). Available from: http://www.unicaen.fr/recherche/mrsh/pudc

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-182

Increasing Papers' Discoverability with Precise Semantic Labeling: The sci.AI Platform

Roman GURINOVICH [a,1], Alexander PASHUK [a], Yuriy PETROVSKIY [a,b], Alex DMITRIEVSKIJ [a], Oleg KURYAN [a], Alexei SCERBACOV [a], Antonia TIGGRE [a], Elena MOROZ [a, c], Yuri NIKOLSKY [a]
[a] *sci.AI, Xpansa, Mustamae Tee 5, 10616, Tallinn, Estonia*
[b] *Odessa National Medical University, Valikhovs'kyi Ln, 2, Odesa, Ukraine, 65000*
[c] *IITP RAS, Bolshoy Karetny per. 19, build.1, Moscow 127051 Russia*

Abstract. The number of published findings in biomedicine increases continually. At the same time, specifics of the domain's terminology complicates the task of relevant publications retrieval. In the current research, we investigate influence of terms' variability and ambiguity on a paper's likelihood of being retrieved. We obtained statistics that demonstrate significance of the issue and its challenges, followed by presenting the sci.AI platform, which allows precise terms labeling as a resolution.

Keywords. Supervised semanticization, word sense disambiguation, paper influence and citation, biomedical text processing, named entity recognition

1. Key objectives of the study and its significance

Over the last two decades, life sciences articles have become substantially more complex, reflecting technological evolution, particularly OMICs experimentation, increasing cooperation between multiple institutions, and involving more advanced math and statistics applied to the data. In many publications, plain unstructured text is supported by algorithms, code, and multiple files of processed and raw datasets with annotated metadata and graphs. With such enhancements in place, experimental articles, per se, might become a driving force of the Literature Based Discoveries (LBD) [1]. Recently, the whole field of "meta-analysis" has arose to describe "dry lab" studies on normalization, unification, and analysis of many similar datasets derived from different labs and projects. However, a number of experimental papers are missed, because they cannot be retrieved from the body of literature by keywords search. Needless to say that scientists are keenly interested in higher discoverability of their published research and referencing to their findings, as citation index becomes an increasingly prevalent metrics in evaluation of their work. The issue can be addressed with proper semantic labeling of the texts as the very first step in global analysis of the research reports.

[1] Corresponding Author: roman.gurinovich@xpansa.com

The state-of-the-art section reflects the ways published papers are algorithmically processed in text mining applications. Such computations rely on preexisting, statistically supported information, while text mining of the scientific literature targets novel findings. This leads to Information Retrieval (IR) and then Information Extraction (IE) underperformance when applied to scientific literature.

The objective of the paper is to consider just one issue of many in biomedical texts processing: false terms recognition caused by ambiguity of the concepts' names and multiple-terms spelling variants. This leads to at least two undesirable effects:

1. Lower recall rate when search engines and aggregators retrieve articles, so the target audience does not receive a full set of relevant papers.

2. Retrieving a paper that is irrelevant to the sought-for concept. For example, reader can query 'cat' with the 'cat' animal in mind but receive texts about the 'CAT' gene.

We address:

a. A global need of initial transformation of the plain text to a machine-readable format; and

b. Uncertainty issue mentioned above;

by releasing the sci.AI system. This system combines automatic metatagging and manual validation of the results by the author or reader and supports generating semantic structures during writing and editorial processing. Human validation eliminates almost any possibility of term misinterpretation in the following IR and IE tasks, because authors can be expected to have a comprehensive understanding of the concepts being mentioned in their papers and can supervise the machine's results.

2. State of the art in the field of biomedical texts semanticization

Computational Linguistics is one of the most dynamic fields with innovations being released almost monthly. Unfortunately, there is no solid state-of-the-art solution for biomedical text labeling yet that unites all the latest advances in each subfield into a single package.

The first subfield is metatagging standards and paradigms. Semantic Web's objects, concepts, knowledge association, and data representation utilize schema.org vocabularies and W3C RDF/XML [2]. A current limitation is the lack of a similar single schema for the life sciences. Former related initiatives here are W3C Scholarly HTML [11] and JATS4R [12]. Still both schemas do not provide a standard namespace for biomedical concepts labeling.

The second subfield is terms labeling or Named Entities Recognition.

Just as in many other areas, deep learning and neural networks (NN) methods are increasingly popular for extracting information from professional texts [7, 8]. NN algorithms are rather generic and can be applied for the text analysis in unsupervised fashion (i.e., to a variety of texts without establishing prior rules generation). However, its precision and recall hardy depend on statistical data and cannot be considered as stable solution for concepts and challenges that have appeared recently. Still, NN demonstrates the highest recognition rates [16, 17] among automatic methods.

Then there are methods of increasing precision by reducing concepts ambiguity by connecting the same concepts in various ontologies.

UMLS (Unified Medical Language System), in combination with MetaMap, provide graph-like links between objects from various ontologies, as widely-accepted solution for the Word Sense Disambiguation (WSD) task in the biomedical domain. Essentially, UMLS represents a metaontology of biomedical terms and concepts. UMLS is extensive and well supported by NIH, and it is in constant development. Future considerations include possibly connecting this data to the sci.AI application. Currently, there are several limitations:

1. Lack of details for specialised ontologies, such as Uniprot and ChEBI.
2. Focus on the indexing task for the NCBI. This leads to the same dropdown in precision and recall of post-publication text processing.
3. It is not a simple plug-and-play solution for the publishing industry [6].

SciGraph by the Neo4j [13] framework allows objects to be interconnected and can be used as a technical basis for future metaontologies.

3. Design and Methodology

Resolving terms' ambiguity and variability represents a significant challenge in text processing. Here, we investigated how these factors affect the paper's influence. Such causality is assumed based on the logic that findings described in the paper can be reused and cited—only if the paper will be discovered by the readers first. To model paper's influence potential mathematically, we defined Paper's Influence as a function of a variable we called Discoverability.

$$Paper's\ Influence\ =\ f(Paper's\ Discoverability) \tag{1}$$

The paper's potential for influence greatly depends on how accurately search engines and aggregators solve the IR task. For further explanation, we will continue with our query scenario from above. A reader is discovering the paper about animal 'cat' after querying string Q_i 'genes of cat', including term t_{ij} 'cat' while meaning biomedical concept c_i 'Felis catus', corresponding to object '9685' in the ontology [14]. Concept c_i then can be referenced with any term (spelling variant) t_{ij} of the set T_i

$$T_i\ =\ \{t_{i1}, t_{i2}, \ldots, t_{iN}\}\ \in\ c_i \tag{2}$$

If we assume that a reader will read the paper if the search engine returned it in response to the query Q_i, then "discoverability" is a synonym of "retrieval". We can then apply two major IR metrics, recall and precision:

$$recall\ =\ \frac{number\ of\ relevant\ papers\ retrieved}{number\ of\ all\ existing\ papers\ about\ the\ concept} = \\ = P(retrieved\ papers\ \supseteq T_i\ |\ existing\ relevant\ papers \supseteq c_i) \tag{3}$$

$$precision\ =\ \frac{number\ of\ relevant\ papers\ retrieved}{number\ of\ retrieved\ papers} = \\ = P(relevant\ papers\ among\ retrieved \supseteq c_i\ |\ retrieved\ papers\ \supseteq T_i) \tag{4}$$

As long as such cases could be found across biomedical terminology, when concept can have several synonyms (variable terms) or single term can refer to several concepts (ambiguous terms), probabilistic precision and recall can be calculated based on the numbers of possible outcomes when querying $Q_i \supseteq t_{ij}$. For example, texts "TNF alpha", "TNFa" and "TNF α" are variants of the object Uniprot [P01375]. This means that if search engine was queried with "TNF alpha", an ideal result would return all documents that contain all three variants. Still, due to existence of the several variants, the there is a probability ≥ 0 that some of them will not be considered.

We can estimate chances of such event using a basic definition of the probability as the ratio of the number of favorable outcomes to the total number of possible outcomes. Term's ambiguity and variability define those numbers of possible outcomes. Finally, when we know precision and recall of the paper's retrieving while searching for the concept c_i, we can answer specific questions about discoverability of the paper in some kind of progression order.

Question 1. How many papers out of all existing literature about concept c_i can be retrieved, when there is set T_i, all terms of which refer to this concept c_i ?

$$\begin{aligned} & recall(variability) \\ & = \frac{number\ of\ retrieved\ papers\ with\ t_{i1} + \ldots\ + number\ of\ retrieved\ \ papers\ with\ t_{iN}}{total\ number\ of\ relevant\ existing\ papers\ about\ c_i} = \\ & = \frac{\Sigma\ retrieved\ papers\ with\ T_i}{total\ \Sigma\ papers\ relevant\ \ c_i} \end{aligned} \tag{5}$$

Question 2. How many papers out of retrieved and containing terms from T_i mention concept $c_i$specifically? As long as only recorded synonyms are proved to exist, we can assume that all synonyms from the ontology and generated variants constitute a full dictionary of the concept, and

$$\begin{aligned} & precision(variability) \\ & = \frac{number\ of\ relevant\ papers\ about\ \ c_i}{number\ of\ retrieved\ papers\ with\ t_{i1} + \ldots + number\ retrieved\ papers\ with\ t_{iN}} = \\ & = \frac{\Sigma\ relevant\ out\ of\ retrieved\ papers\ about\ c_i}{\Sigma\ retrieved\ papers\ with\ T_i}, \\ & \forall t_{ij}, T_i\ \in\ c_i, \\ & \Rightarrow \Sigma\ relevant\ papers\ with\ c_i\ = \Sigma\ retrieved\ papers\ with\ T_i\ \Rightarrow \\ & \Rightarrow precision(variability)\ = 1 \end{aligned} \tag{6}$$

This means that precision for the specific concept does not depend on the number of variants, as long as we assume that all variants are describing the same concept in the event.

Question 3. There is term t_{ij}which refers to the concept c_i or another concept c_m. How many papers out of retrieved and containing term t_{ij}, are talking about the concept $c_i$exactly?

$$\begin{aligned} & precision(ambiguity) = \\ & = \frac{number\ of\ relevant\ papers\ about\ \ c_i}{number\ of\ retrieved\ papers\ about\ \ c_i + \ldots + number\ of existing\ \ papers\ about\ \ c_m} = \end{aligned} \tag{7}$$

$$= \frac{\Sigma\, relevant\ out\ of\ retrieved\ papers\ about\ \ c_i}{\Sigma\, retrieved\ papers\ \ c_i \cap \ldots \cap\ c_m = \{t_{ij}\}}$$

Question 4. Did we receive all papers containing term t_{ij}? (Answer: Yes, obviously. Continuing to ask this question is important for keeping track of the general recall and precision derivation)

$$recall(ambiguity) = \frac{number\ of\ retrieved\ papers\ with\ t_{ij}}{total\ number\ of\ relevant\ existing\ papers\ with\ t_{ij}} = 1 \tag{8}$$

Question 5. What is the overall probability of retrieving a relevant paper for the concept c_i that has many variants $\{t_{i1}, t_{i2}, \ldots, t_{iN}\}$ and some of them $c_i \cap \ldots \cap\ c_m = \{t_{ij}\}$ are also found in the other concepts?

This means a probability of two independent events: A = the concept has spelling variants, and B = those variants can be found in several concepts. Therefore P (A and B) will be multiples of the probabilities above:

$$\begin{aligned} &recall(variability\ and\ ambiguity) = recall(variability) * recall(ambiguity) = \\ &= \frac{\Sigma\, retrieved\ papers\ with\ T_i}{total\ \Sigma\ papers\ relevant\ \ c_i} * 1 \end{aligned} \tag{9}$$

$$\begin{aligned} &precision(variability\ and\ ambiguity) = \\ &= precision(variability) * precision(ambiguity) = \\ &= \frac{\Sigma\, relevant\ out\ of\ retrieved\ papers\ about\ \ c_i}{\Sigma\, retrieved\ papers\ \ c_i \cap \ldots \cap\ c_m = \{t_{ij}\}} * 1 \end{aligned} \tag{10}$$

If operating only with the number of variants per concept, then prior probability of variants occurrence can be approximated as uniformly distributed, as long as actual frequency of terms occurrence in the papers will be retrieved in the next steps. This means that, in the first approximation, occurrence = {True, False} of the term can be sufficient variable to estimate the minimum expected probabilities:

$$prior\ recall(variability\ and\ ambiguity) = \frac{1}{Number\ of\ terms\ variants\ per\ concept} \tag{11}$$

$$\begin{aligned} &prior\ precision(variability\ and\ ambiguity) = \\ &= \frac{1}{Number\ of\ concepts\ per\ single\ term\ variant} \end{aligned} \tag{12}$$

As long as the Number of the terms per concept ≥ 1 (each object has at least a main name) and Number of concepts per single term variant ≥ 1 (each term is related to at least one object) and they are in the denominator of the retrieving probabilities above—terms' variability and ambiguity will always reduce (at least, will not increase) recall and precision, respectively, when searching for the paper.

In order to estimate influence of the existing terms' uncertainty on the papers discoverability, we have searched for:

1. homographs across Uniprot, ICD-10, ChEBI, MeSH, Drugbank, and Gene Ontology databases;

2. possible spelling variants for the same objects;

3. actually used terms' variants in the 26782464 Pubmed, 26404 Bioline and 5426 eLife papers.

MeSH Categories G–Z were not analysed because they contain generic objects, such as countries' names, which are out of scope of sci.AI semanticization for now.

Our research is ongoing and the latest results can be found on the sci.AI webpage [20].

Table 1. Variability in the ontologies and influence on paper's recall

Ontology name	Ontology size, n of IDs	Number of synonyms	Average number of synonyms	Number of synonyms that were found in the papers	Number of synonyms with variants	Average number of synonyms with variants	Number of variants that were found in the papers	Maximum number of found variants per concept	The smallest expected recall
Uniprot	553667	1018837	1.8402	702942	1729359	3.1235	936673	84	0.0119
ChEBI	104854	201061	1.9175	175750	510817	4.8723	67159	81	0.0123
Gene Ontology	46517	173156	3.7224	47073	287903	6.1914	54711	94	0.0106
Drugbank	8221	28980	3.5251	15724	154704	18.8204	28445	243	0.0041
ICD-10	11420	20728	1.8150	9680	30463	2.6675	14883	19	0.0526
MeSH (A-F tree)	23716	199486	8.4114	139078	309189	13.0459	170898	210	0.0048

We had not only considered synonyms that exist in the ontologies but also created a rules-based term variant generator (TVG) to cover a case when the same object, Uniprot [P01375], might be written as "TNF alpha", "TNFa", or "TNF α" in a paper. Next generating techniques groups were utilized:

- orthographic;
- abbreviations and acronyms;
- inflectional variations;
- morphological variations;
- structural recombinations [4, 5, 6].

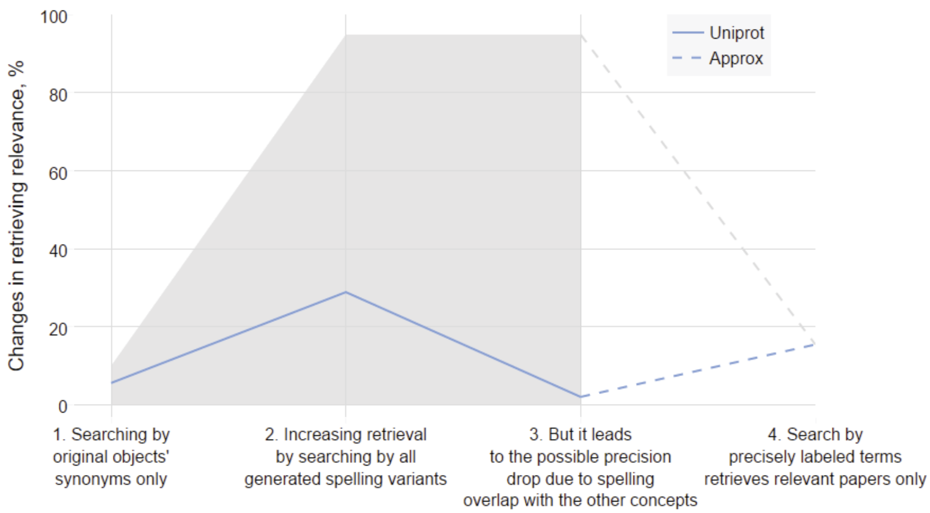
Table 1 shows average number of original terms' synonyms and how much variants were generated. Then we've searched for them in the papers. There is increase of the concept detection of 2.03 - 3 times more when searching for all variants.

Table 2 shows how much objects has terms with identical spellings, i.e. ambiguous terms. Higher overlap within the same ontology than across other ontologies makes algorithmic recognition even more challenging tasks, because algorithms have to distinguish objects within the same class.

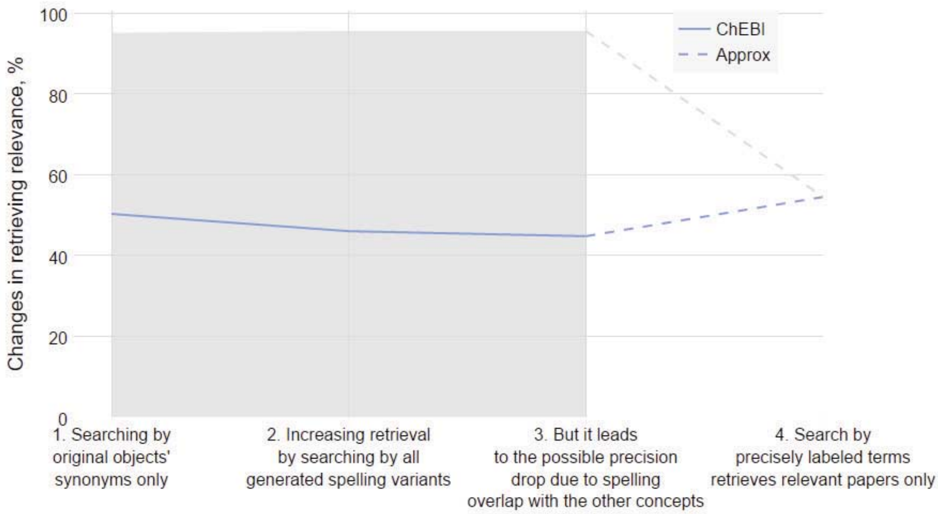
Table 2. Ambiguity in the ontologies and influence on paper's precision

Ontology name	**Ontology size, n of IDs**	**Number of the objects' names with the same spelling within ontology**	**Number of the objects with the same spelling for all variants *within ontology***	**Number of the objects with the same spelling for all variants *across other ontologies***	**Number of the objects with the same spelling for all variants across *the same and other ontologies***	**Maximum number of objects with the same spelling**	**The smallest expected precision**
Uniprot	553667	0	75271	3426	78697	17	0.0588
ChEBI	104854	0	14921	597	15518	1780	0.0006
Gene Ontology	46517	0	2882	1177	4059	9	0.1111
Drugbank	8221	0	20771	478	21249	766	0.0013
ICD-10	11420	0	1882	762	2644	6	0.1667
MeSH (A-F tree)	23716	-	-	-	-	-	-

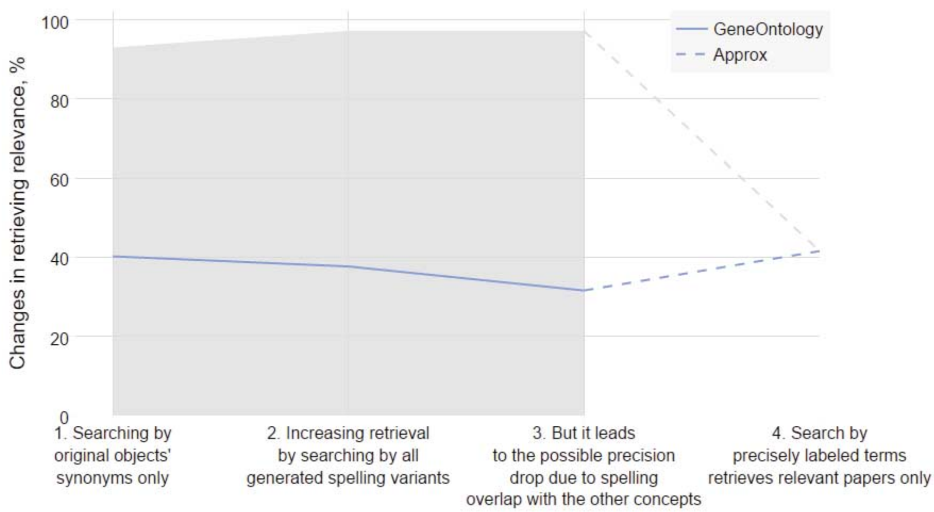
Fig.1 shows overall influence of the variability and ambiguity of the terminology on paper's discoverability.



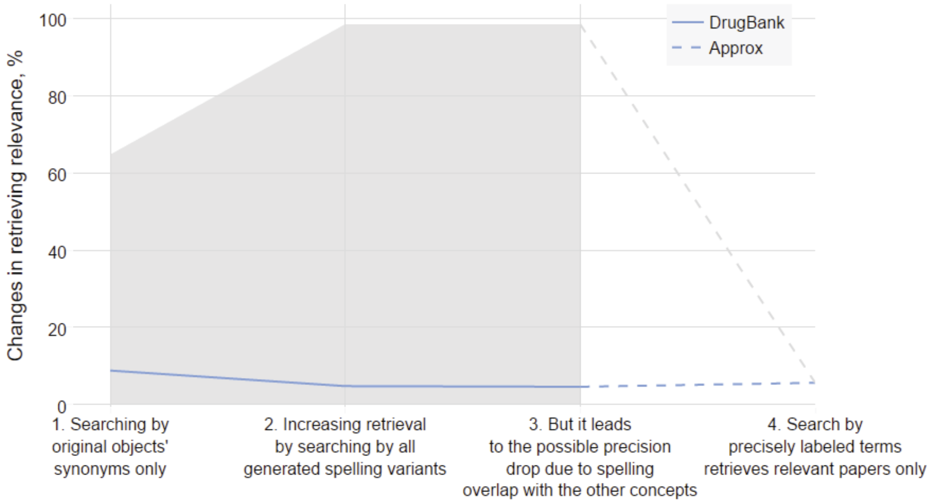
a) Uniprot



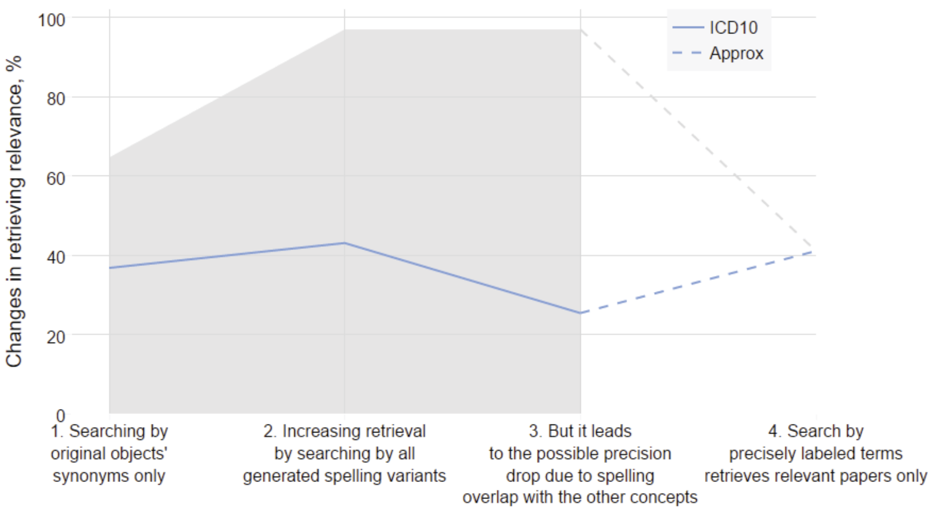
b) ChEBI



c) GeneOntology



d) DrugBank

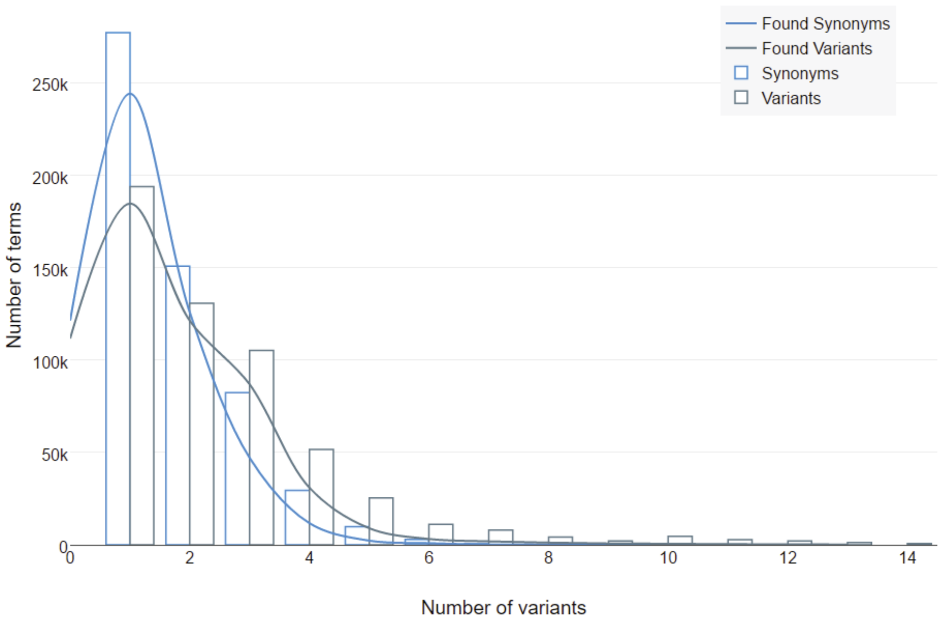


e) ICD10

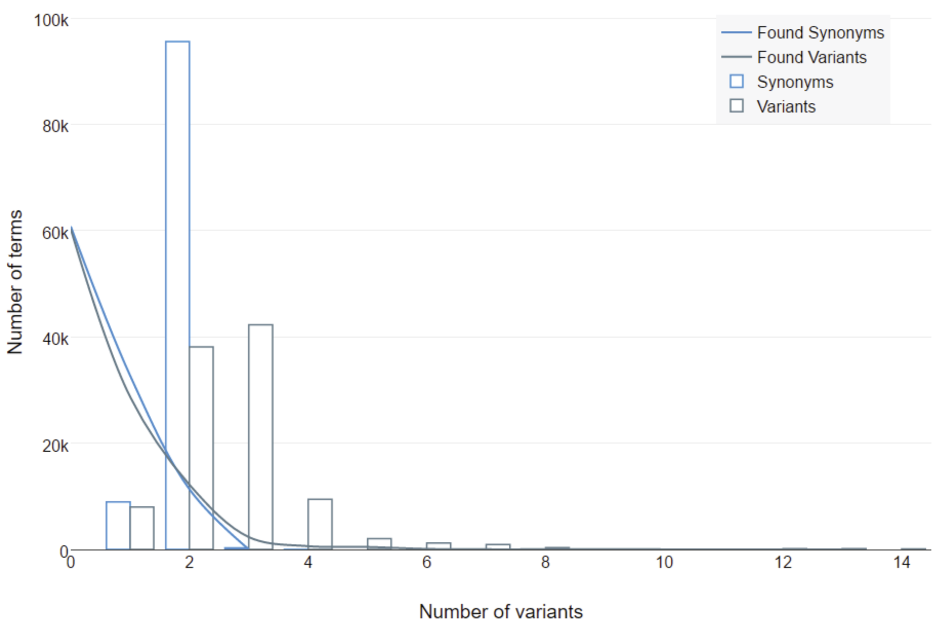
Figure 1. Influence of the terminology variability and ambiguity on a paper retrieval when text contains proteins, chemical elements, genes, drugs, diseases.

When searching by original synonyms only, average likelihood of finding papers is lower than searching by all possible variants. Retrieving higher amount of the papers can be done at the cost of their relevance. Increasing amount of variants leads to the drop of the probabilistic precision. Relevance can be guaranteed only in case of labeling terms and searching by exact ID instead of a string. As long as current literature is not labeled, exact recall and precision can't be calculated for. We used relative changes instead, to visualize scale of the issue across most of the available literature.

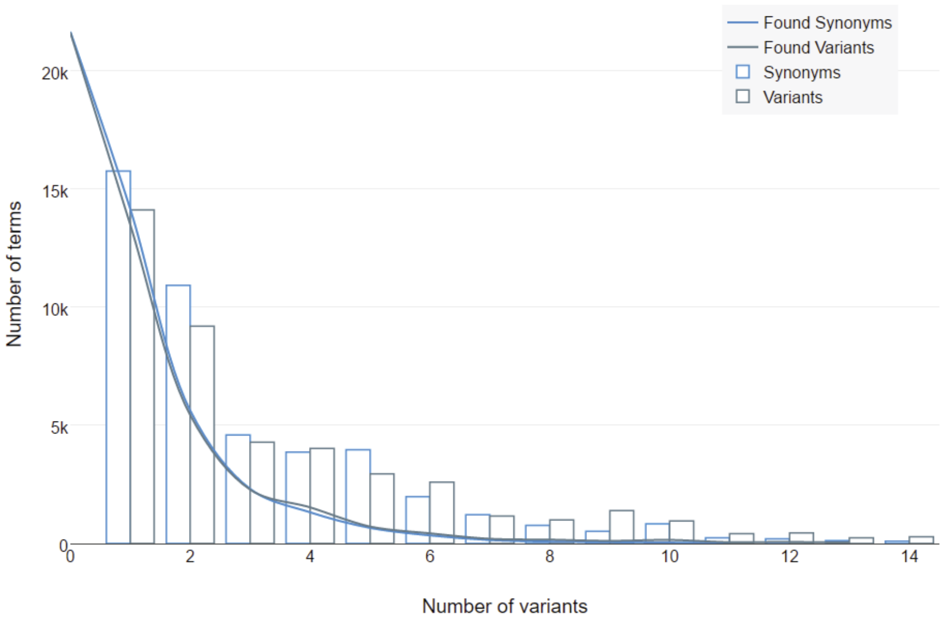
Fig.2 shows the distribution of the variability across ontologies.



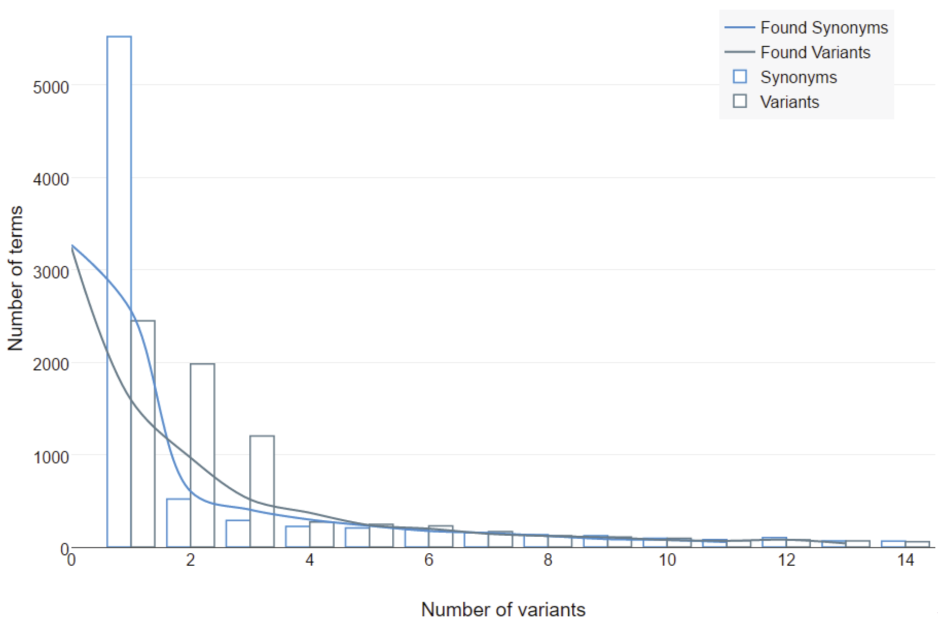
a) Uniprot



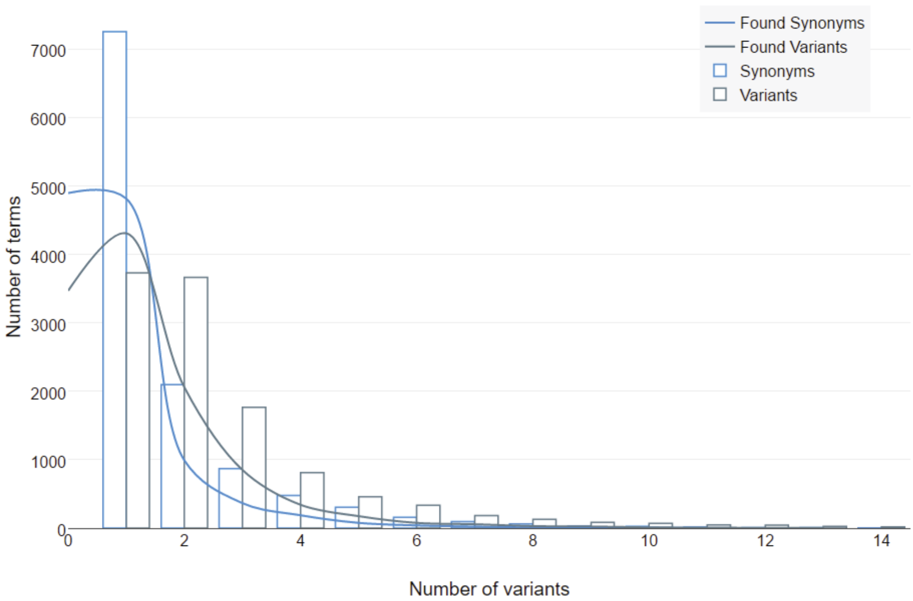
b) ChEBI



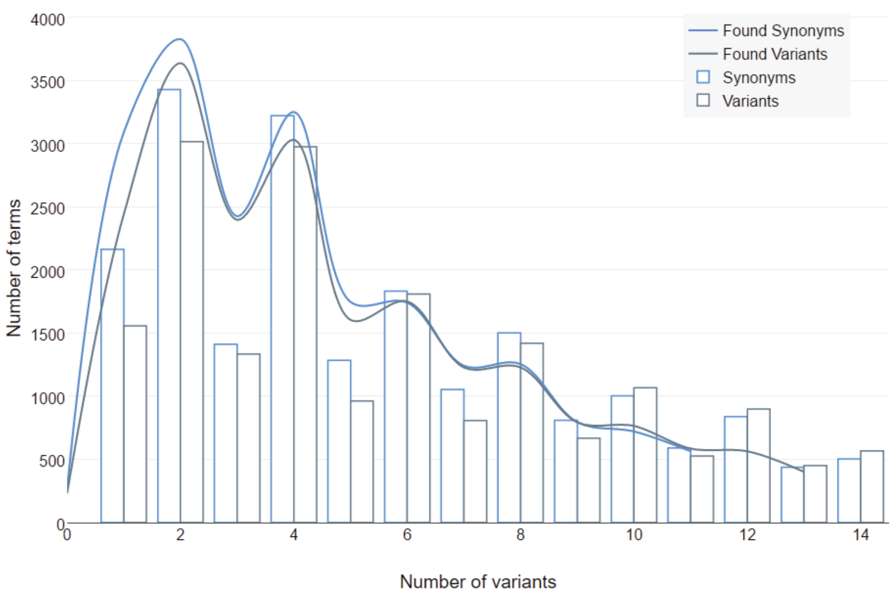
c) GeneOntology



d) DrugBank



e) ICD10



f) MeSH

Figure 2. Number of objects that have specific amounts of original synonyms, generated variants and variants found in the papers.

There are original synonyms in the ontologies, generated variants and how much of them were found in the papers. It shows that there are significant chances to found term's spelling that was never mentioned in the ontology. Thus, it reduces recall of the relevant papers.

We were going to use the MSH WSD Data Set [15] initially for the ambiguity testing purposes, but it turned out to contain generic words only. So, we performed a generic wide search across all ontologies and variants to obtain low-level detalization.

There is also a case of "artificial" ambiguity in ontologies. It is caused by intersection of alternative names and terms' descriptions in attempt of extending variants to increase recall. "Carbon monoxide" example is provided in the Table 3.

Table 3. Example of the objects with the same spelling.

Primary name	Ontology	Matched Objects	Synonyms
Carbon monoxide	ICD10	Accidental poisoning by and exposure to other gases and vapours (X47)	utility gas, utility ga, helium, motor exhaust ga, accidental poisoning by and exposure to other gases and vapours, sulfur dioxide, lacrimogenic gas, **carbon monoxide**, motor exhaust gas, nitrogen oxides
		Intentional self-poisoning by and exposure to other gases and vapours (X67)	utility gas, intentional self-poisoning by and exposure to other gases and vapours, helium, sulfur dioxide, lacrimogenic gas, **carbon monoxide**, motor exhaust gas, nitrogen oxides
		Poisoning by and exposure to other gases and vapours, undetermined intent (Y17)	poisoning by and exposure to other gases and vapours, undetermined intent, utility gas, helium, sulfur dioxide, lacrimogenic gas, **carbon monoxide**, motor exhaust gas, nitrogen oxides
	MeSH	Carbon Monoxide (D002248)	**carbon monoxide**, monoxide, carbon
	ChEBI	carbon monoxide (CHEBI:17245)	**carbon monoxide**, CO
	Drugbank	Carbon monoxide (DB11588)	**carbon monoxide**, lung diffusion test mix nohco, lung diffusion test mixture, lung diffusion test mixture gas, carbon monoxide, compressed air medical G.M., lung diffusion test mix no Ne Co, carbon(II) oxide, Co-HE-O2-N2 mixture, carboneum oxygenisatum, Co-NE-O2-N2 mixture, carbon monox, helium, oxygen, nitrogen L.D.M., CO
CH	GeneOntology	naringenin-chalcone synthase activity (GO:0016210)	**CH**, malonyl-coa:4-coumaroyl-coa malonyltransferase(cyclizing), chalcone synthase activity, chalcone synthetase activity, naringenin-chalcone synthase activity, flavanone synthase activity, DOCS, flavonone synthase activity, DOC, CHS

	Uniprot	Cheilanthifoline synthase (C7195_ESCCA)	**CH**, CHS, cytochrome P450 719A5, cheilanthifoline synthase
		Canavanine hydrolase (CANHY_HELVI)	**CH**, canavanine hydrolase
	ChEBI	methanylylidene group (CHEBI:29432)	**CH**, methanylylidene group
		methylidyne group (CHEBI:29429)	**CH**, methylidyne group
		methanetriyl group (CHEBI:29433)	**CH**, methanetriyl group
	Drugbank	N-Cyclohexyltaurine (DB03309)	CHES, n-cyclohexyltaurine, **CH**
IMP	MeSH = ChEBI	Inosine Monophosphate (D007291)	ribosylhypoxanthine monophosphate, inosinic acid, **IMP**, inosinate, sodium, sodium inosinate, inosine monophosphate, acids, inosinic, monophosphate, ribosylhypoxanthine, inosinic acids, monophosphate, inosine, acid, inosinic
	DrugBank	Imipenem (DB01598)	imipemide, imipenem anhydrou, imipenem anhydrous, n-formimidoylthienamycin, imipenem, **IMP**, imipenem and cilastatin for injection, USP, ran-imipenem-cilastatin, imipenem, n-formimidoyl thienamycin, imipenem and cilastatin, imipenemum, imipenem and cilastatin for injection, -USP, primaxin 250, imipenem and cilastatin for injection USP, (5R,6S)-6-((R)-1-Hydroxyethyl)-3-(2-(iminomethylamino) ethylthio)-7-oxo-1-azabicyclo(3.2.0) hept-2-ene-2-carbonsaeure, primaxin IV 500, primaxin 500, primaxin IV 250/250 add-vantage vial, imipenem and cilastatin for injection, usp, imipenem and cilastatin for injection,-usp, (5R,6S)-3-(2-formimidoylamino-ethylsulfanyl)-6-((R)-1-hydroxy-ethyl)-7-oxo-1-aza-bicyclo[3.2.0] hept-2-ene-2-carboxylic acid, imipenem and cilastatin for injection, tienamycin, imipenem and cilastatin for injection usp, imipenem and cilastatin for injection-USP, imipenem and cilastatin for injection-usp, primaxin IV, primaxin-iv, n-formimidoyl thienamycin, (5R,6S)-3-((2-(formimidoylamino) ethyl) thio)-6-((R)-1-hydroxyethyl)-7-oxo-1-azabicyclo(3.2.0) hept-2-ene-2-carboxylic acid
	GeneOntology. The same	obsolete mitochondrial inner membrane peptidase activity (GO:0004244)	**IMP**, obsolete mitochondrial inner membrane peptidase activity, mitochondrial inner membrane peptidase activity

		mitochondrial inner membrane peptidase complex (GO:0042720)	**IMP**, mitochondrial inner membrane peptidase complex
	ChEBI	IMP (CHEBI:17202)	**IMP**, C10H13N4O8P
	Uniprot. The same for various organisms	Inositol monophosphatase (IMPA1_DICDI)	IMPase, **IMP**, inositol-1(or 4)-monophosphatase, inositol monophosphatase, d-galactose 1-phosphate phosphatase
		Inositol monophosphatase (IMPP_MESCR)	IMPase, **IMP**, inositol-1(or 4)-monophosphatase, inositol monophosphatase
		Inositol monophosphatase ttx-7 (IMPA1_CAEEL)	IMPase, **IMP**, inositol monophosphatase ttx 7, abnormal thermotaxis protein vii, inositol-1(or 4)-monophosphatase, abnormal thermotaxis protein-7, inositol monophosphatase ttx vii, inositol monophosphatase ttx-vii, abnormal thermotaxis protein7, d-galactose 1-phosphate phosphatase, abnormal thermotaxis protein 7, inositol monophosphatase ttx-7, abnormal thermotaxis protein-vii, inositol monophosphatase ttx7

This leads to the necessity of human validation of the same concepts identification. Such functionality exists in sci.AI to validate several ID's from the various ontologies for the same term (Fig. 3).

4. Decreasing retrieving uncertainty with the precise semantic labeling feature of the sci.AI platform

Formalization and statistics above show that uncertainty is not an exception but basic feature of the biomedical text mining. This uncertainty might lead to significant deviations when interpreting academic papers with unsupervised methods only. While it might be acceptable for fiction literature mining, because the major task there is context and sentiments analysis that acts as a smoothing function—such uncertainty might contradict goals of mining STEM research communication, where we are looking for the anomalistic or novel discoveries, exact objects interactions, verification of facts, and relations between statements in various texts. This is why accepting uncertainty might have significant negative consequences on the LBD.

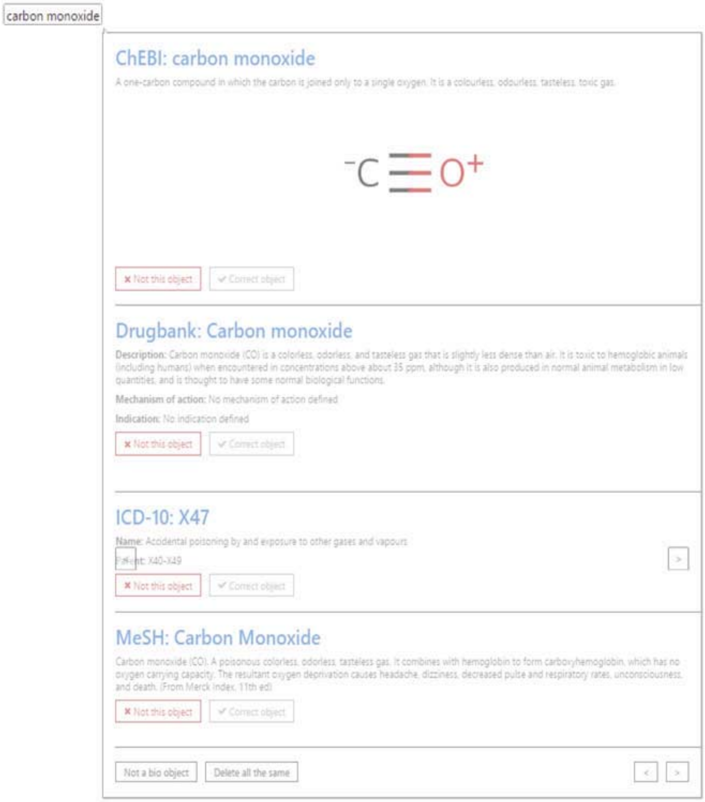


Figure 3. Labeling term with several objects.

In order to address this issue, we implemented the sci.AI platform that has supervised labeling functionality on top of the text mining framework. After initial automatic terms recognition, no matter whether precision is 70% or 99%, users can make final verifications to level up recognition precision to 100%. From the perspective of the search and text mining algorithms, this means removing any uncertainty, which, in turn, leads to exact papers extraction in an SQL-like querying manner. Thus, assuming that author will always label terms correctly, maximum precision and recall will be achieved.

Human-made corrections will be used as training data for the next processings of a text. Such learning with human feedback provides steady path to gradient growth of text mining quality.

Current version of the sci.AI allows to upload text, then performs Named Entities Recognition (NER) task automatically. Author or annotator can validate labeling results via interface and export final structured text to the XML file (Fig.4).

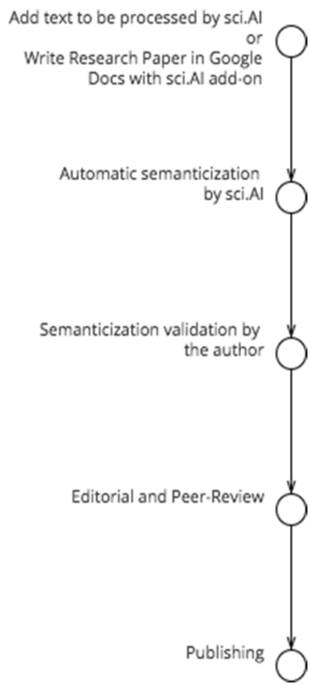


Figure 4. Paper semanticization in sci.AI.

Development roadmap includes release of the next features:

- Machine Learning based analysis to provide the most likeable variants in the first place. This feature will be based on the (a) logs of the terms validation events and (b) statistical co-occurrences of the terms in all available texts. We expect that it will provide approximately 90% recognition rate, as reported by NN researchers [7, 8]. Introduction of the NN prioritisation is expected to reduce cases of the necessary authors intervention to the reasonable minimum. Internal time tracking done by our team members suggest that 10 pages validation time will be reduced from 1 h to 15 min, approximately. Current statistics will be corrected after actual feature release and making more measures for the various annotators and texts with different density of the terms;
- Graph based WSD connections between objects and existing metaontologies data;
- Generating JATS, RDF/XML and RDFa files;
- Validation by the readers, not only by the authors;
- Concepts interactions labeling, for example, protein-protein.

The system can be embedded into the publishing process directly. Both authors and editors can create semanticized versions during submission of even publish new version of the digital paper. Key features of the current production version are as follows:

1. Automated metatagging of biomedical concepts, (named-entity recognition, with further context-dependant semanticization of terms). Current tags contain links to the related objects in the ontology;
2. User-friendly web preprint for tags editing and recognition supervising;
3. Web application for easy integration into the existing publishing process.

sci.AI allows labeling a term with several term-2-ontologyId relationships. For example, '*serotonin*' is the object UID=*D012701* in *MeSH* and UID=*28790* in *ChEBI* ontologies simultaneously. Author or reader can suggest additional UIDs too or can correct existing one. This functionality contributes to the global connectivity between terms and might support UMLS Word Sense Disambiguation (WSD) works.

We expect another possible positive effect for the future papers too. Authors might come to the same single spelling variant of the concept, like "ClNa" only and not "salt" or "NaCl".

As of the end of 2016, sci.AI is in the first phases of its long-term development roadmap from being a semanticization tool to becoming a full-fledged artificial intelligence (AI) tool applied to the life sciences. Wide adoption of this application will extend the publisher's role even further into research results delivery to the intended target audience.

5. Discussion

This paper is the first in our series of researchers about precise semantic labelling of life sciences texts. Our goal was to focus on dependency of the paper's influence on two fundamental factors: ambiguity and variability of terms. In order to avoid excessive complication, we made several assumptions which may bias the results. These simplifications will be addressed in follow-up studies:

1. Prior precision estimation is calculated with assumption that probability of retrieving a paper with concept c_i when searching for ambiguous term t_i might have uniform distribution. In fact, it has a nonlinear distribution, as shown in statistics in Tables 1, 2 and Fig. 2.

2. Prior recall estimation is calculated with assumption that probability of retrieving a paper with concept c_i when searching for term t_i with multiple variants might have uniform distribution. In fact, it has a nonlinear distribution as shown in statistics in Table 1, 2 and Fig. 2.

3. Simplified dependency of recall from ambiguity and precision from variability.

4. Categories of terms variability and implementation of the terms variant generator deserve full comprehensive description in the following research.

6. We intended to show fundamental specifics of biomedical language that makes it is challenging to achieve 100% recognition of terms with unsupervised methods only. Still, there are various NLP approaches including metaontologies like UMLS based disambiguation and statistical methods that significantly improve terms recognition. Those methods are integrated by sci.AI development team and performance of each of them will be evaluated in separate paper.

7. We assumed that there is the same number of concepts and objects within single ontology.

8. "Human factor" was removed from consideration by assuming that author can always correctly label every biomedical concept in own manuscript. Under "precise labeling" we mean "labeling verified by the actual text's author".

9. There are several studies, where researchers propose models of the future paper's success, for example, [19]. Future analysis might take into consideration ambiguity and variability as variables in the prediction models.

10. Part of speech tagging might improve precision of the variants validation. This functionality exists in the sci.AI but was not applied for the statistics calculation.

11. We assume that all possible spelling variants were generated. Further validation is required.

References

[1] J. Preiss & M. Stevenson, "The effect of word sense disambiguation accuracy on literature based discovery". In Proceedings of the ACM Ninth International Workshop on Data and Text Mining in Biomedical Informatics: Medical Informatics and decision making, 2016, pp. 59-63.

[2] "Resource Description Framework (RDF) Model and Syntax Specification", 1999. [Online]. Available: https://www.w3.org/TR/1999/REC-rdf-syntax-19990222. [Accessed: 05-Feb-2017].

[3] Y. Tsuruoka, J. McNaught, and S. Ananiadou, "Normalizing biomedical terms by minimizing ambiguity and variability", BMC Bioinformatics, vol. 9, no. Suppl 3, p. S2, Apr. 2008.

[4] M. Krauthammer and G. Nenadić, "Term Identification in the Biomedical Literature", J. Biomed. Inform., vol. 6, no. 37, pp. 512–26, 2004.

[5] N. P. Cruz Díaz, "An Analysis of Biomedical Tokenization: Problems and Strategies", in Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (Louhi), 2015, pp. 40–49.

[6] A. R. Aronson, "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program", Proceedings. AMIA Symp., pp. 17–21, 2001.

[7] S. Zhang and N. Elhadad, "Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts", J. Biomed. Inform., vol. 46, no. 6, pp. 1088–1098, Dec. 2013.

[8] E. Arısoy, T. N. Sainath, B. Kingsbury, and B. Ramabhadran, "Deep Neural Network Language Models", in NAACL-HLT 2012 Workshop: Will We Ever Really Replace the *N*-Gram Model? On the Future of Language Modeling for HLT, 2012, pp. 20–28.

[9] E. Galeota and M. Pelizzola, "Ontology-based annotations and semantic relations in large-scale (epi)genomics data", Brief. Bioinform., vol. 30, p. bbw036, May 2016.

[10] A. Siu, P. Ernst, and G. Weikum, "Disambiguation of entities in MEDLINE abstracts by combining MeSH terms with knowledge", in Proceedings of the 15th Workshop on Biomedical Natural Language Processing, 2016, pp. 72–76.

[11] "Scholarly HTML", 2014. [Online]. Available: https://w3c.github.io/scholarly-html/. [Accessed: 15-Feb-2017].

[12] "Optimise machine reuse of content tagged in JATS XML", 2016. [Online]. Available: http://jats4r.org/. [Accessed: 15-Feb-2017].

[13] "SciGraph/SciGraph: A Neo4j backed ontology store", 2016. [Online]. Available: https://github.com/SciGraph/SciGraph. [Accessed: 15-Feb-2017].

[14] "Taxonomy - NCBI", 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/taxonomy. [Accessed: 06-Feb-2017].

[15] "Word Sense Disambiguation (WSD) Test Collection", 2010. [Online]. Available: https://wsd.nlm.nih.gov/collaboration.shtml. [Accessed: 07-Feb-2017].

[16] S. N. d. Santos and V. Guimarães "Boosting Named Entity Recognition with Neural Character Embeddings", [Online]. Available: https://arxiv.org/abs/1505.05008. [Accessed: 15-Feb-2017].

[17] G. Lample, M. Ballesteros, S. Subramanian, "Neural Architectures for Named Entity Recognition", in Proceedings of NAACL-HLT 2016, pages 260–270.

[18] C. D. Manning, P. Raghavan and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press. 2008.

[19] I. V. Ponomarev, "Predicting highly cited papers: A Method for Early Detection of Candidate Breakthroughs".

[20] "sci.AI: ELPUB2017 Visualizations", 2017. [Online]. Available: https://sci.ai/elpub2017. [Accessed: 24-Apr-2017].

Appendix A. Examples of the objects' synonyms and generated variants

Variants, that were found in actual papers are marked with * and DOI of one of the retrieved papers.

Primary term: Peroxisome proliferator-activated receptor gamma coactivator 1-alpha

Ontology: Uniprot [PRGC1_HUMAN]
Synonyms: PGC-1-alpha, PPAR-gamma coactivator 1-alpha, PPARGC-1-alpha, Ligand effect modulator 6
Variants: *PGC-1alpha [10.1186/1750-1326-4-10], PPARGC 1-alpha, *PPAR γ-coactivator 1α [10.1038/nutd.2011.3], *peroxisome proliferator-activated receptor-γ-coactivator 1-α [10.1016/j.molmet.2015.09.003], PPARGC-i-α, *PPAR-gamma coactivator 1α [10.1186/1476-511X-10-246], *PPARGC1-α [10.1186/1743-7075-7-88], *PPAR-gamma coactivator 1alpha [10.1155/2008/418765], *peroxisome proliferator-activated receptor γ coactivator 1-alpha [10.1038/srep18011], *peroxisome proliferator-activated receptor-γ-coactivator 1α [10.1210/me.2014-1164], *PPAR γ coactivator 1 α [10.1074/jbc.M115.636878], *peroxisome proliferator-activated receptor γ coactivator 1-α [10.1016/j.molmet.2015.09.003], PGC-i-α, *PGC-1 α [10.1038/ncomms10210], *PPAR-γ-coactivator 1-alpha [10.1371/journal.pone.0055940], *PGC-1α [10.7554/eLife.03245], *ligand effect modulator-6, PPARGC i-α, *PPARGC-1 α [10.1074/jbc.M113.512483], PPARGC ialpha, *PPAR-gamma coactivator 1 α [10.3892/mmr.2013.1714], *PPAR-γ coactivator 1 α [10.1074/jbc.M115.636878], PPAR γ coactivator 1alpha, *ligand effect modulator 6, PGC-i alpha, *peroxisome proliferator-activated receptor-γ-coactivator 1 alpha [10.1038/srep18011], *PPAR-γ coactivator 1α [10.1038/nutd.2011.3], *PPARGC 1alpha [10.1038/nm.2049], PPARGC-ialpha, *peroxisome proliferator-activated receptor γ coactivator 1alpha [10.1111/jnc.12089], PGC iα, ligand effect modulator6, *peroxisome proliferator-activated receptor gamma coactivator 1 alpha [10.1152/ajpgi.00270.2015], *PPARGC 1 α [10.1074/jbc.M113.512483], PGC-i α, *PPAR-γ-coactivator 1 α [10.1074/jbc.M115.636878], *PGC1-α [10.7150/ijbs.7972], *PGC 1α [10.7554/eLife.03245], PPARGC iα, *PPARGC1 alpha [10.1016/j.jnutbio.2009.03.012], *PGC1-alpha [10.2527/jas.2009-1896], *PPARGC1-alpha [10.1016/j.jnutbio.2009.03.012], *PGC1α [10.1016/j.molmet.2015.08.002], *peroxisome proliferator-activated receptor gamma coactivator 1-alpha [10.1152/ajpgi.00270.2015], *PPAR-γ coactivator 1 alpha [10.1371/journal.pone.0055940], ligand effect modulator vi, *peroxisome proliferator-activated receptor γ coactivator 1 alpha [10.1038/srep18011], *PPARGC 1α [10.1371/journal.pgen.1005062], PPARGC-i α, PPARGC i-alpha, *peroxisome proliferator-activated receptor gamma coactivator 1alpha [10.1007/s00125-006-0268-6], *peroxisome proliferator-activated receptor gamma coactivator 1α [10.1038/ncomms3906], *PPARGC-1-α [10.1074/jbc.M113.512483], *PPAR-γ coactivator 1-α [10.1074/jbc.M115.636878], PPARGC-i-alpha, *peroxisome proliferator-activated receptor γ coactivator 1 α [10.1016/j.molmet.2015.09.003], PPAR-γ coactivator 1alpha, *PPAR γ-coactivator 1-α [10.1074/jbc.M115.636878], *PGC 1-alpha [10.1038/sj.ijo.0803567], *PGC-1-α [10.1038/ncomms10210], *PPAR γ-coactivator 1-alpha [10.1371/journal.pone.0055940], PPARGC-iα, *PGC1 alpha [10.2527/jas.2009-1896], *peroxisome proliferator-activated receptor γ-coactivator 1-α [10.1016/j.molmet.2015.09.003], *PPARGC 1-α [10.1074/jbc.M113.512483], *PGC 1 α [10.1038/ncomms10210], PGC-i-alpha, PPARGC i α, *PPARGC-1α [10.1371/journal.pgen.1005062], *peroxisome proliferator-activated receptor γ coactivator 1α [10.1210/me.2014-1164], *PPAR-gamma coactivator 1 alpha [10.1152/japplphysiol.00780.2009], *PPAR-γ-coactivator 1α [10.1038/nutd.2011.3], *PGC 1alpha [10.1186/1750-1326-4-10], *PGC-1 alpha [10.1038/sj.ijo.0803567], *PPAR-γ coactivator 1-alpha [10.1371/journal.pone.0055940], PPARGC-1-alpha, *peroxisome proliferator-activated receptor-γ-coactivator 1alpha [10.1111/jnc.12089], PGC i α, PPARGC i alpha, *PPARGC1 α [10.1186/1743-7075-7-88], *PPAR γ coactivator 1α [10.1038/nutd.2011.3], PGC-iα, PPARGC 1 alpha, *PPAR-γ-coactivator 1-α [10.1074/jbc.M115.636878], *PGC1 α [10.7150/ijbs.7972], *peroxisome proliferator-activated receptor gamma coactivator 1 α [10.3168/jds.2015-9847], *PPARGC1α [10.7554/eLife.18206], *PPAR γ coactivator 1 alpha [10.1371/journal.pone.0055940], *PGC 1 alpha [10.1038/sj.ijo.0803567], PGC i-α, PGC i-alpha, *peroxisome proliferator-activated receptor γ-coactivator 1 α [10.1016/j.molmet.2015.09.003], PPAR-γ-coactivator 1alpha, *peroxisome proliferator-activated receptor-γ-coactivator 1 α [10.1016/j.molmet.2015.09.003], *PPAR γ coactivator 1-alpha [10.1371/journal.pone.0055940], *PPARGC1alpha [10.1186/s13395-016-0083-9], ligand effect modulator-vi, *PGC-1-alpha [10.1038/sj.ijo.0803567], *peroxisome proliferator-activated receptor γ-coactivator 1α [10.1210/me.2014-1164], PGC i alpha, *PPAR-gamma coactivator 1-α [10.3892/mmr.2013.1714], PPAR γ-coactivator 1alpha, PGC ialpha, PPARGC-i alpha, *PPAR γ coactivator 1-α [10.1074/jbc.M115.636878], *PGC 1-α [10.1038/ncomms10210], *peroxisome proliferator-activated receptor gamma coactivator 1-α [10.3168/jds.2015-9847], *PPAR γ-coactivator 1 alpha [10.1371/journal.pone.0055940], PPARGC-1 alpha, *PPAR γ-coactivator 1 α [10.1074/jbc.M115.636878], *PPAR-gamma coactivator 1-alpha [10.1152/japplphysiol.00780.2009], *peroxisome proliferator-activated receptor γ-coactivator 1alpha [10.1111/jnc.12089], *PGC1alpha [10.1677/jme.1.01499], *peroxisome proliferator-activated receptor γ-coactivator 1-alpha [10.1038/srep18011], PGC-ialpha, *PPAR-γ-coactivator 1 alpha [10.1371/journal.pone.0055940], *peroxisome proliferator-activated receptor γ-coactivator 1 alpha [10.1038/srep18011], *PPARGC-1alpha [10.1038/nm.2049], *peroxisome proliferator-activated receptor-γ-coactivator 1-alpha [10.1038/srep18011]

Primary term: Interleukin-1 receptor type 2
Ontology: Uniprot [IL1R2_HUMAN]
Synonyms: IL-1R-2, IL-1RT-2, IL-1RT2,CD121 antigen-like family member B, CDw121b, IL-1 type II receptor, Interleukin-1 receptor beta, Interleukin-1 receptor type II
Variants: interleukin1 receptor type 2, *interleukin-i receptor type-ii [10.1021/ac800928z], *interleukin 1 receptor type-ii [10.1021/ja043466g], *interleukin i-receptor type II [10.1021/ac800928z], interleukin1-receptor type ii, *interleukin-1 receptor type-ii [10.1021/ja043466g], *interleukin-i-receptor type II [10.1021/ac800928z], interleukin-i receptor type-2, interleukin1 receptor type2, interleukin1 receptor type-ii, interleukin-1-receptor type2, *interleukin-1 receptor type ii [10.1021/ja043466g], *interleukin i-receptor type ii [10.1021/ac800928z], interleukin1 receptor type-2, *interleukin i receptor type-ii [10.1021/ac800928z], interleukin1-receptor type II, interleukin-i receptor type 2, *interleukin-i-receptor type-ii [10.1021/ac800928z], interleukin i receptor type 2, *interleukin 1-receptor type ii [10.1021/ja043466g], *interleukin-1-receptor type-ii [10.1021/ja043466g], interleukin1-receptor type-ii, *interleukin 1-receptor type II [10.1021/ja043466g], interleukin1 receptor type ii, interleukin1-receptor type 2, interleukin-i-receptor type2, *interleukin-1 receptor type 2 [10.1038/mi.2015.108], interleukin1 receptor type II, *interleukin i receptor type II [10.1021/ac800928z], interleukin-i-receptor type-2, *interleukin i receptor type ii [10.1021/ac800928z], interleukin1receptor type-ii, interleukin 1-receptor type2, interleukin-i-receptor type 2, *interleukin-1-receptor type II [10.1021/ja043466g], interleukin-1 receptor type2, interleukin1receptor type-2, interleukin 1 receptor type2, interleukin1receptor type 2, *interleukin-1-receptor type ii [10.1021/ja043466g], *interleukin 1 receptor type ii [10.1021/ja043466g], *interleukin 1 receptor type-2 [10.1038/mi.2015.108], interleukin i receptor type-2, *interleukin-i-receptor type ii [10.1021/ac800928z], *interleukin-1-receptor type-2 [10.1038/mi.2015.108], *interleukin-1 receptor type-2 [10.1038/mi.2015.108], interleukin1-receptor type2, interleukin i-receptor type 2, *interleukin 1-receptor type 2 [10.1038/mi.2015.108], interleukin1-receptor type-2, interleukin1receptor type ii, interleukin i receptor type2, interleukin i-receptor type-2, interleukin1receptor type2, *interleukin 1 receptor type II [10.1021/ja043466g], *interleukin 1 receptor type 2 [10.1038/mi.2015.108], *interleukin-1-receptor type 2 [10.1038/mi.2015.108], interleukin1receptor type II, *interleukin 1-receptor type-2 [10.1038/mi.2015.108], *interleukin-i receptor type ii [10.1021/ac800928z], *interleukin-i receptor type II [10.1021/ac800928z], *interleukin i-receptor type-ii [10.1021/ac800928z], *interleukin-1 receptor type II [10.1021/ja043466g], interleukin i-receptor type2, *interleukin 1-receptor type-ii [10.1021/ja043466g], interleukin-i receptor type2, interleukin 1-receptorβ, CD121 antigen-like family member B, interleukin i-receptor-beta, interleukin1 receptor type-II, *interleukin-i receptor type-II [10.1021/ac800928z], interleukin-i-receptor beta, interleukin i receptor-beta, interleukin 1-receptor-β, interleukin i-receptorβ, interleukin1receptor-β, interleukin1receptor type-II, *interleukin-1-receptor-beta [10.1186/1471-2164-11-545], interleukin i receptor β, IL1type II receptor, interleukin-1 receptorbeta, interleukin-1 receptor-β, *IL-1RT2 [10.3389/fnint.2013.00061], interleukin 1 receptor-β, *interleukin 1-receptor-beta [10.1186/1471-2164-11-545], interleukin1-receptor beta, interleukin 1-receptorbeta, *IL 1-type II receptor [10.1084/jem.20020906], *IL-1R2 [10.1038/mi.2015.108], interleukin1receptor-beta, *IL-1R-2, interleukin-1-receptor β, interleukin-i-receptor-beta, *IL-1RT ii, interleukin i receptor beta, *IL-1-type II receptor [10.1084/jem.20020906], *interleukin i receptor type-II [10.1021/ac800928z], interleukin 1-receptor β, *interleukin-1 receptor beta [10.1186/1471-2164-11-545], *interleukin 1-receptor beta [10.1186/1471-2164-11-545], interleukin-1-receptorβ, IL i-type II receptor, interleukin 1 receptorβ, *interleukin-i-receptor type-II [10.1021/ac800928z], interleukin i-receptor beta, interleukin1-receptor β, interleukin1 receptor β, *IL1-type II receptor [10.3389/fcell.2016.00072], interleukin-i-receptorβ, interleukin-i receptor-β, *IL1 type II receptor [10.3389/fcell.2016.00072], *IL-1RT-ii, interleukin i-receptor β, interleukin1-receptor-beta, interleukin-i-receptor-β, interleukin-1 receptor β, interleukin-1 receptorβ, interleukin1-receptor type-II, interleukin-1-receptor-β, *CDw121b, interleukin 1 receptorbeta, interleukin1 receptorbeta, interleukin-i receptor beta, interleukin1 receptorβ, interleukin-1-receptorbeta, interleukin1-receptor-β, interleukin i receptorbeta, CD121 antigen-like family member-b, *IL-1R-ii [10.1177/039463200601900204], *IL-1R 2, interleukin1 receptor-β, interleukin i receptor-β, IL-1RT-2, *interleukin i-receptor type-II [10.1021/ac800928z], IL-i type II receptor, interleukin1-receptorβ, interleukin-i receptorbeta, interleukin-i-receptorbeta, *interleukin 1 receptor beta [10.1186/1471-2164-11-545], interleukin i-receptorbeta, interleukin 1 receptor β, interleukin1 receptor beta, *IL 1 type II receptor [10.1084/jem.20020906], *interleukin 1 receptor type-II [10.1021/ja043466g], *interleukin-1 receptor type-II [10.1021/ja043466g], IL-i-type II receptor, *IL-1R ii [10.1177/039463200601900204], interleukin i receptorβ, interleukin1receptorβ, *interleukin-1-receptor type-II [10.1021/ja043466g], *IL-1 type II receptor [10.1084/jem.20020906], interleukin i-receptor-β, IL-1RT 2, interleukin-i-receptor β, *interleukin 1 receptor-beta [10.1186/1471-2164-11-545], interleukin1 receptor-beta, interleukin-i receptor β, interleukin1receptor β, IL i type II receptor, interleukin1receptorbeta, *interleukin 1-receptor type-II [10.1021/ja043466g], *interleukin-1 receptor-beta [10.1186/1471-2164-11-545], interleukin1receptor beta, interleukin-i receptor-beta, interleukin-i receptorβ, interleukin1-receptorbeta, *interleukin-1-receptor beta [10.1186/1471-2164-11-545]

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-203

Living Digital Ecosystems for Data Preservation

An Austrian Use Case Towards the European Open Science Cloud

Raman Ganguly[a,1], Paolo Budroni[b] and Barbara Sánchez Solís[b]
[a] *University of Vienna, University Computer Center*
[b] *University of Vienna, University Library and Archive Services*

Abstract. This paper will address issues concerning the handling of complex data such as research data, multimedia content, e-learning content, and the use of repositories infrastructures. At the University of Vienna, an ecosystem for digital data preservation and research data management has already been established and will be subsequently be enlarged according to future needs and requirements. in the future. This living digital ecosystem is the foundation for research data management and was implemented from the beginning as a central service according to the FAIR principles as stated in the first HLEG-EOSC [1] report. With the help of ten years of professional experience, a model for digital data preservation was established to address the complexity of heterogeneous data. This was necessary because of different use cases assigned to the interdisciplinary data management team based at the Computer Centre and the Library. The source for the use cases are research projects, their different approach to research and their multifaceted requirements regarding the efficient re-use of data. The usage of this model might be considered as the foundation on which an ecosystem for digital data preservation can be built.

Keywords. visualization of data, repositories infrastructure, digital workflow, research data management, data life cycle

1. Introduction

A solid research data management system is the foundation of open science, open data and open access. Ten years ago, the University of Vienna inaugurated a project with the goal of creating a system which could house digital objects. With the idea of a simple repository to manage data, the project Phaidra (Permanent Hosting, Archiving and Indexing of Digital Resources and Assets) was born. From the beginning, openness was a key motivation and we invited every member of the University, including students, to use the repository. We also provided our technology to other universities and institutions, and so the Phaidra network was created. Today Phaidra is used at research institutions in five different countries.

As more users began to work with the repository, it became apparent that the system should be more flexible and more "agnostic". For these reasons the management started a reengineering process and to rethink the whole setup. Back to the design phase, the management communicated with stakeholders and were confronted

[1] Corresponding Author, Raman Ganguly, University of Vienna, Universitaetsstraße 7 (NIG), 1010 Vienna, Austria, tel.: +43 1 4277 14189, e-mail: raman.ganguly@univie.ac.at,

with a broad range of research data and use cases. The goal became clear: to address as many stakeholder needs as possible. To meet this goal, it was first decided to refactor the technical structure of the repositories to a micro-services architecture, and second, models for data management were designed, which could be used for different use cases.

Furthermore, the management decided to start a nationwide project in 2014, including as many Austrian research institutions as possible. e-Infrastructures Austria[2] was a federally funded program for the coordinated expansion and continued development of data repositories across Austria, and was made possible by a grant from the Austrian Ministry of Science, Research and Commerce (BMWFW). The program enabled the safe archival and lasting availability of electronic publications, multimedia objects and other digital data from the research and teaching fields. Concurrently, topics relating to research data management and digital archiving workflows were being addressed. This project offered the ideal frame, to discuss and evaluate the present data preservation strategies with Austrian and international experts.

2. Models for data management

Using three different models as a guide, the management redesigned the repository infrastructure, an important starting point for the transition from a simple repository concept to a living digital ecosystem concept. Based on the suggestions of stakeholders, they took a close look at the research process regarding data. The data lifecycle became the focus of the first model.

The second model describes a workflow for the ingestion of entering data into an archiving system and making it available for re-use. When implementing data management from the start, future re-use is already included as the next step in the data lifecycle.

The third model was driven by the idea that no one system fits for all types of data. It suggests how data could be evaluated to determine which archiving system is ideal for storage.

2.1. Data lifecycle model

When publishing data, the data volume is usually small and appropriate archiving formats already exist. However, this is only the top of the iceberg – which as becomes evident when looking at data in the research process. The value of publications rests in their proper preservation, as stated in the PARSE Insight report: „Digital preservation of research data here means the careful storage of all research output in such a way that it remains accessible, usable and understandable over the long term." [2]

To get a closer look at this iceberg, the management Phaidra Management created a model based on the Data Publication Pyramid [3], and added data not directly included in publications, such as inconclusive and negative results. This worked from the point of view of the data and not the publications themselves.

To get a closer look at this iceberg, the management Phaidra Management created a model based on the Data Publication Pyramid [4], and added data not directly

[2] e-infrastructures Austria: https://e-infrastructures.at/en

included in publications, such as inconclusive and negative results. The illustration [figure 1] worked from the point of view of the data and not the publications themselves.

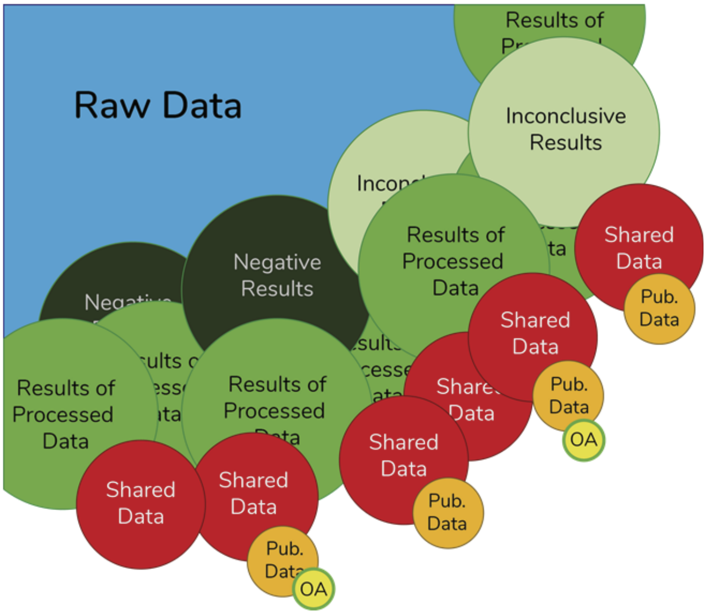


Figure 1. Data life cycle model.

2.2. Digital workflow model

The workflow model is the central model for the ideal of the digital ecosystem and is based on the OAIS environment model from the CCSDS [5]. In this simple model, archives are in the centre, surrounded by the producer, consumer and management. The digital workflow model describes the environment for data management more specifically than the OAIS and defines the points at which data will be transferred from one party to another. The involved parties are the data producer, the archiving manager and the data re-user.

According to the model and to the terms of use of the digital archive, which covers all phases of the life cycle of the data, the data producer is the party who creates and owns the data. It is the data producer's role to define in which quality, how long and in which way and in which context, the data and the related metadata can be re-used. Much clarification is necessary and a data management plan is a useful key instrument for the data producer in answering these three key questions. Data management plans are like a project plan for data and like any other living document should be kept up to date throughout the entire project. They are also a useful tool for data management and data inventory in preservation planning. All that is required is machine-readable output from the data management tool.

Information from the data producer is essential for data management. Data management maintains data quality over a specified time and ensures that only authorized users can access the data. During the ingest process, the data and the responsibility for it are transferred from the producer to the manager. The next transfer of responsibility occurs when the data are delivered to the re-user. For the data re-user, the allowed methods of re-use must be clear, so license agreements must be provided to the re-user and accompany the data.

The FAIR principles [6] (Findable, Accessible, Interoperable, Re-usable) principle should act as a guiding principle for the data re-user. This principle should be adhered to starting, at the latest, at the ingest phase. In this phase, data conversion and enrichment occur. In Phaidra this is possible.

The illustration [figure 2] symbolizes a common legal space for the data. It should be a space where there are common terms of use and data can move without legal barriers from one system to another. Clarification of ownership and license agreements at the ingest process help to create a kind of "Schengen Area" for the populations of data being preserved and managed in this area. Policies, governance, rules of engagement and terms of use for services and data management policies on an institutional level complete the clarification of data usage.

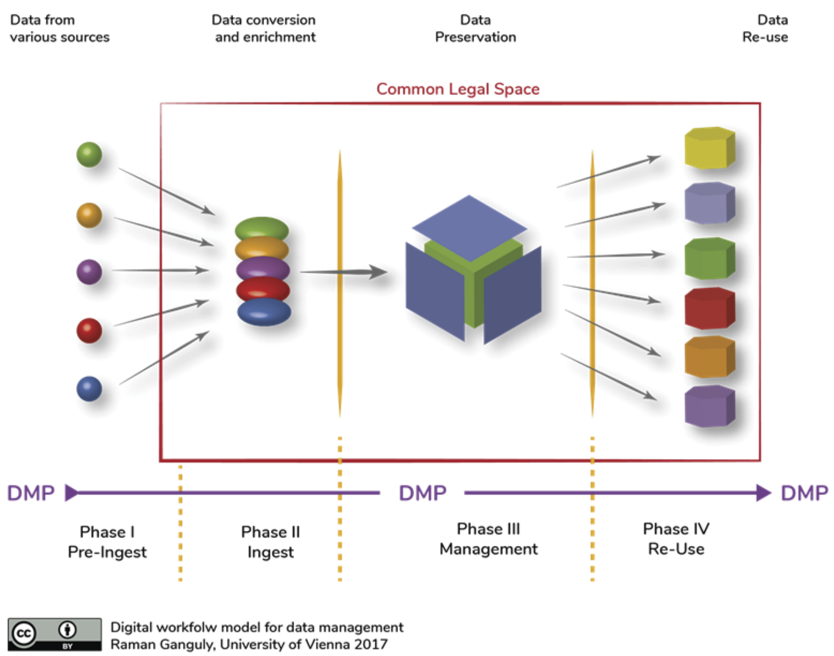


Figure 2. Digital workflow model.

2.3. Evaluation of data

The third model takes a closer look at data management itself and the decision of where to store data. Due to the heterogeneity of research data, one repository or archiving system cannot fit every for all kinds of data. A data manager must decide where to store data in order to *a) maintain quality and b) make data available for re-use*. The attributes of model three can be used to evaluate the data and the archiving system: amount of data, duration of archiving, and complexity of the data format. The attributes of the data should be written in the data management plan, which can then be compared with the features of the archiving system.

The amount of data is easy to measure by counting the files and the file size. Of relevance here is to determine whether there are many small files or only a few large files. This is a major factor when choosing an appropriate storage system. Archiving data is costly and not all data must necessarily be preserved for the long term. For some data, preservation for three to ten years may suffice (e.g for some kind of educational resources), but this should also be carefully planned and executed. The complexity of the data format should be examined from the perspective of data preservation and re-use. Audio and video files are more complex than document files. Databases and software (plus the related contextual and provenance metadata) have special needs in the re-use phase. As the illustrations [figure 3] shows both, the facts of data and the repositories can be added to a grid and compared.

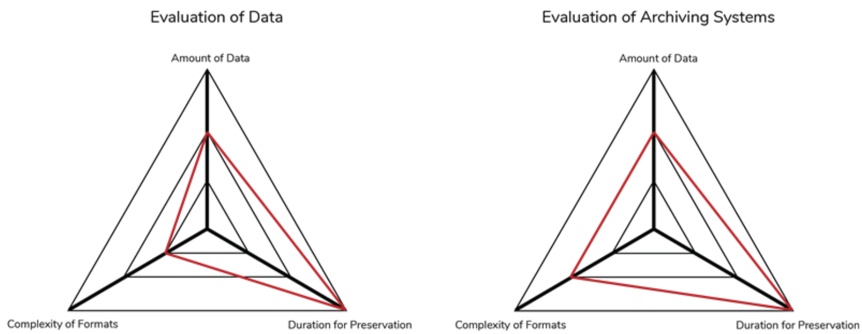


Figure 3. Evaluation of data and repositories.

3. From the repository to the digital ecosystem

The European Science Agenda [6] identifies three layers for data-driven science: data, services and governance. E-Infrastructures cover them all, and is the foundation for data preservation, since data are managed and curated at the infrastructure level. With infrastructure as a foundation, and taking into consideration the three layers proposed by the EU-Commission, services for ingest and re-use are built. This brings value to the infrastructure. Services should be easy to use and appropriate for the use cases of the data producers. Finally, governance is the framework which through appropriate and published policies provides an institutional format for data preservation.

The illustration [figure 4] provides an overview of the strategies used to build a digital ecosystem. Further discussion regarding the infrastructure and service layers will be provided in the following chapters. The governance layer is relevant for designing ecosystems, but is not the main driver and will therefore not be discussed in this paper.

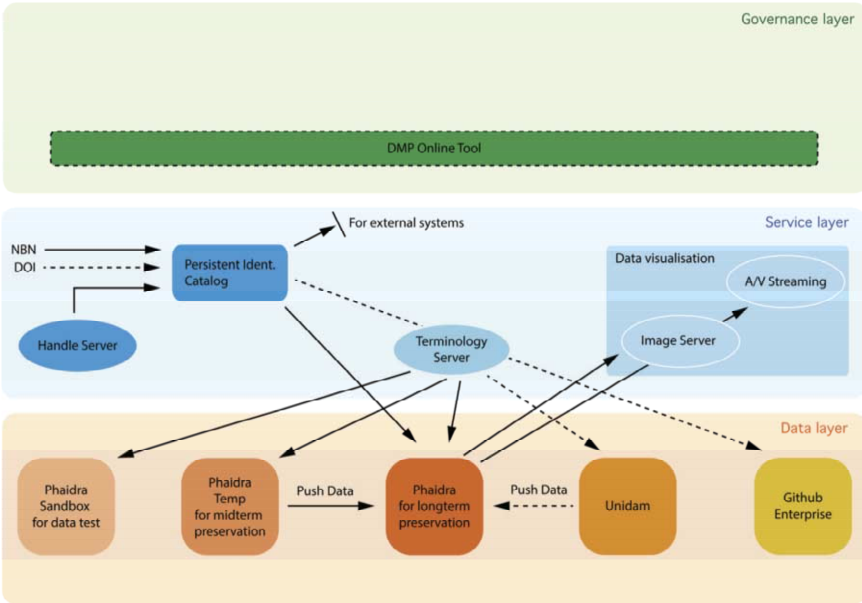


Figure 4. Ecosystem of Phaidra

3.1. The data layer

At the University of Vienna, the management started to create infrastructure for special services and special archiving systems. The team started with Phaidra, the long-term archiving system for generic data, where different bulks of heterogeneous data can be stored. In this repository, all metadata and controlled vocabularies were administrated. Per definition, the long-term archive Phaidra provides a persistent identifier for data which cannot be deleted.

According to certain requirements, some kind of data may be deleted after a defined period of time. Therefore, and in addition, a second repository for midterm archiving was established, where data can be deleted and, in the future, seamlessly transferred to the long-term archive. Currently, an automatic deletion of data after a specified time is not possible. Such a feature requires a better data management plan tool in a machine-readable format a related policy.

As a service, we also provide a repository for testing data, so that users can perform quality testing on data. This repository is the so called Sandbox, and it is mainly a clone of the long-term archiving system. In total, we operate three repositories for generic data with different purposes.

A further repository system, called Unidam, which was first created by two faculties of our university has recently been fully integrated to the central data management infrastructure. This gives repository users the possibility to get more features for their data, particularly in the field of digital humanities.

Based on the nationwide survey “Researchers and Their Data. Results of an Austrian Survey” (2015), which was directed at practically all Austrian researchers (36000 persons), we identified that nearly 25% of research projects use software developed during the process [7]. Looking at the software developed, and using what we know about well-established repositories for this purpose, it was possible to implement a Github Enterprise repository for such research and to integrate it into our ecosystem. This enables data to be linked to a software release, which could also be identified by a persistent identifier.

3.2. The services layer

In the services layer, re-use is the greatest value. For this layer, we reengineered the architecture of our Phaidra repository and integrated an API to enable other applications to dock on Phaidra. This change helped us to integrate an image server for presenting large images over the web and a streaming service for audio and video material, which is stored at the repository.

A further part of the service layer provides tools for managing data. We implemented a terminology server for controlled vocabularies, based on the SKOS [8] standard. This gives our users the possibility to choose controlled vocabularies on a wider range. A handle server creates persistent identifiers throughout the entire digital ecosystem, allowing consistent object referencing.

4. Outlooks

In the future, we plan to integrate a service for data management plans based on the DMP Online Tool [9] from DCC (Digital Curation Centre based at the University of Edinburgh) and the recommendation from RDA (Research Data Alliance) [10] regarding actionable data management plans. These are data management plans which are provided in both a human-readable and machine-readable way. Machine-readable output can further be used in tools for data stewardship. This allows more control over the data, its provenance and context, all relevant for re-use.

Currently, software development takes place in the research community, which poses a challenge regarding infrastructures and coordination. The question is, if software developed by research projects constitute a part of data preservation, and if so, how can software be maintained after a project ends? This challenge shows the need for technical consulting for researchers from the beginning of a project.

Important steps for the digital ecosystem are not only to provide a good working infrastructure, but to connect with the research community and maintain links to other infrastructure projects. Therefore, it is essential to our services to maintain the yet existing links to projects such as OpenAIRE[3] and Europeana[4] and OAPEN[5]. We are in regular contact with GÉANT[6] [link 5], and observe the European Open Science Cloud[7] [link 6] and large Austrian infrastructure projects, such as the Vienna Scientific Cluster.

---

[3] OpenAIRE: https://www.openaire.eu/

[4] Europeana: http://www.europeana.eu/portal/en

[5] OAPEN: http://www.oapen.org

[6] GÉANT: http://www.geant.org/

[7] European Open Science Cloud: https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

Acknowledgements

Web references of named organizations

[link 1] e-infrastructures Austria: https://e-infrastructures.at/en
[link 2] OpenAIRE: https://www.openaire.eu/
[link 3] Europeana: http://www.europeana.eu/portal/en
[link 4] OAPEN: http://www.oapen.org
[link 5] GÉANT: http://www.geant.org/
[link 6] European Open Science Cloud: https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

References

[1] Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud. Available from: https://ec.europa.eu/digital-single-market/en/news/first-report-high-level-expert-group-european-open-science-cloud : 12
[2] PARSE.Insight report from the FP7-2007-223758. At LIBER [homepage on the Internet. cited 2017 March 18]. Available from: http://libereurope.eu/wp-content/uploads/PARSE-Insight_D3-5_InterimInsightReport_final.pdf : 3
[3] Susan Reilly, Wouter Schallier, Sabine Schrimpf, Eefke Smit, Max Wilkinson. Report on Integration of Data and Publications. Opportunities for Data Exchange (ODE). FP7 Grant Agreement number 261530 : 19
[4] The Consultative Committee for Space Data System Practices (CCSDS). Reference Model for an Open Archival Information System (OAIS). Magenta Book, June 2012, CCSDS 650.0-M-2 Available under: https://public.ccsds.org/pubs/650x0m2.pdf : 2-2
[5] Mark D. Wilkinson, Michel Dumontier, Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. nature [homepage on the Internet. cited 2017 March 18]. Available from: http://www.nature.com/articles/sdata201618
[6] J. C. Burgelman. European Open Science Agenda. Presentation. Brussels, 15 January 2016. Available from: https://www.era-learn.eu/events/annual-joint-programming-2015-new-date-2016/topic-3-strategies-for-fostering-open-knowledge-and-open-access-in-research/01_2016OpenScienceAgendaERALEARNconference.pdf : 13
[7] e-infrstructures Austria. Researchers and Their Data. Results of an Austrian Survey. Report 2015. Available from: :24
[8] SKOS Simple Knowledge Organization System. W3C [homepage on the Internet. Cited 2017 March 19]. Available form: https://www.w3.org/2004/02/skos/
[9] DMPonline. [homepage on the Internet. Cited 2017 March 19]. Available form: http://www.dcc.ac.uk/dmponline
[10] Tomasz Miksa. Information integration through Actionable Data Management Plans. RDA 8th Plenary https://www.rd-alliance.org/system/files/documents/RDA_8thPlenary_Miksa.pdf: 2

List of Figures

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-211

Increasing the Discovery and Use of Non-Patent Literature (NPL): Scientific Publications in Patent Examination

Fernando LOIZIDES[a,1] , Barrou DIALLO[b], Andrew POLLARD[a], Aekaterini MAVRI[c]
[a] *University of Wolverhampton*
[b] *Independent Author*
[c] *Cyprus University of Technology*

Abstract. In this work we present findings on non-patent literature use, and specifically scientific publications such as academic articles. We interview patent examiners and observe their prior art searching in order to provide insights into the perceived usage of non-patent literature and produce high level requirements for advancing non-patent literature search tools.

Keywords. Patents, Patent Examination, Non-Patent Literature, NPL, Scientific Literature, Academic Publications, Intellectual Property

1. Introduction and Motivation

Patents give the sole rights to an invention to an individual or company on a unique novel item. Patent applications have increased dramatically yearly in the past few decades [1, 2, 3, 17]. Every year, billions are spent in lawsuits over patent infringement and legal battles, making these publications an expensive commodity [4]. In order for a patent to be accepted and published, a patent application is filed and an examination occurs in which prior art is sought, in order to dispute the novelty of the application. A patent examiner triages [5] through large amounts of publications in order to determine the uniqueness of the invention and make a decision as to whether to publish, require amendments and resubmit, or to reject the patent. Being able to search through the previous publications in this scenario is vital so as not to grant a patent to an already existing invention. Therefore, the search on existing publications needs to be exhaustive.

The main search tools and repository used by patent examiners to search for prior art is currently local databases available to the patent examiners. These can be heavily focused on patent applications and previously awarded patents. There are also however further resources that also classify as prior art. Examples of other types of documents include academic articles and also internet based literature. The last two categories are also deemed as extremely relevant and useful to patent examination [15]. Indeed “the knowledge generated by academic scientists has been deemed one of the most crucial ingredients for technological progress and economic growth” [14] and therefore is vital

[1] Corresponding Author’s e-mail address: Fernando.loizides@wlv.ac.uk

to be included in the searching for prior art. It is therefore "irrelevant if prior art is disclosed in patent literature or alternatively in non-patent literature (NPL), including scientific/technical journals and all other kind of content available e.g. on the internet and elsewhere. For in depth analysis of prior art comparable search efforts should therefore be undertaken for non-patent literature as for patent literature" [11]. In fact, many times instead of merely complimenting the search, other types of published work is sufficient to disprove the novelty or prove the existence of already existing prior art on its own [12].

In this work we present and discuss requirements that exist among patent examiners regarding the search and use of existing scientific publications (prior art) and underlying reasons for these. We define scientific publications in this work to mean peer reviewed academic papers such as those found in conference proceedings or in journals. We bring together work from different existing studies within literature and through ongoing user studies conducted within the European Patent Office in The Hague, Netherlands with interviews from patent. The work presented in this paper provides the foundation for subsequent research and development and examines the factors which contribute to the perceived value of these publications such as their findability and accessibility (for example the lack of open access). The value of our work can be seen in three main aims: a) raising awareness of the importance of these publications within patent literature b) eliciting requirements reasons for the discovery and utilization of scientific literature within patent examinations and c) inform technological solutions which may be created or improved upon in future.

2. Background

There is evidence that the use of academic articles within patent examinations and prior art searching is both useful and being sought out within the patent examination domain. For example even in the early 1990's access to use internet resources were made available to the patent examiners, even to a limited degree [6, 7]. In a 2001 internal report 1/30 search reports by the European Patent Office had reference to an internet document. 1/5 search reports within the European Patent Office cite Non-Patent Literature [19]. Guidelines for citing documents from the Internet and their importance were proposed [8]. The European Patent Office has therefore encouraged the use of and included scientific articles and resources within internal publication databases [9]. The process of using cross-organizational resources and digital libraries in order to "overcome the current information deficit and to fulfill the information need of the experts in the innovation-process" was also accelerated and investigated [10]. The prior work and evidence point to a need for a more thorough and up to date investigation and towards a more formal user-requirements based testing and analysis, before setting up any of the mentioned infrastructures for accessing publications. Authors, readers and publishers have now started to move towards an open access paradigm and encourage the availability of academic articles to the general public. We recognize user needs that need to contribute to the design of a framework and tools for patent application analysis with regards to prior art searching, an area that is receiving exponential need and growth [16].

3. Methodology and Findings

We identify initial high level points to assist both in creating suitable repositories as well as tools for searching and presenting information. The results presented in this section originate from previous published work as well as interview sessions from patent examiners at the European Patent Office. Specifically, 18 examiners were interviewed in depth while a further 14 examiners were observed while undertaking a prior art search followed by semi-structured interview. The examiners varied in experience, from 2 to 31 years (AVG 16 years - STDEV 8). The examiners fell under categories A-G (http://www.wipo.int/classifications/ipc/en/ - accessed April 2017). The average amount of prior art searches conducted per year were 70 with a standard deviation of 30. All patent examiners reported this number increasing yearly, indicating that there is a need for further support for faster prior art searches in the same amount of time. Most examiners had extensive experience with scientific publications, usually from their university studies (such as PhD work). All examiners were familiar with scientific publications.

We began by questioning the examiners of their *perceived importance* of non-patent literature, and specifically scientific literature, to the prior art search process. All but one of the examiners saw importance and relevance to searching through scientific documents. The main value was seen in applied research. The examiners all agreed that some fields relate less to academically published material than others. Examples brought forward of research which is highly relevant to scientific publishing regarding their likely relevance in prior art include Informatics areas such as Bioinformatics, Medical informatics and Machine Learning. Examples where scientific literature is perceived to have less of an impact in prior art search include areas such as the furniture and shoe industry, which are more design oriented. Upon questioning the examiners, if they think examinations include NPL responses varied. NPL searching was reported as taking place in specific fields (consistent with the feedback on whether NPL is considered relevant) and on occasion to a limited degree. The question was put forward to the patent examiner which does not use NPL of whether they "think that if they searched for academic literature the outcomes of the examinations may be different" the examiner answered that "to a small extend maybe, some prior art might be cited even if it is less relevant". All but one examiner reported performing NPL search during their prior art searching under 90% of the time. The one remaining examiner reported searching for NPL on every prior art search conducted. In terms of perceived success half the examiners reported 1/3 or less of the times being able to successfully find relevant scientific publications in their prior art searching. 2 of our pool of examiners did not answer. The remaining examiners reported that "most searches can have relevant literature".

3.1. Searching non-patent full text.

Some platforms allow for full text searching. This is however; still weak compared to functionalities in patent full text databases [15]. This was identified by the examiners to be "cumbersome to get the full article sometimes (need to order the article), even if it became much easier in the last years". The search from a search engine is also often done on the abstract or keywords rather than the full text with the user needing to then manually triage through the full text in order to locate relevant sections. There is a needs to allow a full text search directly from the search engine as well as improve tools for the information seeker to investigate the full text internally efficiently.

3.2. Central searching location vs multiple repositories.

One of the current issues that examiners and information seekers face is "to combine several publication search platforms that must be searched separately, with different search interfaces [15]. Search tools should combine different sources and repositories using a familiar common searching interface (see also [11]). Currently, half the examiners reported using mainly external tools in order to search for NPL rather than internal tools. The main reported sources by the pool of participants included Google Scholar, generic Google searching, YouTube and Orchid.

3.3. Different document types.

There are currently different ways to access articles (PDF, HTML, Word, Plain text) which makes the discovery and use of the articles challenging. A common format upon submission by the authors or tools that convert to a standard such as XML should be made available. This would allow for the easier searching of within document material by search engines as well as creating tools for investigating document content by the information seeker. Where there is no automatic XML type conversion within the publisher, an automated medium tool can allow for on-the-fly XML conversion of the documents. It is also worth noting that there is often information associated with the scientific publications which although relevant and useful, may not always be readily available. As one examiner noted: "We miss often the possibility to search conference materials (powerpoint presentation, handbook of abstracts) that are brought up to our attention later by third party or during oppositions".

3.4. Provide familiar and advanced search options.

This principle is one that holds true as a generic usability principle regarding user interfaces. Searching should provide options to allow the user the freedom to customize the search and be more advanced in how they narrow down what they need. Examples include Boolean logic, keyword highlighting tools based on search and image extraction. Examiners reported on using external (to the patent office) tools in order to search for NPL such as scientific documents. As one of the participant reported, "different tools" require "different query languages". Using these tools produced some difficulties of familiarity. For example, some tools are reported to not "prioritize the documents as you [an examiner] would". External tools are also characterized as a "black box" whereas an examiner would like to know how the search is performed. There are often also problems with NPL searching tools such as the opportunity to structure the search in a logical manner rather than searching and losing the search result flow after another search query is entered. The most common search limitation that was reported for scientific publications, is that of "lack of standardized classification of documents". Unlike patent literature that is classified to a detailed level, scientific publication classification varies. Different conferences and journals have their own classification systems which differ to each other, sometimes to a large extent. The other form of classification on scientific documents is a 'keywords' classification, which is often given by the authors and therefore may be even more diverse than the publisher's. It is therefore almost impossible to have a universal classification system implemented on scientific publications. This makes the ability to be exhaustive and the discoverability of the documents extremely difficult for the examiners. This limitation coupled with the fact that NPL search is "more time consuming compared to patent searching", can decrease "efficiency due to lack of time". This results in a 'satisficing' scenario [18] where examiners "draw the line somewhere in terms of time efforts" when a 'reasonable rather than exhaustive' search is made.

3.5. Specialized presentation and navigation

Examiners made a clear point on how "interfaces are not standardized". Tools for specialized searching should include an interface customized to assist in the rapid searching and assimilation of information. Currently, the databases and interfaces in existence for scholarly searches can rate less than adequate. Examples to address this issue is faster navigation to different 'important identified' sections of a document, image extraction and presentation and faster switching between different documents in a search. Another large requirement from the patent examiners are those of serendipity searching by navigation. This was defined as "starting from one document, to find all related documents to this starting document multidimensionality by filtering for author name and from backward and forward citations".

4. Summary and Future Work

We present work investigating the internal usage and perceived value of non-patent literature within prior art searching. From interviews with patent examiners we are able to produce high level guidelines to inform the creation of software to assist prior art searching for NPL. In future, we aim to present a prototype tool which is designed based on these guidelines and which will be able to elicit further detailed requirements while using a user-centered design approach to continue the development of the interface.

5. Acknowledgments and Disclaimer

We would like to thank the patent examiners for volunteering their free time and for their insights which made this research work possible. The opinions of the authors do not necessarily reflect the opinion of their employers.

References

[1] http://www.epo.org/about-us/annual-reports-statistics/statistics.html#filings (accessed – April 2017)

[2] Hingley, P., & Bas, S. (2009). Numbers and sizes of applicants at the European Patent Office. World Patent Information, 31(4), 285-298.

[3] van Zeebroeck, N., Stevnsborg, N., De La Potterie, B. V. P., Guellec, D., & Archontopoulos, E. (2008). Patent inflation in Europe. *World Patent Information*, *30*(1), 43-52.

[4] List, J. (2008). Free patent databases come of age. *World Patent Information*, *30*(3), 185-186.

[5] Creswell, J. (2006). So small a town, so many patent suits. *New York Times*, *24*.

[6] Badi, R., Bae, S., Moore, J. M., Meintanis, K., Zacchi, A., Hsieh, H., ... & Marshall, C. C. (2006, January). Recognizing user interest and document value from reading and organizing activities in document triage. In Proceedings of the 11th international conference on Intelligent user interfaces (pp. 218-225). ACM.

[7] Jonckheere C. The Internet for search project at the EPO. World Patent Inform 1997;19(2):93

[8] Black T, Hayes G, Lintz P, Purcell AF. Use of the Internet in the USPTO for the patent examination process. World Patent Inform 1997;19(3):153.

[9] Verhulst W. The impact of the Internet on prior-art searching in a patent environment. World Patent Inform 1997;19(2):95.

[10] Archontopoulos, E. (2004). Prior art search tools on the Internet and legal status of the results: a European Patent Office perspective. World Patent Information, 26(2), 113-121.

[11] Landwich, P., Vogel, T., Klas, C. P., & Hemmje, M. (2009). Supporting patent retrieval in the context of innovation-processes by means of information dialogue. World Patent Information, 31(4), 315-322.

[12] Annies, M. (2009). Full-text prior art and chemical structure searching in e-journals and on the internet–A patent information professional's perspective. World Patent Information, 31(4), 278-284.

[13] Bonino, D., Ciaramella, A., & Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. World Patent Information, 32(1), 30-38.

[14] Salampasis, M., & Hanbury, A. (2014). PerFedPat: An integrated federated system for patent search. World Patent Information, 38, 4-11.

[15] Adams, S. (2012). Survey of PCT search reports and the importance of the internet as a source of non-patent literature. World Patent Information, 34(2), 112-123.

[16] Baudour, F., & van de Kuilen, A. (2015). Evolution of the Patent Information World–Challenges of yesterday, today and tomorrow. World Patent Information, 40, 4-9.

[17] Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. World Patent Information, 37, 3-13.

[18] Winter, Sidney G. "Satisficing, selection, and the innovating remnant." The Quarterly Journal of Economics 85.2 (1971): 237-261.

[19] http://www.wipo.int/edocs/mdocs/scit/en/scit_6/scit_6_www_31685.pdf (Accessed - 2017)

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-217

Measurement of Open Access as an Infrastructural Challenge: The Case of Finland

Pekka OLSBO[1]
Open Science Centre, University of Jyväskylä

Abstract. Finland has set numeric goals for the development of open access. However, at the moment, no system is available by which this development could be monitored. Poor quality in the metadata records in universities' research information databases prevents metadata-based analysis of open access publishing progress. This paper shows how the quality problems of Finnish publication data can be resolved through centralizing the services and processes of metadata creation and by improving the interoperability of systems involved in the processes. As a result, this study describes an environment where reliable measurement of open access is possible and presents suggested actions for improving the Finnish publication data collection.

Keywords. open access, institutional repositories, current research information systems, metadata, measurement of open access

1. Introduction

Finland has set goals for achieving open science, following the requirements and recommendations defined by the EU commission [1]. The Finnish Ministry of Education and Culture has established key performance indicators for open science, including targets for open access publishing. According to these indicators, 65% of Finnish scientific publications should be open access by 2017. The percentage should be 75% in 2018, and by the year 2020, 90% of publications should be open access [2].

To achieve these goals, measurable performance, often in the form of numeric data, is required. However, such an outcome means that those overseeing the process would need to base their evaluation on reliable metadata records. Finland, however, does not currently have a national Current Research Information System (CRIS). There are, however, research information systems in each university in Finland. The Finnish research publications portal JUULI collects the metadata created and maintained in the research information systems of the universities [3]. This means that the metadata in the Finnish portal JUULI can be trusted only insofar as the quality of metadata in the research databases of the universities can be trusted.

While the poor quality of the metadata created in Finnish research institutes regarding all publications presents a serious problem [4], establishing a reliable measurement of open access is even more challenging. One element of this challenge

[1] Corresponding Author, P.O.BOX 35, 40014 University of Jyväskylä, Finland; E-mail: pekka.olsbo@jyu.fi.

results from the current Finnish publication data collection instructions that do not require the open access status of a publication as mandatory information

The metadata collected from the year 2015's publications demonstrate the problems in quality. That year, Finnish universities reported 33,720 publications in the Finnish JUULI portal. Table 1 shows that the metadata regarding the open availability of the publications are unreported in almost 50% of cases. On the basis of this information, the only conclusion that can be made with this metadata is that the share of open access publications in Finnish universities is not measurable.

Table 1. The open access status of Finnish research publications in JUULI portal (www.juuli.fi) 2015. Data collected in January 2017.

Category	Amount	%
Not OA	10639	31 %
Gold OA	5321	16 %
Other OA (Green)	1639	5 %
No answer	16121	48 %
Total	33720	

From the year 2016 onward, the division of open access was changed from the previous collection guidelines. Now data is also collected in the category of hybrid open access publications, which makes the definition of the open access status even more demanding.

2. The Challenge of the Quality of the Metadata

The key element in making the measurement of open access publishing reliable is a clear definition of open access. Archambault et al. [5] defined open access as divided into the rational definitions of open access and the operational definitions of open access. More specifically, the rational definitions used in Archambault et al. [5] are based on the propositions made by Peter Suber [6]. The operational definitions of open access refer to definitions established by the organization gathering and evaluating the measurement data of open access. (pp 2-5).

In the case of Finland, the operational definitions of open access are described in the document, Publication Data Collection Instructions for Researchers 2016 [7]. These definitions divide open access to three categories:

1. publications published via an open access channel,
2. open access publications published within a hybrid channel,and
3. publications self-archived in the publication archives of an organization or field of science, whether immediately available or after a reasonable embargo specified by the publisher

This division places a great demand on the quality of the metadata created about the publications of Finnish scholarly authors. In other words, each research publication created in Finland would need to be categorized reliably within the designated categories. Such a requirement for reliably measuring open access causes challenge number one: the quality of the metadata.

The challenge of the quality of the metadata also includes the problem of coverage. In order to calculate the share of open access publications, the total number of publications and total number of open access publications per year in Finland must be known. However, because the exact number of open access articles each year is not accurately collected, only estimates can be made about the open availability of research [8], [9].

The second challenge in measuring open access is quantifying the progress of open access. Archambault et al. [5] described why it is quite difficult to measure the growth of open access (OA):

> The reason is that growth in OA appears as the result of four main forces: (1) historical growth in the interest in OA which translates into new papers being increasingly available for free; (2) the growing interest in OA also translates into actors increasingly making available old papers for free; (3) OA policies that allow for delaying OA to scientific papers with embargo periods produce a concomitant disembargoing of scientific articles that creates additional growth in old papers being made available for free; and (4) the fact that the number of published scientific papers is growing, so even for a stable proportion of OA, the number of OA papers would keep growing (p. ii).

As noted earlier, the key performance indicators for open science in Finland are based on the share of open access publications. This makes Forces 1, 2 and 3 as described by Archambault et al. [5] relevant, thus exerting additional pressure on the need for quality metadata and the interoperability of the systems engaged in reporting processes.

The problems in metadata records generally have been categorized by Yasser [10] and further studied by Tani et al. [11]. From the five categories presented by Yasser [10], three can be applied to Finnish metadata:

> (1) incorrect values, i.e., metadata records contain values that do not represent a given resource correctly even though elements are applied correctly, (2) missing information, i.e., the metadata record is not complete and (3) information loss, i.e., some details characterizing the information are lost due to the conversion of metadata from one scheme to another or due the fact that metadata is not extracted from one system to another (pp. 59-60).

All these problems can be solved by improving the processes of metadata creation and/or the interoperability of the systems involved in the collection process. As Nicholas Joint [12] remarked, the more often libraries become the mediators of green open access, the higher the chance of quality metadata being produced. This is one reason why the entire process of metadata creation and maintenance should be centralized.

3. The Progress of Open Access in Finland, 2012–2016

The publication data from Finnish universities are gathered within the national JUULI portal. However, the metadata in the JUULI portal currently does not support a high-quality analysis of the open access availability of research publications. Nevertheless, by using the metadata, I could speculate on the progress of open access in Finland and analyze the metadata problems presented by Yasser [10] in greater detail. This is especially true when looking more closely at the development of green open access in Finland.

Finland's Ministry of Education and Culture launched the Open Science and Research Initiative [13] to both promote research information availability and to serve as the open science platform for the years 2014-2017. However, despite this goal at the national level, neither specific actions nor decisions resulting in open access publishing progress have taken place. Even at the university level, where many universities operate under a mandate for open access publishing, only a few universities are making a difference in the growth of open access. That is why, in the following analysis, I focus on describing the development in four universities in Finland who are contributing the most to this growth and for which mostly complete data is available. These include: Aalto University, the University of Helsinki, the University on Jyväskylä, and the University of Tampere.

Table 2 presents the number of open access articles in the four case universities for the years 2012–2015. These figures include both green and gold open access. The table indicates that measurable progress in open access publishing has occurred only at the Universities of Jyväskylä and Helsinki. Table 3 describes the total number of article publications from these case universities.

Table 2. The number of open access articles in four case universities. Data from the JUULI portal (www.juuli.fi).

University	2012	2013	2014	2015
Aalto	586	564	556	597
Helsinki	1020	1410	1691	1651
Jyväskylä	722	757	1114	1276
Tampere	843	714	934	767

Table 3. The total number of article publications in four case universities 2012–2015. Data from the JUULI portal (www.juuli.fi).

University	2012	2013	2014	2015
Aalto	3587	3769	3579	3360
Helsinki	10858	10867	10800	9777
Jyväskylä	2690	2847	2793	2738
Tampere	2416	2252	2430	2363

By dividing the number of open access publications by the total number of article publications per year, we can create a chart that depicts the development of open access at each case university (see Figure 1). The main question that this graph raises, however, is how reliable are these numbers? Can anything really be said about the development of open access in Finland, or even these universities, from these numbers?

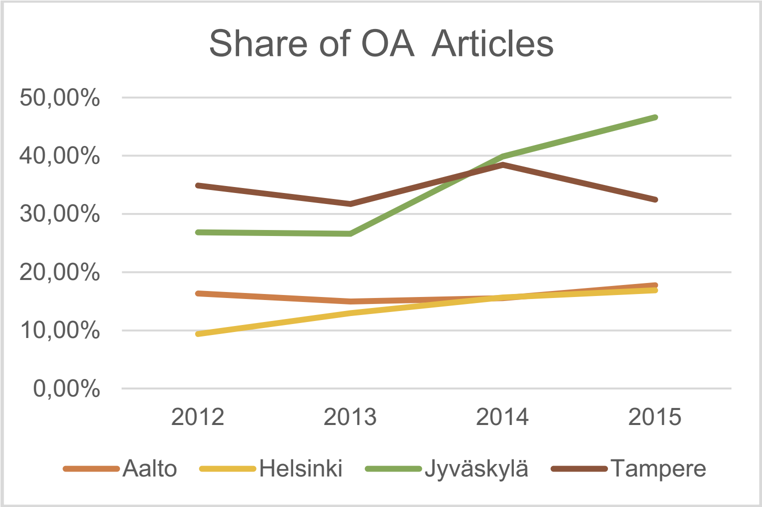


Figure 1. Share of open access articles in four case universities 2012-2015. Data from the JUULI portal (www.juuli.fi).

The reliability of metadata in the JUULI portal can be tested more accurately by looking closely at the development of green open access in these four case universities. During 2012–2015, green open access was categorized in the JUULI portal as "other OA availability." Table 4 shows the number of green open access articles in the four case universities for the years 2012–2015.

Table 4. The number of green open access articles in four case universities 2012–2015. Data form JUULI portal (www.juuli.fi).

University	2012	2013	2014	2015
Aalto	12	8	15	24
Helsinki	157	440	448	549
Jyväskylä	276	279	568	664
Tampere	77	51	179	178

Figure 2 shows the share of green open access articles in each case university. However, the share of green open access is very modest in three of the universities, with the exception being the University of Jyväskylä, which shows rapid development in recent years.

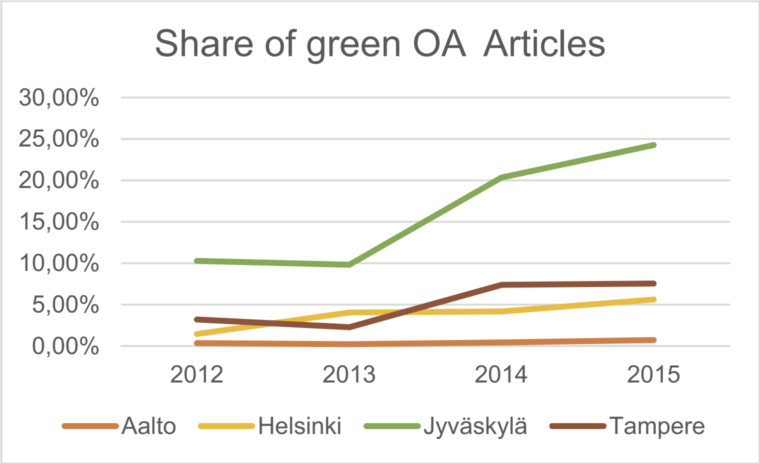


Figure 2. The share of green open access articles in four case universities 2012–2015. Data from the JUULI portal (www.juuli.fi).

In order to test the problems in metadata records categorized by Yasser [10] and Tani et al. [11], the original sources of the metadata must be used. In the case of green open access, this means the institutional repositories and the current research information systems of the case universities. Because both the repositories and the CRISs include data from the year 2016 as well, the analysis, as presented in the figure and tables, can be expanded to cover this additional data.

The data available in the institutional repositories of the universities regarding the number of deposited articles differ significantly from similar data available in the JUULI portal. Table 5 shows the number of deposited articles from the four case universities between 2012 and 2016. This comparison between Tables 4 and 5 demonstrates that no conclusions about the development of open access can be drawn using the data from the JUULI portal.

Table 5. The number of green open access articles in the repositories of four case universities 2012–2016. Data collected from the repositories Aaltodoc [15], Helda [16], JYX [17] and TamPub [18] on February 20, 2017.

University	2012	2013	2014	2015	2016
Aalto	134	162	249	277	448
Helsinki	586	806	978	1166	1598
Jyväskylä	265	371	642	889	1057
Tampere	207	154	236	266	333

I now analyze the problems with the metadata records in the JUULI portal using the three categories as developed by Yasser [10]:

1. *Incorrect values, i.e., metadata records contain values that do not represent a given resource correctly, even though elements are applied correctly.* In comparing the data in the JUULI portal to the original repositories and CRISs for the four case universities, information about the green open access is incorrect or missing in the JUULI portal in a large number of the metadata records. The inaccuracy of a value can be a result of human behavior or a lack of interoperability between the repository and the CRIS.
2. *Missing information, i.e., the metadata record is not complete.* As presented in Table 1, the most significant problem regarding the information about the open availability of publications in the JUULI portal is missing information. Once again, this is due to two reasons: The people responsible for maintaining the CRIS do not fill in the information about green open access, or this information is not extracted from the repository into the CRIS. In some cases, there is only a small qualitative difference between missing information and information loss. Missing information may be due to information loss.
3. *Information loss, i.e., some details characterizing the information are lost due to the conversion of metadata from one scheme to another or due the fact that metadata is not extracted from one system to another.* The basic infrastructure of green open access should be quite simple: Metadata are created or imported into the CRIS; the metadata are then transferred to the repository, where they are enriched with open access information; and then the updated records are imported back into the CRIS. These metadata are then transferred to the JUULI portal. But if the interoperability between the repository and CRIS, or between the CRIS and JUULI, is incomplete, then information is in danger of being lost.

As a result of the analysis above, the causes for the problems in the metadata records reflect two main factors: the human factor and the infrastructural factor. The elimination of problems related to these factors has been one key element for the development of self-archiving and centralized open science services at the University of Jyväskylä.

4. Conclusions: The "Jyväskylä Model" in Practice

The information gathered from research information systems in Finnish universities is not reliable enough for measuring open access development. This is true for both gold and green open access. Therefore, two essential actions must be implemented:

1. Gold, green, and hybrid open access papers must be identified and cataloged within the research information systems with high-quality reliability, and
2. The interoperability of institutional repositories and current research information systems must be improved so that the complete information about the deposited papers in the repository is exported automatically and accurately to the research information system.

As part of the project Finland: A Model Country for Green Open Access, we have studied these processes in the Open Science Centre of the University of Jyväskylä. We have created a model that results in high-quality metadata creation and automated processes between the repository and the CRIS [14]. The basic idea of the "Jyväskylä model" is to centralize all aspects of the self-archiving and open access processes lying within the responsibility of the professionals at the university library. Even within the library, just a few professionals handle the metadata input and the repository–CRIS processes. Researchers do as little as possible and, in some cases, nothing at all. The motivation for this structure is that these processes and workflows need expertise and extensive knowledge about various aspects of publishing and project work, such as the legal aspects, publisher policies, funder mandates, and requirements set by the Finnish government ministries. When experienced professionals attend to collecting the necessary data on publications by the university's researchers, the creation of metadata is much faster and more reliable. Additionally, this system is also cost effective in that fewer people are needed to input and transfer the information.

The infrastructure of the Jyväskylä model is described in Olsbo et al. [15] and it is based on automated movement of metadata between the repository and CRIS. This structure ensures that all changes and additions to the metadata are included in both systems. Consequently, all the metadata transferred to the national JUULI portal is up to date and complete.

This model of centralized publications data gathering, particularly related to open access services, was implemented at the University of Jyväskylä in 2014. Since then, the share of green open access publications at the university has risen from 16% to 51%. Moreover, the staff resources needed for tracking all aspects of university publications, as well as for reporting the research activities to the Ministry of Education and Culture, have been reduced considerably. Figure 3 shows the development of openness of peer reviewed articles in the University of Jyväskylä 2015–2016.

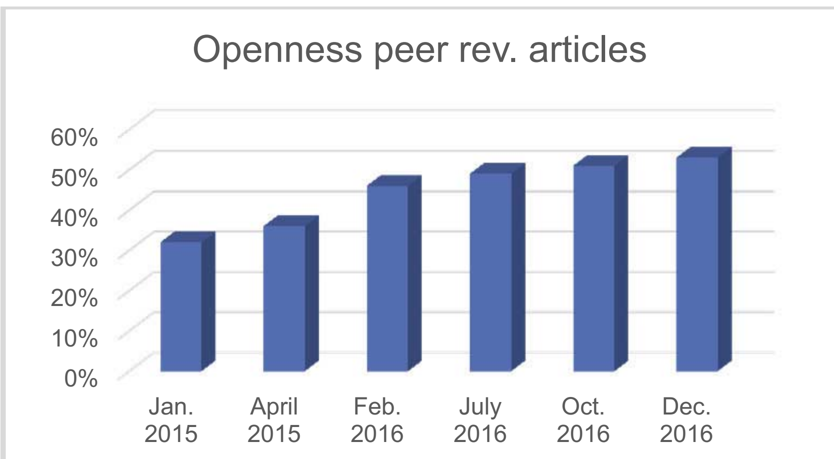


Figure 3. Development of openness of peer reviewed articles in the University of Jyväskylä 2015–2016. Data from the TUTKA research information database (http://tutka.jyu.fi/tutka/).

High-quality metadata and the seamless interoperability of the repository and CRIS ensure the possibility of reliably analyzing the development of open access.

Figure 3 shows the development of open access peer reviewed articles at the University of Jyväskylä for the years of 2015 and 2016.

A prototype has been developed by the University of Jyväskylä for open access monitoring that automatically analyzes the metadata in the CRIS and gathers up-to-date information about the development of open access. The model of practicing and monitoring open access created in the University of Jyväskylä can be implemented at most universities in Finland. When this model is combined with the new national VIRTA Publication Information Service and system architecture [20], it would enable the reliable analysis and assessment of the open access development in Finland.

Acknowledgements

This study was funded by the Finnish Ministry of Education and Culture's Open Science and Research initiative (see more: http://openscience.fi/).

References

[1] European Commission. Commission recommendation of 17.7.2012 on access to and preservation of scientific information. Available at: http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

[2] Niinimäki, S. The Open Science and Research Initiative in Finland. Presentation, Annual JPI Conference, Brussels, 15.1.2016. Available at: https://www.era-learn.eu/events/annual-joint-programming-2015-new-date-2016/topic-3-strategies-for-fostering-open-knowledge-and-open-access-in-research/02_ERALEARN_2020_Niinimki.pdf

[3] Finnish research publications portal JUULI. Available at: http://www.juuli.fi/

[4] Ilva, J. Towards Reliable Data – Counting the Finnish Open Access Publications. Procedia Computer Science, Volume 106, 2017, pp 299-304. http://dx.doi.org/10.1016/j.procs.2017.03.029.

[5] Archambault, E., Amyot, D., Deschamps, P., Nicol, A., & Provencher, F. Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013. Copyright, Fair Use, Scholarly Communication, etc. 2014; Paper 8. Available at: http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1007&context=scholcom

[6] Suber, P. Open Access. Cambridge, Mass.: MIT Press; 2012. Available at: https://mitpress.mit.edu/sites/default/files/9780262517638_Open_Access_PDF_Version.pdf

[7] Ministry of Education and Culture. Publication data collection instructions for researchers 2016, pp. 14-15. Available at: https://confluence.csc.fi/display/suorat/Julkaisutiedonkeruun+tutkijaohjeistukset

[8] Bjork B., Roos A., & Lauri, M. Scientific journal publishing: yearly volume and open access availability. Information Research. 2009; 14(1). Available at: https://helda.helsinki.fi/bitstream/handle/10227/615/bjork_roos_lauri.pdf

[9] Khabsa M. & Giles C.L. The number of scholarly documents on the public web. PLoS One. 2014; 9(5): e93949. pp 59-60. http://dx.doi.org/10.1371/journal.pone.0093949

[10] Yasser, C.M. An Analysis of Problems in Metadata Records. Journal of Library Metadata, 2011; 11:2, pp. 51-62. DOI: 10.1080/19386389.2011.570654. Available at: http://dx.doi.org/10.1080/19386389.2011.570654

[11] Tani, A., Candela, L., & Castelli, D. Dealing with metadata quality: The legacy of digital library efforts. Information Processing & Management. Vol. 49, Issue 6, November 2013, pp. 1194–1205. http://dx.doi.org/10.1016/j.ipm.2013.05.003

[12] Joint, N. Institutional repositories, self-archiving and the role of the library. Library Review, 2006; Vol. 55 Iss: 2, pp.81–84. http://dx.doi.org/10.1108/00242530610649576

[13] Ministry of Education and Culture. Open science and research leads to surprising discoveries and creative insights. Open science and research roadmap 2014–2017. Reports of the Ministry of Education and Culture, Finland 2014: 21. Available at: http://www.minedu.fi/OPM/Julkaisut/2014/Avoimen_tieteen_ja_tutkimuksen_tiekartta_2014_2017.html?lang=en

[14] Suomi rinnakkaistallentamisen mallimaaksi (SURIMA) –hanke [Finland – A model country for green open access project]. Project publications. Available at: https://jyx.jyu.fi/dspace/handle/123456789/52527
[15] Olsbo, P., Hautala, T. & Kananen, J. Modelling centralized self-archiving process in Finland. Poster in Repository Fringe conference, Edinburgh, 1.8.2016. Available at: http://urn.fi/URN:NBN:fi:jyu-201612295282
[16] Aaltodoc – Aalto University publication archive. Available at: https://aaltodoc.aalto.fi/
[17] Helda – Digital Repository of the University of Helsinki. Available at: https://helda.helsinki.fi/
[18] JYX – Jyväskylä University Digital Archive. Available at: https://jyx.jyu.fi/
[19] Tampub – institutional repository of the University of Tampere. Available at: http://tampub.uta.fi/
[20] VIRTA Publication Information Service. CSC – IT Center for Science Ltd; 2017. Available at: https://confluence.csc.fi/display/VIR/Esittelymateriaaleja?preview=/50870665/68724326/VIRTA%20publication%20information%20service%20in%20Finland.pdf

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-227

The Transformation of the Ktisis Repository into a Current Research Information System (CRIS)

Alexia NTINI-KOUNOUDES[a,1] and Marios ZERVAS[a]
[a] *Cyprus University of Technology*

Abstract. Institutional repositories have played a major role in universities worldwide during the last decade. Such systems are developed with the aim to collect and disseminate the research activities of universities. They provide access to and showcase research outputs, and therefore they have become an essential infrastructure for universities. A repository provides the means to properly preserve research outputs and can also be used for research monitoring and assessment. In this case study, we concentrate on the transformation of Ktisis, the institutional repository of the Cyprus University of Technology, into a Current Research Information System (CRIS). A CRIS system records, processes, and presents metrics and figures related with research activity throughout its life cycle. Particular emphasis is given to the results of research activities (publications, patents, research data) and their connection with the environment within which they were created (researchers, organizations, funded programs and projects, research infrastructures, services). In this case study we will describe the procedures followed in order to transform Ktisis into a CRIS system together with the implementation of the integration of ORCID identifiers within the system. Particular attention will be paid to the challenges we came across throughout the process and how we overcame these difficulties and problems. Ktisis is the institutional repository developed and maintained by the Library and Information Services at the Cyprus University of Technology. Ktisis was created in 2008 using the open source software DSpace after the University's Interim Governing Board made the decision that all the research products of academic members must be deposited in the Library. In subsequent years, the mandatory deposit of undergraduate, MSc and PhD theses was also imposed. In early 2015 it was decided that the Cyprus University of Technology (CUT) must become a member of ORCID and the Library undertook the project to integrate ORCID in its systems. At the same time the Library decided that the best way to move forward was to transform Ktisis into a CRIS system using DSpace-CRIS, an extension to DSpace, in order to integrate ORCID with the repository and to take advantage of all the functionalities provided by a CRIS system.

Keywords. repositories, CRIS, ORCID

1. Introduction

Institutional repositories have played a major role in universities worldwide during the last decade. Such systems are developed with the aim to collect and disseminate the

[1] Alexia Ntini-Kounoudes, Library and Information Services, Cyprus University of Technology, Cyprus; Email: marios.zervas@cut.ac.cy.

research activities of universities. They provide access to and showcase research outputs. As such, they have become an essential infrastructure for universities. A repository provides the means to properly preserve research outputs and can also be used for research monitoring and assessment. In this case study, we concentrate on the transformation of Ktisis, the institutional repository of the Cyprus University of Technology, into a Current Research Information System (CRIS). A CRIS system records, processes, and presents metrics and figures related with research activity throughout its life cycle. Particular emphasis is given to the results of research activities (publications, patents, and research data) and their connection with the environment within which they were created (researchers, organizations, funded programs and projects, research infrastructures, and services)[2]. In this case study we will describe the procedures followed in order to transform Ktisis into a CRIS system together with the implementation of the integration of ORCID identifiers within the system. Particular attention will be paid to the challenges we came across throughout the process and how we overcame these difficulties and problems.

Ktisis is the institutional repository developed and maintained by the Library and Information Services at the Cyprus University of Technology. Ktisis was created in 2008 using the open source software DSpace[3] after the University's Interim Governing Board made the decision that all the research products of academic members must be deposited in the Library. In subsequent years the mandatory deposit of undergraduate, MSc and PhD theses was also imposed. In early 2015 it was decided that the Cyprus University of Technology (CUT) must become a member of ORCID[4] and the Library undertook the project to integrate ORCID in its systems. At the same time the Library decided that the best way to move forward was to transform Ktisis into a CRIS system using DSpace-CRIS, an extension to DSpace, in order to integrate ORCID with the repository and to take advantage of all the functionalities provided by a CRIS system.

2. Implementation

2.1. The ORCID Project

The task of identifying researchers and linking them to their research work is difficult and challenging since the researchers can be very active and they frequently move between organizations. Additionally, they often use different variations of their name in their work which means that the information about them needs frequent updating, a complicated procedure in and of itself.

The introduction of CRIS systems comes as a solution for collecting, managing, preserving, analyzing and showcasing the research output of institutions, providing the use of persistent identifiers for uniquely identifying researchers. These identifiers that are being used are the ORCID IDs, which can be used by a researcher throughout his career.

The Library worked on the implementation of ORCID for researchers by creating a portal (http://library.cut.ac.cy/orcid/) where the researcher can create an ORCID or connect an existing one with the University's system, thus enabling the authentication,

---

[2] http://metrics.ekt.gr/en/systimata-cris
[3] http://www.DSpace.org
[4] http://www.orcid.org

reading, updating, adding and synchronizing of research outputs with the Ktisis repository.

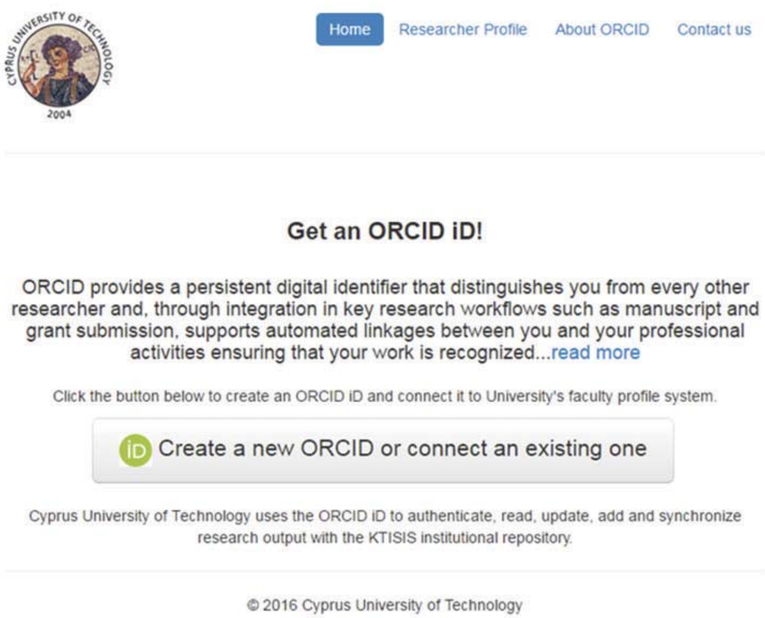


Figure 1. The Library's ORCID portal

When a researcher creates an ORCID through the platform, details such as their name, ORCID id, authorization token and ORCID creation date are written in a log file in the system server which can later be used by the Library for various purposes.

At the moment, the latest version of DSpace-CRIS being used in Ktisis does not support the automatic synchronization of the researchers work between DSpace and ORCID. This is a feature of the system that will be available in subsequent releases. Therefore, we use this file to export the ORCID data of the researcher we are interested in. The available options are to export: the entire profile data for a researcher, bibliographic data, research work data, funding data, or affiliations data. These options are the available information that can be exported through the ORCID. We are using an interface where we supply the ORCID with the researcher we are interested in; the system checks in that log file to see if that researcher already has created an ORCID through Ktisis and if a match is found then it proceeds with the export of one of the available options mentioned above.

However, the problem with the exported file is that it is structured in the ORCID-xml format which makes it difficult to process and transform it into a CSV format suitable for importing into DSpace. The Library at this moment is trying to find a way to convert the exported file with the researcher's works data into a CSV file in order to be able to import this data into Ktisis.

The combination of ORCID and DSpace-CRIS is very powerful as it utilizes the ORCID's ability to improve data quality by connecting researchers with their research while also adding significant value to their work.

2.2. The Transformation of Ktisis to CRIS

Ktisis was created using the open source software DSpace in 2008. In the following years the Library upgraded the system according to new releases, until it was decided that the best way to move forward and satisfy the need for the use of persistent identifiers, and to be able to handle the growing volume and variety of research information, was to transform Ktisis into a CRIS system. The DSpace community had already made available an extension of DSpace, the DSpace-CRIS module, enriching DSpace with CRIS entities and concepts. DSpace-CRIS extends the DSpace data model providing the ability to collect, store, manage, retrieve, expose and exchange data about all research entities (people, organization units, projects, grants, awards, patents, publications, etc.).

The Library collaborated with the company 4Science[5] to transform Ktisis into a CRIS system as previous attempts to perform this task on our own were unsuccessful. The plan was first to clone the live Ktisis system into a testing server so we have a spitting image of the system at that moment. Then all the work needed in order to upgrade Ktisis into a CRIS system was going to be carried out on this staging server and when everything was tested and fine-tuned, the same changes were going to be applied on the live server.

Subsequently, the project was split into 3 phases. The first phase involved the creation of the specifications for the new system and meeting the pre-conditions set, i.e. providing the 4Science staff with SSH and VPN access to the staging environment. The same access was also provided for the production environment.

We asked the Library staff working on Ktisis to stop all activities including the submission of new material. At this point the live system was cloned into the test server so that all the changes would be applied and tested there first.

When these tasks were finished we moved on to Phase 2. The first task in this phase was to configure and create the necessary scripts needed to convert the current repository structure into the structure used in the DSpace-CRIS version. This work was done by the staff of 4Science. When this was completed and applied to the test server, they proceeded with the installation on the staging server of the new configuration leading to the finalized DSpace-CRIS version. A database dump was also taken from the live system that would be imported on the staging server after the installation of the new software was in place.

Before the installation started, we had to perform a number of tasks. We were advised at this point that Tomcat was configured to deal directly on port 80, which is a bit unusual and not the recommended approach since it reduces the performance of the system. The solution to this was to have Apache HTTP run in front of Tomcat. These changes saw the visible improvement of the system performance when the system was up and running. Additionally, it was observed that SOLR[6] was running using the same Tomcat service; therefore, we created a dedicated Tomcat instance for SOLR and noticed again that this provided better performance. SOLR is the indexing server used

---

[5] http://www.4science.it/en/

[6] http://lucene.apache.org/solr/features.html

as a part of Discovery in DSpace to speed up access to content metadata and statistics and it also provides faceting, search results filtering and "More like this" functionality.

We then had to redesign the community hierarchy and data structure. Since DSpace-CRIS is centered on researchers and their work, we changed the community hierarchy structure so that it was not based around faculties and departments anymore but rather on separate entities for each document type we stored in our system (for example articles, books, conference papers, etc.). Then each item was connected to the researchers that had authored it and was allocated to the correct collection.

The data model of DSpace-CRIS entities defines the following 1st level entities: people, organizational units and projects. The first entity we had to think of in terms of design was the organizational units. An organizational unit was created in order to represent the University as a whole and then other organizational units were created to represent each faculty and department. Additionally, we had to list together all the external affiliations that existed in the system database (as collaborations) and make sure that these values were normalized before they were created in the new system. This was important as we wanted to have a system clear of data discrepancies and duplications. Having such a hierarchy provides better statistics for the faculties and departments and also for individual researchers.

When the new community hierarchy was decided we had to provide the mapping between the community hierarchy in the old system and the new one in order to be able to move the items in the database to the correct location in the new CRIS system. Each collection that existed in the old system was mapped to a collection in the new system, using the collection's name and handle in the old system.

At this stage we also had to devise a list of the researchers of the University, their department and faculty and their ORCID IDs. This file was then used in order to import the researcher names into the system, map them to the correct faculty and department and subsequently to create a researcher page for each one of them. Then specific procedures were followed in order to connect their research output already available in Ktisis with their profiles. This way the researchers have access to a dedicated set of functionalities. They can edit their personal information and decide what the visibility of this information will be (public or hidden). The researchers are also able to manage the research output connected to their profiles, such as publications. Furthermore, they can connect their profile with their ORCID and manage the synchronization preferences available in order to transfer information from Ktisis to their ORCID profile, automatically or manually. At the moment only the ability to synchronize the researcher profile is available. The ability to synchronize the researchers' works is currently under implementation by the developers of DSpace-CRIS.

When the installation was complete and the above mentioned tasks were finished, we had to go through the system in order to test if everything was according to specifications. The Library staff spent a lot of time reviewing the system and we were in constant communication with 4Science in order to fix any problems that were observed. We allowed a full month for the testing procedure as we needed to make sure that everything was in the right order. When both sides were satisfied with the results, we proceeded to Phase 3 which included the application of all the configuration changes to the live system. Again we went through a short period of testing the new system until we were in the position to go live.

3. The Day After

When the new system was live, it was decided that it would be presented to the various departments of the University in order to get feedback from the researchers. We were very keen to find out what they thought about the new Ktisis and the new functionalities it provides. Before we started the transformation of Ktisis to a CRIS system, the Library promoted the ORCID functionality and the benefits that it provides to the researchers. It was at this point that the University senate decided that all the researchers of the University must acquire an ORCID, giving the Library greater motivation to go on with the project.

Most of the feedback we received from the researchers was very positive. They really appreciated the ability to have all the information about them and their research work gathered together under their Researcher Profile. Of course there were cases where some researchers were not very keen with the idea of introducing bibliometrics to the system, showing how they rank up in the databases that are being tracked (Web of Science and Pubmed), but these were isolated cases.

Using the system, we realized that there are great functionalities available for administrators as well. Using the DSpace-CRIS software makes it easy to perform tasks that had to be done manually in the past or through running SQL queries on the database, such as exporting statistics for particular departments or organizational units, or batch updating the records through the import/export functionalities available.

Additionally, the task of submitting new items in the system is also easier now. Using only the publication identifiers, the user is able to search through the available databases. A list of matching publications is subsequently shown in order to proceed with the submission process, thus saving time that was needed to manually enter all the details of a publication.

Another important addition in the new system is the introduction of Projects. Projects is another entity of DSpace-CRIS where information about various projects undertaken by the University's researchers can be found. The system has been configured in such a way so that information such as the project title, project abstract (i.e. details about the nature of the project), and some primary data (such as the project coordinator, project start and completion date as well as the current status of the project) are all displayed. The most important functionality here is that the list of all the publications that were carried out during this project can be seen on the project page. This is achieved by connecting the publication with the project through the Dublin core field that was assigned specifically for this purpose.

4. Lessons Learnt

The project of transforming Ktisis into a CRIS system was not an easy and straightforward task. We have tried in the past to perform the transformation on our own since the software is open source and available for everybody to use. However, since the design and structure of DSpace-CRIS is totally different to the simple and easily customizable DSpace software, we did not succeed in creating the system we longed for. There is a lot of work that needs to be done in the configuration, especially for existing systems and since this is a relatively new release, the documentation available was not enough. After two failed attempts to configure the software on test servers and since we considered this improvement to be of great importance for the

University, we decided to ask for help from the original developers of the software and begin a collaboration with them in order to speed things up. Our experience showed that the transformation of a DSpace system to DSpace-CRIS is a procedure that is achievable but it will require a vast amount of time to reach the desired result.

When we started this project we decided that we would keep the infrastructure as it was. Having followed the guidelines of the University's IT team, the way that the software and files had been installed from the very first DSpace installation was how the installed files and middleware were stored on the NFS storage of our network. After a few weeks of using the system, it was observed that the NFS storage was rather slow when writing to disk, thus worsening the system performance. On top of that the database connections were released more slowly, using more RAM. We were not aware that this infrastructure was making our system less robust with reduced performance. The Library will fix this problem by moving the middleware, such as Postgres and Tomcat, and the DSpace-CRIS installation folders onto the local server partition to reduce these issues. Hopefully, this change will increase the performance of the server and minimize the occasional problem of the server being down due to improperly closed database connections.

Another problem that we observed throughout the use of the new system is that we have to be very careful with the submission process of new items. We have already been through the procedure of normalizing the external organizational units that have been created in the system in order to eliminate data duplications and discrepancies. However, we noticed that there are cases when new items are submitted where the submitter (authorized library staff) does not pick up the value from the drop down list that shows the available organizations but instead types in a new one. This has caused problems with the data since the numbers of items belonging to an organizational unit is not correct. It is an easy task to fix the data but difficult to locate the problem. The data can easily be fixed by exporting all the items for the specific "invalid" organizational unit in a CSV format, make the necessary changes of fixing the name of the unit and then importing the file back to the system using the import functionality. Then the invalid organizational unit can be deleted from the administrator user interface.

One of the most important lessons that were learnt throughout this project is that the DSpace-CRIS configuration is messy and very delicate. What we have done in order to avoid any problems on the live system is kept the staging environment in place so that any changes we wish to do are first applied there. If no problems occur, we proceed with applying the change on the live environment. What we plan to do in the near future is to again clone the live system into the staging server as to sync the two systems – since a number of changes were applied directly on the live server, causing a few problems.

5. Future Work

One of the tasks that will be undertaken soon is the creation of translation files for the Greek language for the key messages that are missing from the message catalogue. This is important to us since Ktisis comes both in Greek and English languages and we want to have a proper Greek version of the system. The files produced will be made available to the community through 4Science, the developers of DSpace-CRIS.

Another task that we will work on in collaboration with the DSpace-CRIS developers is the implementation of the capability to import a researcher's work from ORCID to DSpace-CRIS, and vice-versa. This is very important for the Library since using this functionality will make it easier to import all the research output of the University's academics and have a complete listing of their publications.

Additionally, we will continue the effort to keep the researchers' publications up to date by exporting the data from databases such as Web of Science and Scopus and importing them to the system. This, in combination with the new functionality of importing publications directly from a researchers ORCID profile will keep Ktisis up to date.

The Library has put a lot of effort on this project and will continue to work in order to keep the system up to date following new releases of the software and implementing any new functionalities that will be made available in future releases.

References

[1] David T. Palmer, Andrea Bollini, Susanna Mornati, Michele Mennielli, Achieving Visibility with a CERIF Compliant Open Source System. Procedia Computer Science, Volume 33, 2014, Pages 118-123. 12th International Conference on Current Research Information Systems, CRIS 2014 – Managing data intensive science – The role of Research Information Systems in realising the digital agenda.

[2] Nick Sheppard (2010). Learning How to Play Nicely: Repositories and CRIS. Friday, 30 July 2010, Ariadne Issue 63. Retrieved from http://www.ariadne.ac.uk/issue64/wrn-repos-2010-05-rpt

[3] euroCRIS - http://www.eurocris.org/

[4] Miika Nurminen (2014). Preparing for CRIS: Challenges and Opportunities for Systems Integration at Finnish Universities. Paper presented at Open Repositories 2014, At Helsinki, Finland

[5] Josh Brown (2016). ORCID + Pure: a "happy marriage" that makes researchers' lives easier. Collaboration allows researchers and universities to automatically connect their curated data in Pure with a researcher's ORCID iD. Retrieved from https://www.elsevier.com/connect/orcid-pure-a-happy-marriage-that-makes-researchers-lives-easier

[6] CERIF - the Common European Research Information Format. European Scene by Keith G Jeffery (http://ercim-news.ercim.eu/en68/european-scene-qsupport-of-the-research-processq/cerif-the-common-european-research-information-format)

[7] http://duraspace.org/now-available-first-release-dspace-cris-module-dspace-3

[8] https://wiki.duraspace.org/display/DSPACECRIS/DSpace-CRIS+Home

[9] http://www.doria.fi/handle/10024/97722

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-235

Web Technologies: A Survey of Their Applicability to Metadata Aggregation in Cultural Heritage

Nuno FREIRE[a,1], Hugo MANGUINHAS[b], Antoine ISAAC[b], Glen ROBSON[c] and John Brooks HOWARD[d]
[a] *INESC-ID*
[b] *Europeana Foundation*
[c] *National Library of Wales*
[d] *University College Dublin*

Abstract. In the World Wide Web, a very large number of resources are made available through digital libraries. The existence of many individual digital libraries, maintained by different organizations, brings challenges to the discoverability and usage of these resources by potential users. A widely-used approach is metadata aggregation, where a central organization takes the role of facilitating the discoverability and use of the resources, by collecting their associated metadata. The central organization has the possibility to further promote the usage of the resources by means that cannot be efficiently undertaken by each digital library in isolation. This paper focuses in the domain of cultural heritage, where OAI-PMH has been the embraced solution, since discovery of resources was only feasible if based on metadata instead of full-text. However, the technological landscape has changed. Nowadays, with the technological improvements accomplished by network communications, computational capacity, and Internet search engines, the motivation for adopting OAI-PMH is not as clear as it used to be. In this paper, we present the results of our initial analysis of available potential technologies, in particular, the following: IIIF (International Image Interoperability Framework); Webmention; Linked Data Notifications; Sitemaps; ResourceSync; Open Publication Distribution System (OPDS); and the Linked Data Platform.

Keywords. metadata, cultural heritage, linked data, web technology, standards

1. Introduction

In the World Wide Web, a very large number of resources is made available through digital libraries. The existence of many individual digital libraries, maintained by different organizations, brings challenges to the discoverability and usage of the resources by potential interested users.

An often-used approach is metadata aggregation, where a central organization takes the role of facilitating the discovery and use of the resources by collecting their associated metadata. Based on these aggregated datasets of metadata, the central

[1] Corresponding author, Rua Alves Redol 9, 1000-029 Lisboa, Portugal; Email: nuno.freire@tecnico.ulisboa.pt.

organization (often called aggregator) can further promote the usage of the resources by means that cannot be efficiently undertaken by each digital library in isolation. This scenario is widely applied in the domain of cultural heritage, where the number of organizations with their own digital libraries is very large. In Europe, Europeana has the role of facilitating the usage of cultural heritage resources from and about Europe, and although many European cultural heritage organizations do not yet have a presence in Europeana, it already holds metadata of resources originating from more than 3,500 providers[2].

This domain is also characterized by users that often have very specific information needs, which cannot be easily fulfilled by the Internet search engines. The retrieval of resources based on metadata, in combination with the hypertext documents of the World Wide Web, has been a challenge that the search engines have not yet been able to provide an effective solution for, therefore the retrieval of cultural heritage resources via search engines is ineffective.

The technological approach to metadata aggregation has been mostly based on the OAI-PMH protocol, a technology initially designed in 1999. OAI-PMH was meant to address shortcomings in scholarly communication by providing a technical interoperability solution for discovery of e-prints, via metadata aggregation. The cultural heritage domain embraced the solution offered by OAI-PMH, however, the technological landscape around our domain has changed. Nowadays, cultural heritage organizations are increasingly applying technologies designed for the wider interoperability on the World Wide Web. Particularly relevant for our work are those related with the social web, the web of data, internet search engine optimization, and the IIIF (International Image Interoperability Framework).

In this paper, we present the first results of our work in attempting to rethink our technological approach for metadata aggregation, with the goal of finding a solution to make the continuous operation of the aggregation network more efficient and to lower the technical barriers for data providers to give their contribution to Europeana. This paper makes the following contribution to the digital libraries community:

- An analysis of requirements for metadata aggregation based on a large network of data providers – the Europeana Network.
- A functional analysis for innovative use of state of the art technologies.
- A real-world application experience of open standards, thus contributing for their future improvement.

The paper will describe, in Section 2, the technological approach to metadata aggregation most prevalent in cultural heritage. Specific requirements, which guided our technological survey, are presented in Section 3. The Web technologies that were analyzed are presented in Section 4, and Section 5 concludes and introduces potential options for future work.

2. Metadata Aggregation in Cultural Heritage – Past and Present

In the cultural heritage domain, the technological approach to metadata aggregation has been mostly based on the OAI-PMH protocol, a technology initially designed in 1999 [1]. OAI-PMH was originally meant to address shortcomings in scholarly

[2] http://statistics.europeana.eu/europeana (consulted on 4th of January 2017)

communication by providing a technical interoperability solution for discovery of e-prints, via metadata aggregation.

The cultural heritage domain embraced the solution offered by OAI-PMH, since discovery of resources was only feasible if based on metadata instead of full-text [2]. In Europe, OAI-PMH had one of its largest, and earliest, applications in The European Library [3], which aggregated digital collections and bibliographic catalogues from 48 national libraries. It was also the technological solution adopted by Europeana since its start, to aggregate metadata from its network of data providers and intermediary aggregators [4].

However, the technological landscape around our domain has changed. Nowadays, with the technological improvements accomplished by network communications, computational capacity, and Internet search engines, the discovery of resources, such as e-prints, is largely based on full-text processing, thus the newer technical advances, such as ResourceSync [5], are less focused on metadata. Within the cultural heritage domain metadata-based discovery remains the most widely adopted approach since a lot of material is not available as full-text. The adoption of OAI-PMH for this purpose is not as clear as it used to be, however. OAI-PMH was designed before the key founding concepts of the Web of Data [6]. By being centered on the concept of repository, instead of centering on the resources, the protocol is often misunderstood and its implementations fail, or are deployed with flaws that undermine its reliability [2]. Another important factor is that OAI-PMH predates REST [7]. Thus, it does not follow the REST principles, further bringing resistance and difficulties in its comprehension and implementation by developers in cultural heritage organizations.

An additional aspect relevant for our work, is that nowadays, cultural heritage organizations are increasingly applying technologies designed for wider interoperability on the World Wide Web. Particularly relevant are those related with Internet search engine optimization and the International Image Interoperability Framework [9]. Regardless of the metadata aggregation process for Europeana, cultural heritage institutions are already interested in developing their systems' capabilities in these areas. By exploring these technologies, the participation in Europeana of these institutions may become much less demanding and possibly even transparent.

The cultural heritage domain has some specific characteristics, which have heavily influenced how metadata aggregation has been conducted in the past. We consider the following to be the most influential:

- Several sub domains compose the cultural heritage domain: Libraries, Archives and Museums (the term LAM is often used to refer to the three sub domains).
- Interoperability of systems and data is scarce across sub-domains, but it is common within each sub-domain, both at the national and the international level.
- Each sub-domain applies its specific resource description practices and data models.
- All sub-domains embrace the adoption and definition of standards based solutions addressing description of resources, but to different extents. A long-time standardization tradition has existed in libraries, while this practice is more recent in archives and museums.

- Several of the adopted standards tend to be flexible towards data structure. Standards based on relational data models, for example, are rare in cultural heritage, while XML-based data models are common.
- Organizations typically have limited budgets to devote to information and communication technologies, thus the speed and extent of innovation and adoption of new technologies is slow.

In this environment, a common practice has been to aggregate metadata, under an agreed data model that allows the data heterogeneity between organizations and countries to be dealt with in a sustainable way. These data models typically address two main requirements:

- Retaining the semantics of the original data from the source providers
- Supporting the information needs of the services provided by the aggregator

These two requirements are typically addressed in a way that keeps the model complexity low, with the intention of simplifying the understanding of the model by all kinds of providers, and to allow for a low barrier of implementation of data conversion solutions, by both providers and aggregators.

Another relevant aspect of metadata aggregation is the sharing of the sets of metadata from the providing organizations to the aggregator. The metadata is transferred to the aggregator, but it continues to evolve at the data provider, thus the aggregator needs to periodically update its copy of the data. In this case, the needs for data sharing can be described as a cross-organizational data synchronization problem.

In the cultural heritage domain, OAI-PMH is the most well established solution to address the data synchronization problem. Since OAI-PMH is not restrictive in terms of the data model to be used, it allows the sharing of the metadata per the adopted data model of each aggregation case. The only restriction imposed by OAI-PMH is that the metadata must be represented in XML.

In the case of Europeana, the technological solutions around the Europeana Data Model (EDM) [8] have always been under continuous improvement. However, the solution for data synchronization based on OAI-PMH has not been reassessed since its adoption.

The Web Technologies, presented in the following sections, address mainly the data synchronization problem, since the common data model based on EDM is intended to remain in usage. EDM does not impose any obstacles in the choice of Web technologies for this purpose, the data synchronization can be addressed with a wide variety of technologies. This comes from EDM following the principles of the Web of Data, and that it can be serialized in XML and in RDF formats.

3. Requirements for Cross-Organizational Data Synchronization

The synchronization of data sources is a general problem, for which computer scientists have provided many possible solutions. The type of solution applicable to each case is greatly influenced by the requirements of the application scenario, mainly in terms of data consistency guarantees and synchronization latency.

We focus on the scenario of data synchronization across data sources from different organizations. We define the requirements for the solution by considering the characteristics of the cultural heritage domain, mentioned in the previous section, along with some particularities of the metadata aggregation carried out in the Europeana network of data providers and aggregators.

The solution must allow an aggregator to collect structured metadata about the digital resources that a cultural heritage organization (the provider) wants to make available in Europeana. A solution should address the following requirements:

- The set of resources for aggregation is specified by the provider, and may comprehend all the resources of a digital library, or just a subset.
- The set of aggregated resources may evolve over time; therefore, the synchronization process must provide efficient mechanisms for incremental aggregation that will happen over time.
- The synchronization process between the provider and Europeana must be automatic and efficient, in terms of computation and network communication.
- The synchronization mechanism must be scalable to the level of the largest datasets nowadays available in Europeana, which are in the range of 2-5 million resources.
- The solution should be simple to adopt by data providers. One of the following aspects would make a solution simple to adopt:
    - It is based on technologies already in use by data providers;
    - It has very simple technical requirements for implementation;
    - Open source and free tools exist for deploying the solution.
- The solution may be more technologically challenging on the aggregator's side than on the data providers', since the aggregators are often better prepared to address more complex technical implementation issues of information systems.

In the context of the above requirements, the following section will present the Web technologies that we identified as possible solutions.

4. Web Technologies for Metadata Aggregation

Most of the technologies described in this section were designed for fulfilling the needs of general use cases, and are applicable across several domains. Some of these can completely fulfil the requirements of metadata aggregation, while others only do so partially, and need to be combined with other technologies. Not all technologies have been explored, in our work, to the same level of detail, but, in this section, we describe all those that we have identified as being applicable.

4.1. International Image Interoperability Framework

The International Image Interoperability Framework, commonly known as IIIF, is a family of specifications that were conceived to facilitate systematic reuse of image resources in digital image repositories maintained by cultural heritage organizations. It specifies several HTTP based web services [9] covering access to images, the presentation and structure of complex digital objects, composed of one or more images, and searching within their content.

IIIF strength resides in the presentation possibilities it provides for end-users. From the perspective of data acquisition, however, none of the IIIF APIs was specifically designed to support metadata aggregation. Nevertheless, within the output given by the IIIF APIs, there may exist enough information to allow HTTP robots to crawl IIIF endpoints and harvest the links to the digital resources and associated metadata.

To study the feasibility of data acquisition via IIIF, several experiments and case studies have been undertaken, and are currently in progress. The early experiments revealed that IIIF contains all the necessary elements for automatic harvesting of metadata. Some of these elements are, however, not of mandatory implementation, thus they will not be available in many IIIF endpoints. The following elements of IIIF APIs must be provided by data providers, to enable Europeana to harvest:

- Structured metadata: the typical metadata available in the output of IIIF is intended for end-user presentation, thus it is unable to fulfil the requirements of ingestion in Europeana. This limitation may however be overcome by using the optional links (i.e. seeAlso) to structured metadata, as specified in IIIF. These enable crawlers to harvest metadata in any format provided, such as EDM, Dublin Core, etc.
- IIIF Collection indicating the resources for Europeana: In IIIF, it is not required that the endpoint implements a mechanism to make publicly known all the digital objects that it makes available. However, such mechanism may be implemented, and, optionally, the IIIF provider may implement a IIIF Collection that lists the digital objects it holds, or just those intended for delivery to Europeana. By making this collection known to Europeana, all the digital objects referenced in the collection can be crawled, and their metadata harvested by Europeana.

There is one piece of information that IIIF does not provide, which is the modification timestamp of the digital objects. This aspect has an impact in the efficiency of the harvesting process, but only becomes relevant in very large collections, with sizes in the hundreds of thousands of digital objects. In the typical size of the collections delivered to Europeana, within the thousands or tens of thousands, the loss in efficiency is not significant nowadays, due to high availability of bandwidth and computational capacity.

To overcome this issue of harvesting efficiency in large collections, other technologies may be used in conjunction with IIIF. Examples are Sitemaps, HTTP Headers, and notification protocols, such as Webmention and Linked Data Notifications, which we are also being evaluated in our work and are described in this document. This issue of harvesting efficiency has been brought to the attention of the IIIF community, and we are engaged in the discussions for achieving a standard mechanism, or recommendations, which will address it within the IIIF community.

The results so far indicate that data acquisition via IIIF is feasible, and presents little technological barriers for data providers that already have an IIIF solution in place for their own purposes. In the Europeana side, once a IIIF crawler tool is integrated with its aggregation management system, ingestion of IIIF data sources can be carried out under the same process of nowadays.

4.2. Webmention

Webmention is a technology that addresses the general problem of allowing Web authors to obtain notifications when other authors link to one of their documents [10].

Webmention is currently published at W3C as a First Public Working Draft. We could not accurately determine how widely adopted Webmention is nowadays, but many resources can be found in the World Wide Web, from software implementations, running services, and many discussions on its use.

The notification mechanisms provided by Webmention, can be used to mediate the communication between the systems of aggregators and the data providers. Webmention presents the following positive aspects:

- A very simple technological solution;
- Any of the parties may initiate the exchange of information.

There are, however, some negative points regarding Webmention:

- No deployments of Webmention are known to exist in CH institutions;
- The notifications do not allow data to be transmitted, so it must be complemented with other technology, such as the example of linked data, which is described further ahead in this section;
- The notifications may lack semantic meaning (e.g. type of notifications) required for some aggregation operations;
- The application of Webmention, for metadata aggregation, diverges somewhat from what Webmention was designed for. If Europeana uses it for this purpose, further elaboration of specifications will be necessary to define how Webmention is meant to be used.

Due to the lack of a mechanism to transmit data in Webmention notifications, we see its application only in combination with other technologies. For example, in combination with existing linked open data (LOD) that data providers already have in place. Webmention would allow data providers to indicate to Europeana, which resources from their LOD dataset should be aggregated by Europeana.

Webmention could also be applied in a similar way to aggregate metadata from IIIF endpoints. The underlying approach may be the same as for LOD. But in this case, the notifications sent by the data providers to Europeana, would contain links to IIIF resources (manifests), and Europeana would use a IIIF crawler to harvest the metadata from the IIIF endpoint.

4.3. Linked Data Notifications

Linked Data Notifications [11] (LDN) is similar in functionality to Webmention, but it is built having the Web of Data in mind, while Webmention is focused in the Web of Documents. LDN is being designed on top of the W3C's Linked Data Platform (see below), and its notifications have richer semantics than the simple notifications of Webmention. Another promising aspect of LDN is that the notifications may carry data, thus allowing for a more straightforward way of fulfilling metadata aggregation than Webmention. We engaged with the LDN editorial group, and are currently providing feedback to the LDN specifications, considering the metadata aggregation use case.

4.4. Sitemaps

Sitemaps [12] allow webmasters to inform search engines about pages on their sites that are available for crawling by search engine's robots. A Sitemap is an XML file that lists URLs of the pages within a website along with additional metadata about each URL (i.e., when it was last updated, how often it usually changes, and how important it is, relative to other URLs within the same site) so that search engines can more efficiently crawl the site. Sitemaps is a widely-adopted technology, supported by all major search engines. Many content management systems support Sitemaps out-of-the-box, and Sitemaps are simple enough to be manually built by webmasters when necessary.

Considering the application of Sitemaps in the context of Europeana, for data acquisition, it presents the following positive points:

- A simple technology with low barriers for implementation, even for small organizations.
- Already in use in several cultural heritage organizations, where it is applied for search engine optimization of their websites and digital libraries.
- It is extensible; thus, it can be adapted to Europeana specific requirements. For example, Google has Sitemap extensions for images and for videos, each one defining a set of metadata elements for its media type.

A Sitemap is an XML file, which is prepared per the Sitemap Protocol [12]. In digital libraries, Sitemaps typically contain all the links to the landing pages of the digital objects within the digital library.

These kinds of Sitemaps are widely used, thus already existing Sitemaps could be used by Europeana for metadata aggregation, using a WebCrawler such as those used by Internet search engines. Starting by following the links in a Sitemap, and processing structured data within HTML (e.g. microdata, Schema.org, linked data available by content negotiation), an Europeana Crawler may discover the digital cultural heritage objects, as well as metadata.

Besides its typical use for Internet crawlers, Sitemaps may also be deployed by Europeana and data providers in conjunction with other technologies, which would allow for simple ways of sharing data. For example, Sitemaps could be made available by data providers, in order to inform Europeana of the digital objects to be aggregated and when they are updated.

Sitemaps, present two clear benefits: a very low technological barrier, and data providing organizations often have in-house knowledge about XML and/or Sitemaps. Sitemaps are a key technology applied for Internet search engine optimization, thus it is already in use within data providers' websites and digital libraries for making their resources discoverable in Internet search engines. Providing metadata to Europeana by using Sitemaps would substantially reduce the implementation effort needed by data providers.

4.5. ResourceSync

ResourceSync [5] is a NISO standard that enables third-party systems to remain synchronized with a data provider's evolving digital objects, supporting both metadata and content. ResourceSync is based on the Sitemaps protocol and introduces extensions that enable its functionality for accurately and efficiently synchronizing the content of digital objects. Additionally, to Sitemap's capabilities, it allows data sources to:

- specify groups of resources, instead of each one individually.

- specify alternative ways to download the resources, as for example, as a bundle in a zip file.
- specify what has changed at a time.
- specify alternative ways to download just a set of changes
- link resources to metadata that describes the resources
- link to older versions of resources
- specifying alternative download mechanisms, such as alternative mirrors.

This detailed synchronization information provided by ResourceSync allows for much more efficient ways of keeping resources synchronized between a source and a destination.

The extra functionality of ResourceSync over Sitemaps, also increases the technical barriers for its adoption. At the time of writing of this document, we have not yet been able to locate a case of ResourceSync deployment in the cultural heritage domain. Most applications of ResourceSync are in grey literature repositories, which are usually out of scope of cultural heritage.

Since the current focus of Europeana is in acquisition of metadata, ResourceSync may offer more than is necessary, and be an unnecessary challenge for implementation by data providers. Still, ResourceSync is an important technology to follow, particularly as the aggregation of content as well as metadata is starting to gain more attention within the Europeana Network.

4.6. Open Publication Distribution System

Open Publication Distribution System (OPDS) is a syndication format for digital publications which enables the aggregation, distribution, and discovery of books, journals, and other digital content by any user, from any source, in any digital format, on any device. The OPDS Catalogs specification [13] is based on the Atom syndication format and prioritizes simplicity. OPDS usage can be found in eBook reading systems, publishers, and distributors. Publishers and libraries have been early adopters of OPDS. We could not yet determine how widely used OPDS is within the Europeana network.

4.7. Linked Data Platform

Linked Data Platform [14] specifies the use of HTTP and RDF techniques for accessing and manipulating resources exposed as Linked Data [6]. Several cultural heritage institutions publish as linked data the metadata regarding their resources. Although these data sources can be accessed and processed for aggregation, they are not available in a uniform and standard way. This requires a lot of manual effort for aggregators to processing the data, presenting a serious obstacle to an efficient and sustainable aggregation process. Within the many aspects specified by the Linked Data Platform, some provide the necessary standardization for an efficient aggregation based on linked data sources.

5. Conclusion

In conclusion, several technological solutions from the Web are available and look promising for simplifying the implementation of the metadata aggregation scenario in cultural heritage. The next steps of this work will aim to assess the actual usage and existing knowledge of these technologies, within the cultural heritage institutions. Future work, on the technical software side, will address how these technologies may be used for designing crawling robots that aggregate the metadata. We expect that with crawling algorithms, which make use of Web technologies, the technical barriers and operational costs may be lowered, leading to more sustainable metadata aggregation networks.

Acknowledgments

This work was partially supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, and by the European Commission under the Connecting Europe Facility, telecommunications sector, grant agreement number CEF-TC-2015-1-01.

References

[1] Lagoze C, van de Sompel H, Nelson ML, and Warner S. The Open Archives Initiative Protocol for Metadata Harvesting, Version 2.0. 2002. Available from: http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm
[2] van de Sompel H, Nelson ML. Reminiscing About 15 Years of Interoperability Efforts. D-Lib Magazine. vol. 21, n. 11/12. 2015. doi:10.1045/november2015-vandesompel
[3] van Veen T, Oldroyd B. Search and Retrieval in The European Library: A New Approach. D-Lib Magazine, vol. 10, n. 2. 2004. ISSN 1082-9873.
[4] Pedrosa G, Georg P, Concordia C, Aloia N. Europeana OAI-PMH Infrastructure. Project Europeana Connect deliverable D5.3.1. 2010.
[5] National Information Standards Organization. ResourceSync Framework Specification. 2014. Available from: http://www.niso.org/apps/group_public/download.php/12904/z39-99-2014_resourcesync.pdf
[6] Berners-Lee T. Linked Data Design Issues. W3C-Internal Document. 2006 Available from: http://www.w3.org/DesignIssues/LinkedData.html
[7] Richardson L, Ruby S. Restful Web Services. O'Reilly. 2007.
[8] Europeana v1.0, The EDM Definition V5.2.7. Online from: http://pro.europeana.eu/web/guest/edm-documentation
[9] Stuart S, Sanderson R and Cramer T. The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images. Archiving 2015, 2015. Available from: http://purl.stanford.edu/df650pk4327
[10] Parecki A (ed.). Webmention. W3C Candidate Recommendation, 2016. Available from: https://www.w3.org/TR/webmention/
[11] Capadisli S, Guy A (eds.), “Linked Data Notifications”, W3C Working Draft, 2016. Available from: https://www.w3.org/TR/ldn/
[12] Sitemaps XML format. Available from https://www.sitemaps.org/protocol.html
[13] The openpub community. OPDS Catalog 1.1 specification. (2011). Available online: http://opds-spec.org/specs/opds-catalog-1-1
[14] Speicher S, Arwe J, Malhotra A. Linked Data Platform 1.0. W3C Recommendation, 2015. Available from: https://www.w3.org/TR/ldp/

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-245

Governing Knowledge Commons: Applications of an Open Knowledge Broker in Caribbean Disaster Management

Maurice MCNAUGHTON[a,1] and Lila RAO[a]
[a] *Mona School of Business and Management, The University of the West Indies*

Abstract. This paper is based on research conducted as an initiative under the Open and Collaborative Science in Development Network (OCSDNet) to explore new innovative mechanisms that can enhance collaborative disaster recovery planning, knowledge management, and learning in the Caribbean. The need for enhanced knowledge management to mitigate disaster risk through the sharing of information and knowledge is a strategic imperative of the Caribbean Disaster Management community. We employ a preliminary conceptual application of the Knowledge Commons/IAD Framework to illustrate how this kind of institutional analytic process can illuminate and inform strategy, governance and desirable collective action, as well as the merits of alternative enabling technologies. The study contributes to arguments challenging the neutrality of infrastructure for collective action. It highlights the importance, and perhaps imperative, of an institutional approach to the design and implementation of socio-technical systems.

Keywords. knowledge commons, disaster recovery planning, governance, controlled vocabularies, Caribbean

1. Introduction

The Open Science paradigm is broadly characterized by a multitude of emergent trends and influences that continue to disrupt and transform conventional notions about the process, mechanisms and roles involved in knowledge creation and dissemination. Factors ranging from the increased participation of the non-scientific public, democratic parity in access to knowledge resources and greater efficiency in knowledge production through open models of collaboration, manifest the changing nature of scientific practice, and is underpinned by the pervasive influence of ICTs [1]. One of the key technological artifacts that is increasingly prevalent in the Open Science paradigm is the emergence of the Knowledge Commons.

Commons, as a general term, refers to *a resource* shared by a group of people, whereas knowledge, as a specific type of resource, refers to a broad set of intellectual and cultural artifacts. Ostrom [2] deliberately emphasizes the essence of "Commons" as being not simply an inanimate pooled-collection of artifacts, but rather "*a commons is a shared resource that is vulnerable to social dilemmas"*. This notion is captured by Frischman et al [3] in their definition of the knowledge commons as *"the*

[1] Corresponding Author, Maurice McNaughton, Mona School of Business and Management, The University of the West Indies, Mona, Kgn 7, Jamaica; E-mail: maurice.mcnaughton@uwimona.edu.jm.

institutionalized community governance of the sharing and, in some cases, creation, of information, science, knowledge, data, and other types of intellectual and cultural resources" thus emphasizing the governance aspect and institutional character of the commons.

This paper is based on research conducted as an initiative under the Open and Collaborative Science in Development Network (OCSDNet) [2] to explore new innovative mechanisms that can enhance collaborative disaster recovery planning, knowledge management, and learning in the Caribbean. A related paper by Davis at al [4] describes the technical architecture of an online knowledge broker, that was developed, based on Linked Open Data Standards (RDF/XML, SKOS, SPARQL, etc.), as a structured, hierarchical thesaurus of relevant concepts relating to Caribbean Disaster Management. The knowledge broker provides a technical solution for the integration of silos of knowledge as it relates to disaster management/recovery, by providing a common semantic reference for knowledge resources and artifacts distributed throughout the region, thus allowing for the emergence of a distributed *knowledge commons.*

Much of the emergent thinking about the characteristics of the knowledge commons, it's utility and requisite governance practices is drawn from references to the corresponding theories about natural resource systems governed as commons such as fisheries, grazing pastures, forests, and irrigation systems, etc. Ellinor Ostrom and various collaborators establish much of the intellectual pedigree in which this work is rooted [5], [6]. Notwithstanding the now acknowledged conceptual flaws, Hardin's seminal "Tragedy of the Commons" (1968) is one of the important anchors in this intellectual tradition. With many of these contributions based on "thought experiments" and conceptual frameworks, there is still a call for empirical studies of real-world commons that help to illuminate when and how knowledge commons governance work—and when it doesn't [2], [3]. As Hess & Ostrom succinctly notes, *"[] study after study demonstrates, there is no one solution to all commons dilemmas"*. This paper thus responds to the call for more empirical research on emerging knowledge commons and studies that help to validate the application of analytic frameworks such as the knowledge commons research framework [3] in different contexts.

The primary objective of this paper is to explore and evaluate the effects that the introduction of the electronic knowledge broker [4], can have on the efficacy and coordination efficiency of the knowledge commons in the Caribbean Disaster Management community. The specific context within which this research is situated provides an interesting domain for examining the characteristics, governance and patterns of interactions within a knowledge commons. Given the common Caribbean vulnerability to, and experience with natural disasters, there is a strong regional commitment to collaboration around comprehensive disaster management and the sharing of knowledge resources, artifacts and coordination.

However, as the research highlights, even within a domain where the actors have strong intersecting interests, minimal competitive incentives, and are generally amenable to open and collaborative solutions to common problems, there are persistent barriers that constrain the effectiveness of knowledge commons. For example, currently in Disaster Recovery Planning, a number of institutions/entities in the region are developing documents and databases related to disaster management/recovery. While there seems to be an active willingness to share these resources, the primary challenge

[2] http://ocsdnet.org/

with the efficacy of this de facto *"knowledge commons*" is standardization, coordinated production and knowing what knowledge resources exists ("*How do we know what we know?*"). There is no central knowledge authority or directory that someone can go to find out what resources are available and thus they continue to exist in silos with limited sharing. Perhaps, references to the "*tragedy of the anticommons*" which contemplates the circumstances under which resources are inefficiently underutilized rather than over utilized, as in the more familiar commons setting, may offer some insights in this regard.

The remainder of this paper is structured as follows: *Section two* provides a brief overview of some of the salient literature on Knowledge Commons, and mechanisms for evaluating their attributes and governance characteristics. We examine, in particular, Ostrom's Institutional Analysis and Development Framework (IAD) and derivative applications [3] [8] [9] [10], to various types of commons and related communities; *Section three* applies the IAD to a specific empirical context, the Caribbean Disaster Management community, and presents a conceptual comparative institutional analysis of the relative merits of two distinct technology approaches to building an effective knowledge management system. In *Section 4*, we highlight the merits of this type of institutional analysis and the insights and implications it suggests for approaches to the planned implementation of a knowledge management system for the regional disaster management community. Some concluding observations and considerations for continued research are outlined.

2. Literature Review

2.1. Knowledge Commons

The notion of the *Commons* as an economic institution in its own right was elevated after Garret Hardin's seminal "The Tragedy of the Commons", published in 1968 [7]. Although Hardin's "Tragedy" predicts a failure of collective action, absent one of the traditional governance mechanisms of private property or public goods, the subsequent work of commons scholars in the natural resource arena led by Ellinor Ostrom and various collaborators suggest otherwise, and have established much of the intellectual pedigree in which this academic stream is rooted [5], [6]. These studies lead to the identification of a general-purpose set of design principles (See [5], 90-102), if commons as a governance institution is to effectively address the essential core issues of *equity*, *efficiency* and *sustainability*.

Fascination with this discourse has been amplified in recent times with the widespread adoption of ICTs and the pervasive effects of the Internet in particular, leading to the emergence of the *Knowledge Commons* as a specific variant of the Commons, manifesting idiosyncratic characteristics and a more complex set of social dilemmas [3], [9]–[11]. Frischman et al [3] offer a definition of the knowledge commons as "*the institutionalized community governance of the sharing and, in some cases, creation, of information, science, knowledge, data, and other types of intellectual and cultural resources*". They emphasize the idea that while this applies collectively to resources, a group or community of people, the *Commons* does not denote the elements themselves, but rather the institutional arrangement of these elements. This is a particularly important nuance for the knowledge commons in which typically, participants not only share existing resources but also engage in defining and producing those resources, making it impractical, if not impossible, to separate the attributes of the managed resources from the attributes of the community that produces and uses them [3]. In the case of Caribbean Disaster Management, the community itself, is as valuable as the knowledge artifacts, by way of expertise, empathy and responsiveness in the aftermath of a Disaster.

Several other distinct idiosyncrasies of the knowledge commons, as compared with natural commons, are highlighted by Frischman et al [3] as follows:

a) While natural resource commons typically seek to allocate consumption and preserve resources, knowledge commons must address the coordination requirements across a range of production, dissemination and consumption activities. Hence social dilemmas in these '*constructed commons'* often exist moreso in incentivizing the contribution and sharing of knowledge rather than the risk of congestion or overconsumption.
b) Given the nonrivalrous and nonexcludable character of knowledge resources, knowledge commons must confront questions of openness and the dynamics of sharing both within and outside the defined commons community
c) Nonrivalry of knowledge and information resources may be associated with rivalrous inputs and outputs such as time, money and reputation, leading to potentially contentious issues in acquiring and distribution these resources among members of the commons community.

These nuances become apparent and relevant to our subsequent empirical Case analysis.

2.2. Evaluating Knowledge Repositories

According to [3], these distinctive attributes of the knowledge commons require significant engagement with the specific narrative of the community, given the basic assumption that each actor begins from a position of rational self-interest. We therefore turn to several studies of electronic information repositories such as knowledge management (KM) systems and document management (DM) systems that primarily address the behaviors of participants as the unit-of-analysis, and their motivations/incentives to collaborate, contribute and share knowledge artifacts, and provide important insights to the social dilemmas confronting knowledge commons [12]–[14].

Kankanhalli et al. [12] point to the failure of KM initiatives due to the reluctance of employees to share knowledge through knowledge management systems (of which their particular focus is on electronic knowledge repositories (EKR)). They use both the social exchange theory and the social capital theory to formulate and test a theoretical model to explain EKR usage by knowledge contributors. The model was validated through a large scale survey of public sector organizations and it was found that self-efficacy and enjoyment in helping others significantly impact the EKR usage by knowledge contributors.

Bock and Kim [13] seek to develop an understanding of the factors affecting the individual's knowledge sharing behavior in an organisation context. They recognize the importance of knowledge sharing and the growing interest in this area by many senior management. They draw on a number of theories including the social exchange theory, self-efficacy, and theory of reasoned actions. They surveyed employees of large public organisations and found that expected associations and contribution are the major determinant of an individual's attitude toward knowledge sharing. Interestingly they found that expected rewards was not significant to the attitude toward knowledge sharing.

Chiu et al. [14] study focuses on virtual communities and the willingness of members to share knowledge with other members. It is imperative that for these virtual communities to have any value the richness of the knowledge is key, and that this knowledge must be member generated. Thus, it is essential that studies must be conducted to explain why individuals elect to share or not to share knowledge with other community members when they have a choice. They also used the Social Capital Theory but with the Social Cognitive Theory to develop and test a model for investigating the motivations behind people's knowledge sharing in virtual communities. The study found that community-related outcome expectations play an important role underlying knowledge sharing in terms of both quantity and quality, while personal outcome expectations have a negative but insignificant effect on quantity of knowledge sharing.

The insights derived from these studies inform the design of a baseline survey of the Caribbean Disaster Management community and their knowledge attitudes and practices as described in Section 3.2.

2.3. Analyzing Knowledge Commons – IAD Framework

The Institutional Analysis and Development (IAD) framework was developed by Elinor Ostrom and collaborators to provide a general-purpose comparative method of institutional analysis for studying commons arrangements in the natural environment [5] According to Hess and Ostrom [2], the framework is well suited for analysis of resources where new technologies are developing at a rapid pace, such as the new ICTs that have redefined knowledge communities. It is not only applicable to descriptive analysis, but "... *is an appropriate place to start when trying to think through the challenges of creating new form of commons such as a new digital repository within an organization."*

Frischman et al [3] have taken account of the distinct idiosyncrasies of the knowledge commons in their adaptation of the IAD to the Knowledge Commons Framework (see fig 1.), which reflects the more complex relationships among knowledge resources, community participants, and governance structures. This is evident, for instance, in the case of Caribbean Disaster Management, where knowledge artifacts such as Disaster Recovery Plans are a function of both standards set by the Governing body as well as the unique local circumstances of the countries participating in the CDM community. [3] describes the application of the framework to the examination and analysis of specific empirical contexts using a systematic cluster of inquiries. Schweik [9] illustrates the application of the approach in the comparative analysis of Open Source Commons institutions, while Morell [10] applies the approach in examining online communities engaged in the creation of digital commons. [10] considers the enabling ICT infrastructure as part of the governance mechanism, rather than as an aspect of the resource characteristics of the commons, hence challenging the neutrality of infrastructure for collective action. This idea is particularly interesting for this study, where, in *Section 3.4.,* we employ the IAD systematic approach to comparative institutional analysis, while examining the relative merits of alternative technology solution approaches to the knowledge management challenges of the Caribbean Disaster Management community.

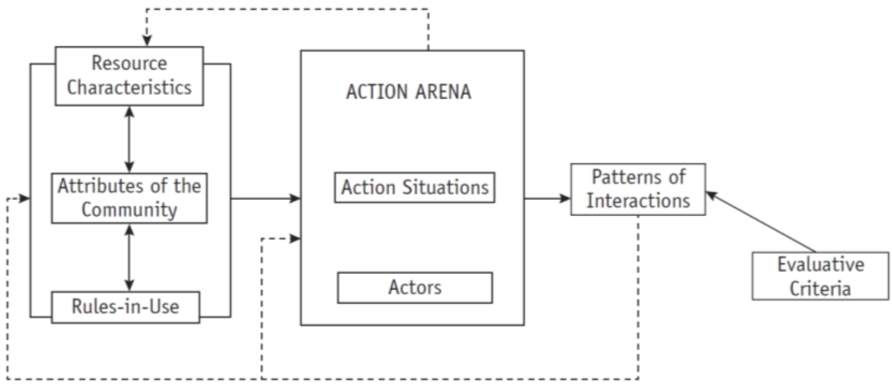


Figure 1. The Knowledge Commons (IAD) Framework. *Source*: Frischman et al [3]

3. Case – Caribbean Disaster Management

3.1. Institutional Context

In the Caribbean, the Disaster Management community is well organized, with the Caribbean Disaster Emergency Management Agency (CDEMA) designated as a regional inter-governmental agency for disaster management in the Caribbean Community (CARICOM). CDEMA's mandate is to fully take up its role as facilitator, driver, coordinator and motivating force for the promotion and engineering of Comprehensive Disaster Management (CDM) in all Participating States. CDEMA is supported and actively engages with a network of national disaster management agencies. For example, in Jamaica the Office of Disaster Preparedness and Emergency Management (ODPEM) is the main body responsible for coordinating the management of the various types of disasters while in St. Vincent and the Grenadines it is the National Emergency Management Office (NEMO) who is assigned the role of activating the community on a countrywide basis to deal with disasters. CDEMA is governed through a regional structure[3] consisting of the Council, a Technical Advisory Committee (TAC) and the Coordinating Unit:

- The CDEMA Council is the highest level of governance of CDEMA and determines policies and major decisions on its operations.
- The Technical Advisory Committee (TAC) is the technical and programmatic advisory arm of the CDEMA and comprises of the National Disaster Coordinators and representatives of specialized regional organizations, whose programmes are directly related to the regional disaster management agenda
- The Coordinating Unit is managed by an Executive Director who is appointed by the Council, and has responsibility for the management and administrative functions of the organization, including research, education and ICT.

Given its mandate CDEMA represents a key knowledge actor within the CDM Community and knowledge management (KM) is seen as one of the strategic priorities for the community, as reflected in Outcome 2 of the Comprehensive Disaster

[3] http://www.cdema.org/index.php?option=com_content&view=article&id=47&Itemid=60

Management (CDEMA, 2014), which speaks to the need for enhanced knowledge management to understand disaster risk through the sharing of information and knowledge. Two significant institutional initiatives within CDEMA's 5-year strategic program in keeping with these priorities are:

a) "Enhancing Knowledge & Application of Comprehensive Disaster Management (EKACDM)", a research project that has as one of its key outcomes "*The creation of a regional network which generates, manages, and disseminates knowledge on disaster management*".
b) CDEMA is currently in the process of implementing the Caribbean Risk Information System (CRIS)[4] as a multifaceted virtual platform that will host an electronic repository of risk management data and information in response to the perceived gaps in the existing de facto knowledge commons in the CDM Community.

3.2. Baselining the CDM Knowledge Ecosystem

While the Caribbean Disaster Management Community has a comprehensive and mature institutional framework in terms of governance mechanisms and reporting relationships throughout the Caribbean, there are deficiencies in the management and sharing of knowledge resources across the Community as acknowledged by the following excerpt from CDEMA's 2014 – 2024 Strategic Plan, page 14 (highlights added for emphasis):

> In a broader context, information about disaster preparedness in case of an emergency, disaster management plans, policies and guidelines have been in existence and accessible for many years. However, communities have been severely affected by disasters due to lack of adequate coping capacity. This may be attributed to **limited access to resources to address risk exposure**. **In order to enhance the information sharing and management of the knowledge generated from various sources, it is highly essential to closely network the organizations/institutions and moreover people working at the community level to increase resilience**. The network of institutions will create a common platform and **enable its stakeholders and people to capture, organize, share and reuse the knowledge** generated in the area of disaster management.

To unpack this issue further, a baseline survey was designed to assess the state of attitudes and practices of members within the Disaster Management community to share knowledge resources with other members of this community. This is an important step of the process in determining the effect the implementation of the Electronic Knowledge Repositories will have on the willingness of the members the Caribbean Disaster Management community to share and contribute to the knowledge resources of the community.

Given this objective, a number of the constructs from existing literature [12]–[14] were synthesized into a survey instrument with 16 constructs. Kankanhalli et al. [12] was important as the focus on their study, like this one, was on Knowledge Management Systems. Chiu et al. [14] studied virtual communities and as the

[4] http://www.cdema.org/index.php?option=com_content&view=article&id=1577&Itemid=576

Caribbean Disaster Management Community is dispersed throughout the region at different levels it was thought that many of the constructs would be relevant. Finally, Bock and Kim [13] include a construct for IT usage and this was deemed to be extremely important given the diversity of members in the Caribbean Disaster Management Community. This questionnaire (*available on request*) will be administered to the members of the Disaster Management Community throughout the Caribbean region with the support of the Caribbean Disaster and Emergency Management Agency (CDEMA).

3.3. Conceptual Application of the Knowledge Commons (IAD) Framework

The appropriately termed Institutional Analysis and Development Framework finds application in both *analysis* of existing types of commons, as well as in commons *design* for future desired states. Hess/Ostrom [8] suggest, for instance, that the "Action Arena" (see Fig 1) is the *"appropriate place to start when trying to think through the challenges of creating a new form of commons such as a new digital repository within an organization"*.

3.3.1. CRIS – Caribbean Risk Information System

In response to the need for enhanced knowledge management as a strategic priority, the Caribbean Risk Information System (CRIS) is conceived as a multifaceted virtual platform that hosts risk management data and information accessible to stakeholders to facilitate analysis, research, greater awareness of risk management and climate change adaptation in the region. Prior regional efforts at similar ICT initiatives have been unsuccessful due to a number of perceived challenges:

- Absence of data and information sharing protocols among agencies
- Lack of standardization (minimum) of datasets leading to incompatibility of databases to facilitate research, etc.
- Irregular or un-sustained hosting capacities
- Poor access to, or limited understanding of use of, data and information

CRIS is conceptualized to have the following functional components:

- Provides access to DRM and CCA information developed internally by CDEMA and externally by regional and international development partners;
- Maintains a repository of key disaster risk management documents for each CDEMA Participating States including policies, plans, procedures;
- Provides access to geospatial data with specific focus of demonstrating how risk information can be incorporated into developing hazard specific maps and aid development-oriented planning;
- Maintains project reports to assist in research and information sharing;

A key required technology component of CRIS, therefore, is a *document management system (DM)* to provide the electronic repository for policies, plans, procedures, reports, standards and other knowledge artifacts. A related initiative to explore new innovative ICT mechanisms [4] led to the development of an alternative technical solution referred to as a *knowledge broker (KB)*. The knowledge broker provides for the integration of silos of knowledge within the disaster management domain, by providing a common semantic reference for knowledge resources and artifacts distributed throughout the region, thus allowing for the emergence of a distributed *knowledge commons.*

3.3.2. Comparative Institutional Analysis - Knowledge Broker vs Document Management System

Our conceptual application of the Knowledge Commons/IAD Framework within this interesting institutional domain seeks to illuminate and inform strategy action, governance and desirable patterns of interactions, as well as the merits of alternative enabling technologies, if the Caribbean Disaster Management community is to realize the goal of enhanced knowledge management and to allow a greater focus on the contingent effects of the enabling ICT infrastructure. As such, we interrogate just a subset of the systematic cluster of inquiries articulated by Frischman et al [3] in their adapted knowledge commons (IAD) framework. This is consistent with the observation [11] that "*efforts to apply the IAD framework in the realm of the information or knowledge commons so far have focused on discrete subparts of the problem or applied only subparts of the IAD framework, which seems sensible and perhaps inevitable*". Our focus is therefore on individual actors in "action situations" and the governance mechanisms that will influence their decision-making and actions with respect to desirable outcomes for the CDM knowledge commons. In the interest of space, a brief synopsis of the key aspects of the analysis is presented in Table 1.

Table 1: Analysis of CDM using the IAD Framework

IAD Components	Definition/Description	Questions/Narrative – CDM Knowledge Commons
Actors	Key members of the community participating in the knowledge commons	Key Actors in this knowledge commons setting include: CDEMA, the centralized governing body; the country level agencies responsible for disaster preparedness and emergency management; National disaster coordinators and representatives of specialized regional organizations (including Multilateral agencies and Academia), whose programmes are directly related to the regional disaster management agenda; organizations/institutions and people working at the local community level. CDEMA, the designated inter-governmental agency for disaster management in CARICOM, oversees a comprehensive and mature institutional setting.
Action Situations	Decisions & Actions by Community members and the related incentives/outcomes, assessed at various levels of interaction: *Constitutional, Collective and Operational*	The related strategic Goal is to enhance the information sharing and management of the knowledge generated from various sources, especially the organizations and institutions and people working at the local community level, to increase resilience. A common platform is required that will enable its stakeholders and people to capture, organize, share and reuse relevant knowledge. Achieving this goal will require Actions at several levels, with related incentive mechanisms and governing rules. *Constitutional*: Within its designated authority, the CDEMA Council determines overarching policies and major decisions relating to the community. Mobilizing the knowledge commons requires strong leadership, visioning and advocacy at the highest level *Collective:* To induce the institutional change towards a more effective knowledge commons requires a

		combination of standards, guidelines and incentive mechanisms *Operational:* At this level, a vibrant knowledge commons will manifest active contributions, sharing and use/reuse of knowledge artifacts by the community
Rules-In-Use/Governance Mechanisms	The formal rules/informal practices that govern and direct behavior in the knowledge commons	Knowledge resources within the CDM domain include policies, plans, procedures, reports, standards and other reference artifacts. Many of these resources are produced at local community, national, regional and international levels. The technology choices in terms of a centralized document repository *versus* a distributed knowledge broker that indexes resources wherever they reside, can significantly influence the degree of participation and interaction and ultimately the effectiveness (i.e. of *equity*, *efficiency* and *sustainability*) of the commons. *Openness* with respect to the knowledge resources is not a binary state, rather it is a spectrum that exists on multiple-dimensions, specifically: *Discoverability, Accessibility, Reusability, and Transparency; Discoverability* has been, perhaps, the most evident gap in relation to disaster management knowledge resources in the Caribbean *(How do we know what we know?*). The absence of a central knowledge authority or directory has severely limited the discoverability of knowledge resources. The *knowledge broker* has been demonstrated as an effective mechanism for the integration of DRP knowledge silos currently dispersed throughout the region. *Openness* to community participation in relation to the creation of, rights to, and use of knowledge resources (including access by community *outsiders*). The design of the technology platform is also significant in enabling or precluding varying levels of participation access such as standards setting, decentralized production, self-management of contributions, etc. The ability of participants to determine their level of commitment to participation can broaden the community by encouraging more casual contributors. More formal rules can be established through activities such as document standards (e.g. country/sector-level disaster recovery plans) and well as licensing regimes for knowledge artifacts within the commons.

4. Conclusion and Future Research

This paper contributes a very preliminary and decidedly limited *(deliberately so)* application of the Knowledge Commons (IAD) Framework to examine the institutional context for knowledge management within the Caribbean Disaster Management community. Even within this limited scope, the IAD has demonstrated its utility in evaluating key decision issues, especially as it relates to the implementation of the enabling technology platform. The comparative analysis of the *knowledge broker,* a controlled vocabulary semantic server, as the enabling technology versus more conventional centralized document management systems presents some distinct advantages, especially as it relates to emergent attributes of the knowledge commons, such as openness, modularity and decentralized production, self-management of contributions and infrastructure provisioning.

This case when fully developed is likely to provide additional empirical support for the arguments advanced by [10], challenging the neutrality of infrastructure for collective action. It highlights the importance, and perhaps imperative, of an institutional approach to the design and implementation of socio-technical systems, especially those spanning porous organizational boundaries, versus the typically techno-centric approaches based on the classical Systems Development LifeCycle. The acknowledged failure of prior attempts at implementing similar regional ICT initiatives underscore the need for a different approach that recognizes the importance of institutional design as part of ICT projects to addresses the non-technical social dilemmas that typically confront such initiatives.

The next phase in this research is to administer the survey instrument developed as part of this study to the institutional members of the Caribbean Disaster Management community which includes the National Disaster Coordinators and representatives of specialized regional organizations, whose programs are directly related to the regional disaster management agenda. The researchers have been invited to participate in the next meeting of CDEMA’s Technical Advisory Committee (TAC) and will use the opportunity to discuss some of the institutional design considerations arising from this analysis. Ultimately the goal is to ensure that this research and the ongoing engagement with the Caribbean disaster management community can inform and influence the design and implementation of CRIS as a key enabling platform for an effective and sustainable CDM knowledge commons.

References

[1] B. Fecher and S. Friesike, "Open Science: One Term, Five Schools of Thought," RatSWD Work. Pap., 2013.

[2] C. Hess and E. Ostrom, "Introduction: An Overview of the Knowledge Commons," in Knowledge as a Commons: from Theory to Practice., MIT Press, 2006.

[3] B. M. Frischman, M. J. Madison, and K. J. Strandburg, "Governing Knowledge Commons," in Governing Knowledge Commons, B. M. Frischmann, M. J. Madison, and K. J. Strandburg, Eds. Oxford University Press, 2014.

[4] J. Davis, A. Nicholson, L. Rao, and M. L. McNaughton, "Towards an Open Knowledge Broker for Disaster Recovery Planning in the Caribbean," in Uciencia 2016, 2016.

[5] E. Ostrom, Governing the commons: The Evolution of Institutions for Collective Action. Cambridge university press, 1990.

[6] E. Ostrom and C. Hess, "Understanding Knowledge as a Commons: From Theory to Practice." MIT Press, 2006.

[7] G. Hardin, "The Tragedy of the Commons," Science (80-.)., vol. 162, no. 3859, pp. 1243–1248, 1968.

[8] C. Hess and E. Ostrom, "A Framework for Analyzing the Knowledge Commons: a chapter from Understanding Knowledge as a Commons: from Theory to Practice.," 2005.

[9] C. M. Schweik, "Toward the Comparison of Open Source Commons Institutions," Gov. Knowl. Commons, pp. 255–279, 2014.

[10] M. F. Morell, "Governance of Online Creation Communities for the Building of Digital Commons: Viewed through the Framework of Institutional Analysis and Development," in Governing Knowledge Commons, Oxford University Press, 2014, p. 281.

[11] D. H. Cole, "Learning from Lin: Lessons and Cautions from the Natural Commons for the Knowledge Commons," 2014.

[12] A. Kankanhalli, B. C. Y. Tan, and K.-K. Wei, "Contributing knowledge to electronic knowledge repositories: an empirical investigation," MIS Q., pp. 113–143, 2005.

[13] G.-W. Bock and Y.-G. Kim, "Breaking the myths of rewards: An exploratory study of attitudes about knowledge sharing," Pacis 2001 Proc., p. 78, 2001.

[14] C.-M. Chiu, M.-H. Hsu, and E. T. G. Wang, "Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories," Decis. Support Syst., vol. 42, no. 3, pp. 1872–1888, 2006.

[15] Caribbean Disaster Emergency Management Agency, "Regional Comprehensive Disaster Management Strategy and Programming Framework 2014-2024," 2014.

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-257

What Is at Stake? Public Participation and the Co-Production of Open Scientific Knowledge

Hugo FERPOZZI[1]
University of Buenos Aires; Open and Collaborative Science in Development Network; Centre for Science Technology and Society

Abstract. Openness has become an explicit subject across science policy and scholarly practice, where it is often vindicated in a rhetoric of optimism. In political discourse, as much as in the scholarly literature, open access to research data and publications is expected to enable what policy has typically failed to achieve by other means: that is, to overcome material, class, and political barriers that stand in the way of knowledge circulation. However, whether openness in science is a good thing or not also seems to depend on what is being opened, to what extent and for whom. In this paper I draw on different critical areas of Latin American science, technology and society studies (LASTS) to suggest that the current dominant views around open science can be limiting, as much as they could be enabling, more inclusive dynamics of access to and uses of scientific knowledge, especially in the peripheral (or non-hegemonic) contexts of science. These limiting views around openness, I argue, are linked with restrictive conceptions about science and its products: scientific activity is understood, by this token, as an invariably universal enterprise. In consequence, science outputs are conceived as self-contained knowledge products, and the processes and practices that account for their production and use are only partly taken into consideration. The aim is hence to elaborate on different forms of participation and exclusion to the processes of knowledge production which could help us understand how different stakeholders become engaged or excluded in the production of knowledge. To do so, I take the case of genomic research and drug development for neglected diseases as my empirical background. The argument draws on two concepts from LASTS. The first one is cognitive exploitation, according to which scientific outputs are used in for-profit contexts by third-parties, but without compensating the original producers. In this way, it is not only producers, users and appropriators of knowledge who become key in the dynamics of knowledge circulation, but also those acting as intermediaries. The other concept is integrated subordination, which refers, on the one hand, to the dynamics by which peripheral regions collaborate with elite research networks, and the difficulties that stand in the way of industrializing scientific knowledge, on the other. These difficulties spawn from the lack of capacities, but also from adherence to international research agendas, which are not necessarily connected with those required to attend to social needs in peripheral contexts. By putting into question the nature and the limits of openness, and by re-examining the types of knowledge at stake (beyond research data and publications), the actors, and their involvement, I suggest other ways in which open scientific knowledge could become effectively used.

Keywords. participation, open access, co-production of knowledge, policy, discourse

[1] Hugo Ferpozzi, Centro de Ciencia, Tecnología y Sociedad, Valentín Virasoro 732 3, 1405 Ciudad de Buenos Aires, Argentina; E-mail: hugo.ferpozzi@gmail.com.

1. Introduction

Scientific knowledge has long been understood as *universal*. From the *philosophes* of the Enlightenment to Robert K. Merton's first systematic attempts into the sociology of science [1], at least, scientific knowledge was more often than not idealized as belonging to the final commitment to universal human reason. Scientific universality, as I understand it, was (and still is) intended to impart at least two interrelated norms: scientific knowledge has to transcended personal *ownership*, but it also has to be reducible to *context-free* accounts of the reality.

More controversial depictions of science emerged towards the 1970s when the social sciences and the humanities gave up on the concept of scientific knowledge as something necessarily true and invariably universal. "Post-Mertonian" waves of science studies broke into the sites of knowledge production, then, to gather traces of contingency, arbitrariness, negotiation, chance, belief and secrecy that govern the making of scientific knowledge and order scientific practices. "Laboratory studies", as these were dubbed later on, owed partly to the post-structalists' wider reaction to realism and the autonomous individual that served as an underpinning for the universal (Western) logos, although the starting point for these new waves of science studies can also be traced back to less radical efforts (including Thomas Kuhn's *Structure of scientific revolutions* [2] and David Bloor's *Knowledge and social imagery* [3]), aimed at dragging scientific knowledge into the realm of relativism, precisely, because of its association with human history and the social.

The deeper epistemological implications of making scientific knowledge an object of social inquiry shall not (and could not) be discussed here further, as it remains, yet, a matter of dispute within the field of science, technology and society studies (and mostly everywhere else) [see, for instance, 4–9]. Instead, in this paper I shall elaborate on the overall critique to the two aforementioned norms that are expected to regulate the pretended universality of scientific knowledge, and which still seem to be vindicated in the dominating discourses of open science. My critique can be synthesised as follows: the ownership of scientific knowledge can be realized through its *effective use* and de facto exclusiveness in spite of its formal openness; in addition to this, the possibility of producing and effectively using scientific knowledge is for the most part *context-dependent*, in spite of its intended universality. In putting forward this critique I draw on several proposals stemming chiefly from Latin American science, technology and society studies (LASTS). The overall orientation of this paper is to identify the implications of ownership and the contextual factors in scientific openness, especially in its ability to attend to social needs in peripheral societies and developing regions. To do so, I focus mainly on the case of Chagas disease research and open-access drug development initiatives. The data comes from my doctoral dissertation [10] and from the outputs of the Open and Collaborative Science in Development Network project "Can open and collaborative science meet social needs?". This line of enquiry would be, in the last instance, aimed at elaborating on different forms of participation and exclusion in science that are seldom contemplated in the analyses of open science and its related initiatives. In doing so, I present a set of conceptual tools that can help us understand (and potentially foster) the participation of stakeholders in the face of producing usable scientific knowledge. These tools are derived from the concepts of *cognitive exploitation* and *integrated subordination*, both found in LASTS.

The next section introduces the case study and the idea of *cognitive exploitation* to understand forms of knowledge ownership and exclusion that might subsist precisely because of scientific openness, rather than in spite of it.

2. Neglected Tropical Diseases Research and the Exploitation of Open Scientific Knowledge

Chagas disease or *American Trypanosomiasis* is endemic in the Americas and affects over 15 million individuals. While its forms of transmission are mostly vector-borne and occur in rural areas with deficient housing conditions, migratory processes of the last 40 years have render the disease as a public health concern in urban areas traditionally and non-endemic regions [11–16]. Chagas is also classified by the World Health Organization (WHO) as one the 17 neglected tropical diseases, meaning it prevails in tropical conditions and lacks of effective, affordable, and widely-available treatment options.

In spite of this discouraging scenario, Chagas has been a target of sustained international research efforts, aimed, in part, at making up for the lack of commercial interest shown by pharmaceutical firms [17–20]. Since its launching by the WHO in 1975, support for research and development activities in Chagas and other similar diseases has come, mainly, from the Special Programme for Research and Training in Neglected Tropical Diseases (TDR). Other research centres and funding bodies include the Rockefeller Foundation, The Bill & Melinda Gates Foundation and Doctors Without Borders.

The problem of cognitive exploitation, as well as the potential problems with open access in Chagas disease research, surfaces with the completion of the Trypanosome cruzi Genome project (TcGP), an initiative aimed at mapping the genomic sequence of the Chagas' causing organism. Almost in parallel with the Human Genome Project, the TcGP was devised as a means for fostering the development of medical applications against Chagas disease. All the obtained sequences and the research data from the project is stored in open access, publicly available databases: GenBank is one [21], although other genomics resources have been specifically dedicated to Chagas disease research and oriented towards drug development effort such as TDR Targets and TriTrypDB.

TDR Targets, for instance, operates as a web-accessible open access resource developed "to facilitate the rapid identification and prioritization of molecular targets for drug development, focusing on pathogens responsible for neglected human diseases" [22]. The project was envisaged soon after the completion of the TcGP and owed, partly, to the valuation made by TDR working groups on therapeutic options available for neglected diseases, especially after the completion of the TcGP and other *Trypanosomatidae* genome projects [10]. Through the TDR, the WHO managed to set up and define the initial outlines of TDR Targets, but also expected other stakeholders –representatives from research laboratories and the pharmaceutical industry – to become responsible for the project funding and execution. In its initial planning, however, open access was not a requirement specified by the WHO but a collective proposition made by the participating researchers [23].

2.1. Cognitive Exploitation: Production, Use and Contextualization of Collaborative Scientific Outputs

In previous work, I followed the hypothesis that open access resources such as TDR Targets could be subject to processes of *cognitive exploitation* [10]. The concept of cognitive exploitation refers to the utilization of non-profit knowledge outputs in for-profit operations without providing compensation for the original knowledge producers [24]. In this way, pharmaceutical firms can, at least in principle, take advantage of the knowledge produced from publicly funded research efforts without having to face the costs of the initial (and more uncertain) development stages themselves.

Cognitive exploitation is not a phenomenon exclusive to science or to open access. A typology of the processes involving cognitive exploitation has been proposed by Kreimer and Zukerfeld [24] according to the kinds of knowledge at stake: these include *scientific* but also *indigenous*, *labouring*, and *informational* (*digital*) knowledge. Different types of cognitive exploitation involve different classes of *producers* and *appropriators* of knowledge, as well as different *mediators* and *intermediaries* that operate under certain *regulatory frameworks.* Here, however, I am concerned chiefly with the exploitation of scientific and digital knowledge.

The exploitation of the scientific and the digital kinds of knowledge have been extensively addressed in the scholarly literature, although, logically, not always by means of the term “cognitive exploitation” (and not necessarily through the approach followed here). In what respects to scientific knowledge, for instance, Lefèvre [25] understands science as a form of *universal labour* that is freely appropriated by private producers once it is stable enough to become profitable. Codner, Becerra & Díaz [26], on the other hand, proposed the term *blind technological transfer* to refer to the utilization of publicly funded research publications in patent documents; however, their study only samples the area of biotechnology at the University of Quilmes in Argentina. In a different sense, documents from the International Union for Conservation of Nature [27] have drawn attention to the possible abuses of open data in the field of wildlife conservation, while Fecher & Friesike [28] note the widespread utilization of unpaid workforce within open scientific practice.

A more extensive debate has taken place around the exploitation of digital knowledge. The issues stem mostly from the dynamics of collaborative production and digital economy brought about with more “recent” phenomena such as user-generated content, social networking platforms, and free software licences [29–36].

Zukerfeld [31], for that matter, makes a distinction between different subtypes of exploitation of digital knowledge based on the production of *data*, *software* or *contents*. In any case, central to the idea conveyed with the exploitation of digital knowledge is what the author understands to be the material economy of digital knowledge and its inherent “double freedom”:

> Whereas the usual voices (from management literature to hackers) emphasize one freedom (the shiny side of copying and sharing informational goods), we think we are unwittingly discussing about two very different but inseparable freedoms. Here is where Marx comes back. One of the key factors for the birth of Capitalism has been what Marx called the double freedom of labor power. On the one hand, the worker is freed from the feudal order, free to move and free to sell his labor-power where, when and how he wants to. By the time of Marx, this had been the only freedom mentioned by Political Economy, Contractualism and Liberalism. But, on the other hand the worker is also freed

> from the means of production, as it is well known… Marx highlights the necessity of two contradictory freedoms. In the first case, freedom refers to empowerment; in the second, to the lack of power [p. 146].

The dynamics through which free knowledge is, in these two senses, incorporated into the capitalist machinery is therefore characterized as processes of *inclusive appropriation*: while freely produced digital knowledge is only made profitable by the third-parties, it nonetheless remains non-rival and, in this cases, also non-exclusive [31]. Processes of cognitive exploitation, in sum, necessarily entail asymmetrical exchanges that take place under contingent legal frameworks and exclude physical coercion as a means [24,29]. The next section relates the problem of cognitive exploitation with the problem of drug development for neglected tropical diseases.

2.2. Cognitive Exploitation in the Field of Neglected Tropical Diseases Research

Open access genomic databases represent particularly meaningful resources of scientific knowledge that may be problematised under the concept of cognitive exploitation, especially in what respects to drug development for neglected diseases. Three main points can support this claim:

- First, genomic databases are both *products and means of producing* scientific and digital knowledge, the latter of which encompasses the three subtypes defined above as data, software and contents. While the last decade has made out of openness a more explicit subject across science policy and scholarly practice, the focus, I argued, is usually put on the scientific outputs represented by research data and publications (contents). This means, in other words, that the debates on its constraints and opportunities have remained tied to the problem of access to outputs [28,37–42]. For example, without recurring to the idea of cognitive exploitation, the publishing industry has been long criticized (now especially in the digital era) for hindering the dissemination of scientific knowledge, indirectly favouring economically privileged actors, and making profits out of scientific production at the expense of knowledge sharing (43). More recently, however, private publishers have began to align with open access-based business models as long as it will not endanger traditional editorial privileges and profits [44–47]. This does not necessarily hold true for drug development-related databases.
- Second, the potential intermediaries and (or) appropriators of the knowledge produced from genomic databases oriented to drug development goals are primarily pharmaceutical firms. In spite of this, the motivations sustaining neglected tropical diseases research are also knowledge-driven and supported via government and NGO funding (as is the case with TriTrypDB and TDR Targets). The ability, interests, and frameworks available to enforce intellectual property protection, on the one hand, and the capacities to industrialize scientific knowledge, on the other, are markedly asymmetrical.
- Third, and related to the last point, neglected tropical diseases prevail, by definition, in tropical and subtropical contexts, meaning that their incidence is significantly higher in the contexts of development. This not only reinforces its asymmetrical position in relation to the leading centres of technological development and scientific research, but also pose very different interests in

> terms of knowledge production and in defining what is at stake. Neglected tropical diseases constitute a subject for scientific research but also a social and political issue in both endemic and non-endemic contexts. By circumscribing what the issue in the last instance *is*, then, global health organization and research centres can pre-define certain solutions as possible, and therefore restrict the types of knowledge and stakeholders than may (or can) become involved.

In spite of appearing as a suitable target for cognitive exploitation, our research [10] suggests that no commercial utilization of knowledge produced from genomic databases for neglected tropical diseases is actually taking place. In these cases, however, this might be occurring for all the "wrong" reasons, as there is no locally produced open scientific knowledge being industrialized, be it by means of asymmetrical exploitation or not. In other words, unrestricted access to the research outputs (as represented in the double freedom of open digital knowledge) can be a necessary condition for utilizing scientific knowledge, but certainly not a sufficient one. As a matter of fact, in the field of drug development for neglected tropical diseases, the costs and uncertainty of initial research efforts –not compensated by the relatively low purchase power of the affected populations – did not appear as the main factor hampering the development of new drugs.

The two points discussed above can also synthesise what is conveyed in the idea of *integrated subordination*. This idea refers to the dynamics by which peripheral regions succeed in integrating to the mainstream research networks, but fail to industrialize scientific knowledge and connect scientific research with local social needs. As it has been proposed in LASTS [48], this failure spawns from insufficient technological capacities, but also from the adherence to international research interests and agendas which are not necessarily connected with local needs and demands in spite of its rhetoric of "social relevance".

3. Open Science in (Asymmetrical) Contexts: Which Knowledge Outputs and For Whose Needs?

One of the limitations in conceptualizing open science, I argued, is its engagement with the products of science rather than with its processes. Fecher & Friesike [28] offer an initial approach to this limitation by making a distinction between the *democratic* and the *public school* of open science: while the former is concerned with *access* to knowledge, the second one is concerned with processes and *accessibility*.

The idea of accessibility involves at least two different aspects. One of them is communication and exchanges between lay and scientific actors, an aspect which has often been conceptualized as a problem of "conveying" scientific knowledge "to" lay audiences. The other aspect has to do with what has been often portrayed as *citizen* science, and refers to the participation of non-experts in certain processes of data collection or analysis. These views, I argue, pose at least two problems.

The first one has to do with the possession of skills and capacities that are required to utilize scientific knowledge. Arza & Fressoli [37] refer to accessibility in this same sense as "the lack of [other] more informal restrictions, such as the specific skills, capacities, and capital resource required to understand or utilize the products of open science" [p. 3]. I would add that it is not just the possession *cognitive* or *material* resources what defines actual accessibility to scientific knowledge, but also –and even more importantly– *political* and *symbolic* resources.

The second problem has to do with the stage of knowledge production at which participation is enabled in science. While citizen science contemplates non-expert participation, the core stages of knowledge production –such as defining the problem and its approach, the research priorities and type of knowledge outputs expected – remains, for the most part, exclusive to expert circles. For sure, the problem of public participation in science has long been analysed in science technology and society studies, moving beyond limited notions of citizen science described above. Diverse approaches and models have been proposed to understand and facilitate the engagement of non-expert and scientists in co-producing scientific knowledge. These could not, of course, be discussed here in detail, as they normally do not touch the specific issues that stem from open access [7,49,50].

Nonetheless, in the field of neglected diseases, the precise implications of open access have been critically addressed by Masum and Harris in an institutional document titled "Open source for neglected diseases: Magic bullet or mirage?" [51]. The authors review a series of initiatives linked with drug development to analyse, along with the dimension of access, the dimensions of *collaboration* and *governance*. In the field of neglected diseases, they note, drug development is complicated further due to the recent expansion of legal and market regulations [p. 1]. For example, while 1394 new drugs had been commercialized between 1975 and 1999, only 16 where destined for neglected diseases, and even those few "new" drugs put on the market have been proved to be deficient, or simply modified copies of pre-existing drugs [17].

In the face of commercial disadvantage, then, open access could theoretically contribute to overall productivity in drug development, mainly by means of facilitating decentralised operations and data sharing [28,52,53]. However, there are significant differences between the dynamics of pharmaceutical business and other domains where open access and practices occur more "naturally", such as in the realm of software development. Masum and Harris, for instance, describe how the two business dynamics are very different in terms of regulation, risks and costs, or even in terms of safety and time requirements. On the other hand, while software firms can rely on *copyright protection*, in principle, without major difficulties, biomedical and pharmaceutical firms, instead, depend on extensive clinical trials and costly *patent filing* processes on the road to putting a new product on the market. An open access approach, they argue at last, works well with *discovery* or *pre-competitive* stages of biomedical research, but has been rarely been known to succeed during the phases of *technology transfer* and *delivery*.

From the obstacles standing in the way of drug development, it is clear that the signifier "open science" may convey fundamentally dissimilar meanings across the various techno-scientific spaces engaged in drug development. As the authors illustrate this polysemy:

> What is the "source code" at each stage of neglected-disease research? While some working in synthetic biology make the analogy of DNA as source code, the situation is actually more complex. In software, the source code *is* the product, while in biology, there are many relevant levels of description and analysis, from DNA to structural genomics, protein interactions, metabolism, and so forth—all interacting in complex ways and requiring a long and expensive process to go from description to approved product [52, p.3].

This contextual approach suggests that the difficulties that hold back the utilization of locally produced, open scientific knowledge do not just illustrate "mismatchings" between the dominating discourses on open science, on the one hand, and its realization through effective scientific practices, on the other. It also suggests that these notions, in their limited and universalising conceptualization of access, have omitted the contextual and dynamic factors affecting the production and social use of scientific knowledge. These contextual factors do necessarily have an effect on scientific openness and its ability to meet social needs, and depend on possessing the required skills and capacities to face changing constraints and regulations, as well as with the possibilities of engaging stakeholders in the different stages of research. In other words, there are different dynamics of openness and accessibility that exceed the question of accessing (or opening up) scientific data and publications.

In the case of TDR Targets, the tensions that result from the asymmetrical position between the producers, the potential appropriators, and the intended end-users of scientific knowledge in the specific context of drug development for neglected tropical diseases have been noted by researchers, firm directors and representatives from health-related NGOs, although the potential solutions they might pose are different. As a laboratory director with more than ten years of experience in developing genomic databases for neglected tropical diseases put it:

> I never worked for a private [pharmaceutical] firm. But from my meetings, and from all the experience I had from different people in different levels, I believe in the first place that there's no market. So even if firms are interested in us doing all the work and then taking advantage of the results, as you say, the day they get those results they are going to realize that they are developing a product that maybe didn't cost them a lot, but that they have to sell to demographics with no income ... Governments have to buy it for them, and every government is different, just as everything else. I don't know if it's a great business or not ... It's different, for example, if someone has diabetes, which needs to be treated permanently ... a person suffering from Chagas, instead, you treat them, they're cured, and it's over, they won't be taking that drug ever again … In addition to that, firms... invest and need to get the money back in ten, twenty years, which is what patents last for. ... And even then, I hear this isn't their motivation; the motivation is another kind of intangible benefit, which has to do with public image. ... One of the things being blamed [for the lack of new drugs] is excessive regulation... controls, safety issues... there are many and every drug has a problem. So if you have a life or death situation, say, if you're dying from cancer, they'll get approval.... But for other things they may face many problems [23].

Here, the role of *mediators* and *intermediaries* – necessary to enable a successful industrialization of scientific knowledge – is, again, clearly not limited to the possession of technological or cognitive resources alone. The WHO appears in this case as the mediator: it is the *obligatory passage point* – recurring to Callon's terminology [54] – through which pharmaceutical firms, research centres, government offices and funding bodies become, at least in principle, enrolled in collaboration. Pharmaceutical firms, instead, are the *intermediaries* capable (again, in principle) of introducing the *translations* necessary for industrializing knowledge. Translation, as it has been shown, entails a few things more than just technical capacities: it entails practical know-how and expertise on how market and health regulations, public expectations, and intellectual property protection might or might not work for a certain disease, product, population, government, and so on.

3.1. Opening Up Participation: Defining What Is at Stake

At this point it might be useful, at last, to rethink participation as part of a more comprehensive idea of openness.

In its role as a mediator, the WHO, through the TDR, has favoured a biomedical approach to the problem of neglected tropical diseases –and arguably more so in the case of Chagas. A working paper issued by the TDR in 1979 depicted the issue of Chagas disease as problem that was rooted in the lack of biomedical advancements. Conversely, the possibility of intervening and improving the living and environmental conditions of the affected population conditions were, in this same paper, seen as uncertain, so the necessary path to attending to the problem of Chagas disease had to be pursued through a knowledge-driven approach from the biomedical sciences [55].

The same type of justification was put forward again 15 years later with the launching of the TcGP in the 1990s [56]. As a matter of fact – and in spite of its ongoing rhetoric of social relevance –, it is until today that the research dynamics around Chagas disease adhere to the rule of universal mainstream science: political intervention remains outside the scope of "legitimate" scientific involvement, and the affected populations –discursively regarded as the "beneficiaries" of research efforts – rarely become engaged in relevant spaces of scientific and political decision-making [10,57,58].

In this sense, even if openness does enable unrestricted access to the research data and publications in the field of neglected tropical diseases research, it is hard to see, in the face of the aforementioned limitations and exclusions to the overall scientific process – including the very definition of and the approach to the problem itself –, how the dominant dynamics of openness could help in fostering scientific outcomes that effectively connect with local social needs and demands.

4. Conclusions

In this paper I suggested that the dominant views on open science could be limiting the local utilization of scientific knowledge in non-hegemonic contexts of science. These limiting views around openness, I argued, are linked with restrictive conceptions about science and its products, which fail to account for the practices and processes involved in their production and use. Instead, I discussed the case of genomic research and drug development for neglected tropical diseases, trying to show how different classes of resources and stakeholders – actual, ideal or potential – become engaged or excluded. To do so, I drew on two concepts put forward by Latin American Studies of science, technology and society. The first one was cognitive exploitation, a concept that allows to detect how and which scientific outputs are used in for-profit contexts by third-parties, but without objectively compensating the original producers. The second concept revolved around the idea of integrated subordination, according to which peripheral regions may successfully integrate international research networks but fail, at the same time, to industrialize the scientific knowledge that would attend to local social needs. In the case of drug development for neglected tropical diseases no processes of actual exploitation of scientific knowledge has been found to take place.

However, the concept of cognitive exploitation allowed to detect how regulatory frameworks, actors and relations shaped (and hindered) the dynamics of knowledge production. On the other hand, the issues underlying the inability to meet social needs did not stem here from mere access restrictions to material or cognitive resources. Instead, it was political, legal, and symbolic resources (typically possessed by mediators and intermediaries) which played a more crucial role and posed strong contextual asymmetries between international and local stakeholders, on the one hand, and between experts and potential beneficiaries of knowledge, on the other.

These underlying issues suggest, in sum, that capacities and participation need to be put in context and conceived as an inseparable aspect of access when discussing open science and its ability to meet social needs in non-hegemonic spaces.

References

[1] Merton RK. La ciencia y la estructura social democrática. In: Teoría y estructura sociales. México: Fondo de Cultura; 1992.

[2] Kuhn TS. The Structure of Scientific Revolutions. Chicago: University of Chicago Press; 1962.

[3] Bloor D. Knowledge and social imagery. Routledge; 1976.

[4] Latour B. Why Has Critique Run Out of Steam? From Matters of Fact to Matters of Concern. Crit Inq. 2004;30(2):225–48.

[5] Latour B. Reensamblar lo Social: una introducción a la teoría del actor-red. Buenos Aires: Manantial; 2008.

[6] Jasanoff S. Ordering knowledge, ordering society. In: Jasanoff S, editor. States of Knowledge: The Co-Production of Science and the Social Order. New York: Routledge; 2004. p. 13–45.

[7] Collins HM, Evans R. The Third Wave of Science Studies: Studies of Expertise and Experience. Soc Stud Sci [Internet]. 2002 Apr 1 [cited 2014 Mar 4];32(2):235–96. Available from: http://sss.sagepub.com/cgi/doi/10.1177/0306312702032002003

[8] Callon M. Four Models for the Dynamics of Science. In: Jasanoff S, Markle G, Petersen J, Pinch T, editors. Handbook of Science and Technology Studies. Thousand Oaks, California: Sage; 1994. p. 29–63.

[9] Rheinberger H-J. A Reply to David Bloor: “Toward a Sociology of Epistemic Things.” Perspect Sci [Internet]. 2005 Sep [cited 2014 Feb 21];13(3):406–10. Available from: http://www.mitpressjournals.org/doi/abs/10.1162/106361405774287973

[10] Ferpozzi H. Coproducción de conocimientos y de problemáticas científicas: Dinámicas en la producción de conocimiento científico en la investigación genómica y biomédica sobre la enfermedad de Chagas (1993-2016) [doctoral dissertation]. (Buenos Aires: AR) University of Buenos Aires; 2017.

[11] Schmunis GA. Epidemiology of Chagas disease in non-endemic countries: the role of international migration. Memórias do Inst Oswaldo Cruz [Internet]. 2007 Oct 30;102 Suppl:75–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17891282

[12] Coura JR. The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions - A comprehensive review. Mem Inst Oswaldo Cruz [Internet]. 2015;110(3):277–82. Available from: memorias.ioc.fiocruz.br

[13] Coura JR. Chagas disease: what is known and what is needed – A background article. Mem Inst Oswaldo Cruz. 2007;112(1):113–22.

[14] Hotez PJ, Dumonteil E, Betancourt Cravioto M, Bottazzi ME, Tapia-Conyer R, Meymandi S, et al. An Unfolding Tragedy of Chagas Disease in North America. PLoS Negl Trop Dis [Internet]. 2013 Oct 31;7(10):e2300. Available from: http://dx.plos.org/10.1371/journal.pntd.0002300

[15] Manne-Goehler J, Umeh CA, Montgomery SP, Wirtz VJ. Estimating the Burden of Chagas Disease in the United States. Dumonteil E, editor. PLoS Negl Trop Dis [Internet]. 2016 Nov 7;10(11):e0005033. Available from: http://dx.plos.org/10.1371/journal.pntd.0005033

[16] Requena-Méndez A, Aldasoro E, de Lazzari E, Sicuri E, Brown M, Moore DAJ, et al. Prevalence of Chagas Disease in Latin-American Migrants Living in Europe: A Systematic Review and Meta-analysis. Rodrigues MM, editor. PLoS Negl Trop Dis [Internet]. 2015 Feb 13;9(2):e0003540. Available from: http://dx.plos.org/10.1371/journal.pntd.0003540

[17] Yamey G. The world’s most neglected diseases : Ignored by the pharmaceutical industry and by public-private partnerships. Br Med J [Internet]. 2002;325(7357):176–7. Available from:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1123710/

[18] Perez-Molina JA, Martinez Perez A, Norman FF, Monge-Maillo B, Lopez-Velez R. Old and new challenges in Chagas disease. LANCET Infect Dis. THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, OXON, ENGLAND: ELSEVIER SCI LTD; 2015 Nov;15(11):1347–56.

[19] WHO. Research priorities for Chagas disease, human African trypanosomiasis and leishmaniasis. [Internet]. World Health Organization technical report series. 2012. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23484340

[20] WHO. Neglected tropical diseases [Internet]. 2016. Available from: http://www.who.int/neglected_diseases/diseases/en/

[21] Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res [Internet]. 2013 Jan 1;41(D1):D36–42. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks1195

[22] Magarinos MP, Carmona SJ, Crowther GJ, Ralph SA, Roos DS, Shanmugam D, et al. TDR Targets: a chemogenomics resource for neglected diseases. Nucleic Acids Res [Internet]. 2012 Jan 1;40(D1):D1118–27. Available from: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkr1053

[23] Anonymous source. Genomics and bioinformatics laboratory director. Personal Communication. 18 Jun 2016.

[24] Kreimer P, Zukerfeld M. La Explotación Cognitiva: Tensiones emergentes en la producción y uso social de conocimientos científicos, tradicionales, informacionales y laborales. In: Kreimer P, Vessuri H, Velho L, Arellano A, editors. Perspectivas latinoamericanas en el estudio social de la ciencia, la tecnología y el conocimiento. Mexico DF: Siglo XXI; 2014. p. 178–93.

[25] Lefèvre W. Science as Labor. Perspect Sci [Internet]. 2005 Jun [cited 2014 Feb 16];13(2):194–225. Available from: http://www.mitpressjournals.org/doi/abs/10.1162/106361405774270539

[26] Codner DG, Becerra P, Díaz A. Blind Technology Transfer or Technological Knowledge Leakage: a Case Study from the South. J Technol Manag Innov [Internet]. scielocl; 2012;7:184–95. Available from: http://www.scielo.cl/pdf/jotmi/v7n2/art15.pdf

[27] IUCN. Rules of Procedure IUCN Red List Assessment Process 2013–2016 [Internet]. 2012 [cited 2016 Aug 3]. Available from: http://jr.iucnredlist.org/documents/Rules_of_Procedure_for_Red_List_2013-2016.pdf

[28] Fecher B, Friesike S. Opening Science: One Term, Five Schools of Thought [Internet]. Bartling S, Friesike S, editors. Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly. Cham: Springer International Publishing; 2014 [cited 2014 Jul 30]. 17-48 p. Available from: http://link.springer.com/10.1007/978-3-319-00026-8

[29] Dolcemascolo A, Yansen G. Informational cognitive exploitation: Concealed relationships behind prosumers' activity on World Wide Web. Sociol Lav. 2017 Mar;(145):61–77.

[30] Petersen SM. Loser Generated Content: From Participation to Exploitation. First Monday [Internet]. 2008 Mar 3 [cited 2014 Mar 31];13(3). Available from: http://journals.uic.edu/ojs/index.php/fm/article/view/2141

[31] Zukerfeld M. Inclusive Appropriation and the Double Freedom of Knowledge: on the Capitalist exploitation of non-for profit software, contents and data producers. Sociol Lav. 2014 Feb;(133):145–58.

[32] Lessig L. Code 2.0 [Internet]. Nueva York: Basic Books; 2006. Available from: http://codev2.cc/download+remix/Lessig-Codev2.pdf

[33] Hardt M, Negri A. Postmodernization, or The Informatization of Production. In: Empire. Cambridge, Massachusetts: Harvard University Press; 2001. p. 280–303.

[34] Fuchs C, Sandoval M. Digital Workers of the World Unite! A Framework for Critically Theorising and Analysing Digital Labour. tripleC. 2014;2(12):486–563.

[35] Vuorinen J. Ethical codes in the digital world: comparisons of the proprietary, the open/free and the cracker system. Ethics Inf Technol [Internet]. 2007 Feb 28;9(1):27–38. Available from: http://link.springer.com/10.1007/s10676-006-9130-2

[36] Rullani E. Le capitalisme cognitif : du déjà vu ? Multitudes [Internet]. 2000 [cited 2014 Apr 2];2(2):87–94. Available from: http://www.cairn.info/revue-multitudes-2000-2-page-87.htm

[37] Arza V, Fressoli M. Benefits and obstacles of openness in science: an analytical framework illustrated with case study evidence from Argentina [Internet]. Buenos Aires; 2016. Available from: https://stepsamericalatina.org/wp-content/uploads/sites/21/2016/08/Arza-Fressoli_24082016.pdf

[38] The Netherlands EU Presidency. Amsterdam Call for Action on Open Science [Internet]. Amsterdam; 2016. Available from: https://english.eu2016.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science

[39] BOAI. Budapest Open Access Initiative [Internet]. 2002 [cited 2017 Jan 23]. Available from: http://www.budapestopenaccessinitiative.org/read

[40] OANA. The Vienna Principles: A Vision for Scholarly Communication in the 21st Century [Internet]. Working Group “Open Access and Scholarly Communication” of the Open Access Network Austria (OANA). 2016. Available from: http://viennaprinciples.org/v1/
[41] Bartling S, Friesike S. Towards Another Scientific Revolution. In: Bartling S, Friesike S, editors. Opening Science The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly. New York: Springer Open; 2014. p. 3–15.
[42] Nentwich M. Web 2.0 and academia. In: Proceedings of the 9th Annual IAS-STS Conference “Critical Issues in Science and Technology Studies” [Internet]. Graz, Austria; 2010 [cited 2014 Jun 9]. p. 66–78. Available from: http://www.researchgate.net/publication/49579798_Web_2.0_and_academia/file/9fcfd505755e3366b0.pdf
[43] Campanario JM. The peer review system: many problems and few solutions. Rev española Doc Científica. 2006 Dec 27;25(3):267–85.
[44] Nature. Editorial. 2016.
[45] Anonymous source. Former authority at the University of Liège. Personal communication. 10 dec 2016.
[46] Perrone I. Internet y las publicaciones científicas. In: Cafassi E, editor. Internet: políticas y comunicación. Buenos Aires: Biblos; 1998.
[47] Rodriguez J. Edición 2.0. Los futuros del libro. Barcelona: Melusina; 2007. 255 p.
[48] Kreimer P. ¿Dependientes o integrados? La ciencia latinoamericana y la nueva división del trabajo. Nómadas. 2006;(24):199–212.
[49] Lemos MC, Morehouse BJ. The co-production of science and policy in integrated climate assessments. Glob Environ Chang [Internet]. 2005 Apr;15(1):57–68. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0959378004000652
[50] Chilvers J, Longhurst N. Participation in Transition(s): Reconceiving Public Engagements in Energy Transitions as Co-Produced, Emergent and Diverse. J Environ Policy Plan [Internet]. 2016;7200(February):1–23. Available from: http://www.tandfonline.com/doi/full/10.1080/1523908X.2015.1110483
[51] Masum H, Harris R. Open Source for Neglected Diseases: Magic Bullet or Mirage? Washington, DC; 2011.
[52] Oduola A, Wayling S, Morel C. Genomics and bioinformatics for tropical diseases. TDRnews [Internet]. Geneva; 2002;(68):6–7. Available from: http://www.who.int/tdr/publications/documents/tdrnews-issue-68.pdf
[53] Lenoir T. Science and the Academy of the 21st Century: Does Their Past Have a Future in an Age Computer-Mediated Networks? In: Voß kamp W, editor. Ideale Akademie: Vergangene Zukunft oder konkrete Utopie? Berlin: Akademie Verlag; 2000. p. 113–29.
[54] Callon M. Algunos elementos para una sociología de la traducción: la domesticación de las vieiras y los pescadores de la bahía de St. Brieuc. In: Iranzo JM, Blanco JR, González de la Fe T, Cotillo A, editors. Sociología de la ciencia y la tecnologia. Madrid: CSIC; 1995. p. 259–82.
[55] Zabala JP. La enfermedad de Chagas en la Argentina. Investigación científica, problemas sociales y políticas sanitarias. Bernal: Universidad Nacional de Quilmes; 2010.
[56] Trypanosoma cruzi Genome Consortium. The Trypanosoma cruzi genome initiative. Parasitol Today [Internet]. 1997 Jan;13(1):16–22. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0169475896100740
[57] Kreimer P. Co-producing Social Problems and Scientific Knowledge: Chagas Disease and the Dynamics of Research Fields in Latin America. In: Merz M, Sormani P, editors. Sociology of Science Yearbook. Dordrecht: Springer; 2015. p. 173–90.
[58] Kreimer P, Zabala JP. Chagas Disease in Argentina: Reciprocal Construction of Social and Scientific Problems. Sci Technol Soc [Internet]. 2007 Mar 1 [cited 2013 Sep 20];12(1):49–72. Available from: http://sts.sagepub.com/cgi/doi/10.1177/097172180601200104

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-269

The University Cheikh Anta Diop of Dakar (UCAD) Science Shop "Xam-Xamu Niep Ngir Niep" (Knowledge of All for All)

Dieyi DIOUF[1]
School of Librarians, Archivists and Documentalists (EBAD), University Cheikh Anta Diop of Dakar (UCAD)

Abstract. Traditionally, universities in the North as well as in the Global South concentrated their activities on two main missions: Teaching and Research. A "third mission" of universities called "service to the community", defined as its social responsibility to contribute to development, is now promoted to researchers [1] [2] [3]. Several studies have shown that scientific and local knowledge play an important role in the process of sustainable development by creating an operational interface between researchers, students and non-profit organizations [4] [5] [6]. In order to fully accomplish this mission for the benefit of local communities, researchers are getting involved in Science shops, which were established in the Netherlands in the 1970's. Glen Millot [15] speaks of "third sectors" in reference to the role Science Shop plays. Indeed, Science shops are dynamic mediators of cooperation between communities, NGOs, citizens and researchers. Science Shos teams receive demands from civil society or organizations and helps translate them into research programs or scientific issues that students and researchers treat and make the results available to communities. This presentation will firstly focus on a definition of some useful concepts. Then, the second part will deal with the origin of Science Shops and their evolution before analyzing the process of setting up the UCAD Science Shop "Xam-xamu niep ngir niep" (Knowledge of all for all).

Keywords. Science shop, openess of research, participatory research, social responsibility, civil society, local knowledge, UCAD

1. Introduction

A Science Shop is a permanent device, usually integrated into the structure of a university. It allows the university to get closer to the people, working together with the civil society organizations, scientists and students. Science shops and institutional repositories are the preferred tools of Open Science as they allow collaboration, sustained exchanges between civil society and academics, but also a better sharing of knowledge [7].

Traditionally, universities had two very specific missions: teaching and research. But, Researchers in Information and Communication Sciences evoke now a "third mission" of universities called "Service to the community", which can be defined as a university's social responsibility to contribute to Development [8].

[1] Corresponding Author, E-mail: dieyi.diouf@ucad.edu.sn/dieyidiouf@yahoo.fr

This third mission is part of the process of building a knowledge society in which knowledge flows freely in the service of the common good [9] [10] [11] [12].

To ensure the “scientificity” of knowledge at university, the dominant normative framework of conventional science requires researchers to refrain from interference by non-scientists in the definition and development of research programs and projects. The university is supposed to deal with local development issues such as the lack of hygiene, corruption, misappropriation of public funds, politicization of the administration which undermines access to public services [13]. In fact, African universities usually ignore peasant or civil society organizations that are at the heart of the debate on local development.

Indeed, opening up research to civil society can make a significant contribution to sustainable development by creating an operational interface between researchers, students and non-profit organizations [4] [5] [6]. To illustrate these ideas in their study of 21 science shops, Leydesdorf & Ward [11] showed that cooperation within these shops can promotes inclusive development.

On the other hand, the dialogue between universities in the South and the North could also have the positive effect of reviving reflection on modes of knowledge [14].

In this presentation, defining some concepts seems useful first of all. Then, we will give a brief overview of the origin of science shops and their evolution before analyzing the process of setting up the UCAD Science Shop.

2. Definition of Concepts

2.1. Xam-Xamu Niep Ngir Niep

The UCAD Science Shop is named *“Xam-xamu niep,”* which literally means “knowledge of all”. It is then associated with the expression *“ngir niep”* which means *“ for all”,* which we call the spirit of association and sharing. In the Wolof language, *Xam-xamu niep ngir niep* translates to the linking between the university and society, and the UCAD Science Shop proposes to formalize and perpetuate these relationships which the university must maintain with society.

2.2. Open Science

In the context of SOHA project, Open Science is defined as a science in the service of local populations of a territory. It is a science that takes into account all types of knowledge to serve society as a whole without discrimination. In other words, instead of ignoring or marginalizing non-scientific knowledge (traditional, local, political, daily, etc.), Open Science is a science that opens up to them. According to this definition, Open science can be considered as a common good which belongs to humanity [7].

2.3. Science Shops

Science Shops are structures that offer citizen groups inexpensive access to scientific and technological knowledge and research so that they can improve their social and environmental conditions. The main types of projects carried out by the Science Shops are bibliographic studies, scientific research, technological developments or adaptations of existing technologies to meet specific needs [15].

2.4. Scientific Third Sector

The Scientific Third Sector is a term that covers a wide variety of actors and processes related to a participatory approach to research. Some forms of organization of these actors around concrete projects have developed and have acquired institutional recognition, at least on a European or international scale. The Science Shops are an example of success at the European level but unfortunately, after a first ephemeral experience in the 1980s, it was difficult to re-emerge in France [15] and elsewhere.

3. History of Science Shops

Created in the 1970's in the Netherlands by students wishing to help a group of citizens worried about the quality of the water of a lake on which they were bordering, Science Shops have spread to more than a hundred locations around the world, but mostly in Europe [9] [10] [11] [12].

Since the early 2000s, this openness to society has been defined as a major element of the social responsibility of universities [1] [2] [3]. As a form of Responsible Research Innovation (RRI), the Science Shop entails engaging all actors (from individual researchers and innovators to institutions and governments) through inclusive, participatory methodologies in all stages of R&I processes and in all levels of R&I governance (from agenda setting, to design, implementation, and evaluation). Science Shops thus emerged to provide access to knowledge or respond to the research needs of organizations that do not have the means to carry out such research, and are also lacking the connection with the world of research considered too focused on academic and technological innovation [9] [10] [11] [12].

With the exception of South Africa, the continent of Africa was largely left out of Science Shop development until November 2016. Conscious of the importance of these tools and considering their high absence in the continent, nine Science Shops were created in conjunction with the SOHA projects in December 2016 (one in Haiti and eight in sub-Saharan francophone Africa: Benin, Burkina Faso, Cameroon, Guinea, Mali, Niger, DRC, Senegal).

Currently, 3 Sciences Shops are set up in Benin (Alowotodji), Haiti (SPOT) and Senegal (Xam-Xamu niep ngir niep). This was one of the goals of the SOHA project, to convey the message that social responsibility of African universities is crucially needed to connect them to local populations for a local development agenda constructed by citizens.

Below are examples of various calls we made to Senegalese academics and civil society in the Saly workshop on March 6th 2017 to begin the UCAD Science Shop outreach activity of awareness raising:

> You are looking for an internship associating your training with a societal problem, the UCAD Science Shop offers you the opportunity to complete your courses by a field research, concrete and useful to society. This practice will be the opportunity to put your knowledge to the benefit of a structure. (See adhesions on the Science shop Web site) online: http://nianitambabn.com/wordpress/

Are you a part of an association, a foundation, an NGO or any other structure involved in the themes Health, Environment or Social Economy? The UCAD Science Shop accompanies you by connecting you with teachers, researchers and students.

4. Context and Justification

4.1. Brief History of UCAD

The University of Dakar was founded on the 14th of February, 1957 and inaugurated in the December of 1959. The University of Dakar is the oldest and most important structure of higher education currently in Senegal. Since the 30^{th} of March 1987, it was baptize Cheikh Anta Diop University of Dakar, following the death of the great historian bearing the same name. Apart from the central administrative departments of the President's Office, it involves thirty higher-level research institutions, comprising six faculties, six higher education institutes, nineteen university institutes, and one Inter-State school of Science and Veterinary Medicine, which is scientifically dependent on the University.

4.2. UCAD Vision

UCAD's vision is to be a successful university serving the economic, social and cultural development of Senegal and Africa, while remaining rooted in the value systems that underpin the Senegalese Nation. It proclaims its openness to solidarity and complementarity with the rest of the world.

4.3. UCAD Missions

In order to highlight its values in the form of objectives, UCAD has set itself the following fundamental missions (some of them related to the science shop):

- To train highly skilled, scientifically and technically skilled personnel, tailored to the African context and the contemporary world, aware of their responsibilities to their peoples and capable of serving them with dedication;
- Develop research in all disciplines of science, technology and culture;
- Mobilize all intellectual resources for the economic and cultural development of Senegal and Africa and participate in the solution of national and continental problems;
- To acquire the most advanced knowledge and methods of investigation in all disciplines of science, technology and culture and to involve them in the development of knowledge and the creation of new methods of investigation, adapting to national and more generally African realities and requirements;
- Working with practitioners to promote traditional knowledge, the circulation of knowledge and information, and the coordination of initiatives to contribute to scientific progress and labor productivity;
- To develop, criticize and disseminate new knowledge, to constitute a place of interaction and cooperation between the world of work and economic, technical, administrative and scientific decision-making centers;

- Study and develop the paths of an endogenous and self-centered development strategy, including participation in the design, implementation and evaluation of national, subregional and regional development plans, etc.

This internationally renowned university has very important students (100,000 students), about 1,500 teachers-Researchers and 1,300 administrative and technical personals [16].

These statistics show a plethora of students with a low degree of training. We have noted an important pedagogical reform resulting from the National Consultation on Teaching with two big changes: the introduction of the LMD[2] reform in 2003 and the reform of Doctoral studies in 2005.

All these researchers produce many research results annually that are ignored by civil society but surely could help to develop one aspect of their needs. For these reasons, we think that Science Shop could play a bridging role between the university and the communities.

5. Motivation for the Creation of a Science Shop in UCAD

5.1. Poor Community Access to University Expertise

At UCAD, with the exception of some doctors who are working with medicinal plants, there is not enough dialog and collaboration between the local population and researchers who remain in laboratories and base their teaching on European realities. Then members of civil society, generally illiterate, are not implicated in scientific activities. This is reflected on the science shop logo: Academics are on the tree (using scientific knowledge) and the civil society, population in general is under the high tree representing the gap between them and academics (they are looking at them and they seem very far from the civil society who is dreaming to reach them one day and ask for help to put an end to the numerous problems they are facing on a daily basis).

5.2. Low Awareness of Local Knowledge in Teaching

As every faculty works alone in his laboratory, each one ignores what the other does. Teachers do not pay enough attention to local knowledge in their teaching programs. So the lack of platforms for coordination and sharing is a handicap. The civil societies are also not interested by what is being done in universities and so they ignore each other and think that no one can do something meaningful for the others.

5.3. Scientific Partition of Academics

There is not enough openness of academics to the civil society. So, they don't include other forms of knowledge in their teachings and always use the French language, French realities and practices. They know what the populations need because they come from and belong to the society, but their needs are not the priority. And the population also thinks that it is not possible to communicate their needs to academics. Another aspect that happens very often is that, the population would not divulgate

[2] LMD refers to licence-maitrise-doctorat (bachelor-masters-doctorate)

some local knowledge. They keep it jealously and die without communicating even with the family. There are many considerations that prevent them from sharing indigenous knowledge and these are still not well understood.

5.4. Low Policy Priority of Administration

The slow advancement and difficulties we are facing in this Science Shop project is a good illustration of the low policy priority of the university administration. As a result, researcher are managing projects themselves or looking for partners outside the university without any support. However, if we succeed in a project, everyone begins to be interested and wants to be associated.

5.5. Low Capacity of Actors

At UCAD, some teacher-researchers are still very distant from ICTs. A few of them avoid publishing online, using social networks, or sharing their publications online. For these reasons, we are organize an Open Access day each year. We managed to give them USBs, pins, tee-shirts marked with"Open Access", etc. So, while they have a very low capacity in new methods and practices, the situation is improving because many texts promoting open access are supported by CAMES, the National Assembly of Universities themselves.

Based on the obstacles above, we believe that it is necessary and even essential to interconnect and to bring the Senegalese academic world (who has competence in research and innovation) and other societal actors who need concrete solutions or inspiring subjects or topics to develop one sector of life.

It is in this context that the creation of a science and knowledge shop in a university like UCAD is meaningful. Thus, the Science Shop will put the skills and knowledge available at UCAD to the service of projects and the needs of the civil society organizations.

By highlighting a pedagogy based on action and collaboration "outside the walls", by enhancing the students' ability to carry out concrete projects at the service of the community, this educational system is based on a brilliant idea: it invites students to carry out, free of charge, as part of their training and/or curriculum, research projects or practical projects in response to needs expressed by the State, the associations, the civil society or by the university.

The term "shop" is therefore misleading. Indeed, civil society organizations and students mobilize other forms of knowledge in service of the project being carried out. These are not volunteers or para-university projects that students would choose to do on their own in addition to their courses. The projects carried out are integrated into the training and pedagogy and form part of the university credits or the units of value of a program.

6. Specific Objects of the Science Shops

6.1. Connecting Universities and Society Through Projects with a Territorial Social Impact

The team of the Science Shop will play an important role of mediation between teachers-researchers and the civil society. If a member of an NGO or an enterprise or civil society has a project, he can post it on our platform. And the Science Shop team will contact for example, a student who was looking for a project or a teacher wanting to study this aspect with a student. So the science shop is a relay for each actor.

6.2. Help The State, Civil Society Organizations and Businesses Find a Solution to Their Different Concerns in a Solidarity Through the Intervention of Motivated Students and Teachers-Researchers

The same interaction will be done for all of these actors. Very often, communities can't pay an expert to develop projects for them. For example, some civil society administrations ask the university to send them students for some projects because they can't recruit someone. So the students only help for 2 or 4 weeks and these students were given one note (credit) as if they were in the classroom. This community-based work is training students for their professional work in the future.

6.3. Encourage Students to Find Answers to Territorial Concerns Enabling Them Both to Implement the Knowledge Acquired in Their Courses and to Acquire and Experiment Open Science Through Scientific Work

This object will train the students to become great citizens and they will improve their training in practice. This kind of exercise will make them much more familiar with their qualifications. They will learn by doing. This is the best way to understand and get more experience beyond theory.

6.4. Co-Build All Together (University and Civil Society) a Sustainable Local Development

If researchers remain in their laboratories and try to find answers for the communities, they will achieve very weak and slow local development. To succeed a better sustainable development, researchers and all the others actors of society must work together in an inclusive way. Development can't be constructed by one part of the population but by all actors.

7. Targets/Beneficiaries

7.1. Target - Students (Doctorate, Master), Teachers, Researchers, Academic Authorities

Every year, students are seeking someone to guide them in their work (research or Dissertation). With the low number of teacher-researchers and lake of accompanying policy, the academic community actors are the first targets of this project. So, if teachers, researchers and academic authorities have a need but can't find motivated students, the Science Shop platform is best place to look first and the team will find solutions together.

7.2. Main Target - Senegalese State, NGOs, Civil Society Organizations and Actors, Enterprises, Local Populations

The Senegalese state has many development projects but they are not associated with the populations for which theses projects are destined. So, experts are generally called from other countries. They are highly paid yet they ignore completely the realities of the local communities.

7.3. Secondary Target: Municipalities, Ministries, Students of African and Foreign Universities, Local Populations, the Diaspora

Municipalities, Ministries and other local administrations usually have very low budgets to pay an expert to lead a project of development for them. So students and researchers will help to do it and this will be a great experience of collaboration between all actors.

8. Activities to Carry out

8.1. Awareness Campaigns and Organization of Seminars, Colloquia, Training Workshops (Authorities, Research Communities, Students, Companies, Society)

The team and pioneers we have in each faculty are going to work widely to sensitize first in order to get the adhesion of more academics, the communities, students and enterprises.

Then, if funded, we will start with short conferences and conduct workshops for faculty after faculty, followed by the high schools of UCAD, and finally the institutes. If they accept the project as done by academics during the Saly workshop, we can then introduce the organization of seminars, colloquia, and training workshops.

8.2. Campaigns to Popularize the Tool

The campaigns can be done during workshops at the same time or during campaigns of sensitization. We will have some materials such as external discs, USBs, video projectors, and Wi-Fi (USB) to avoid the connection cut.

8.3. Conception Planning

Conception planning is an important for a project based on a technical platform. This would allow the team members to gauge the level of understanding of the contributors and redirect them before the scheduled training. These planning sessions are also an opportunity to elucidate certain points that can be raised such as the remuneration of actor, their focal points and decisions on who will do the basic work in their faculty.

8.4. Internal and External Evaluation of the Activities and the Teaching Tool

The team must evaluate internally the activities that they will lead. The evaluation allows the team to understand what has been done, where the team failed, and how to manage to adjust and reorient better. It is an important occasion to justify, to improve, and to follow or redirect activities when there is a problem. Extern collaborators can also evaluate activities to improve them as it can be difficult to judge oneself.

9. Methodology

The project will be decentralized, open and participatory in order to demonstrate the concrete practice of Open science while observing transparency in the ethics of research. Our methodology will be based on the following actions, but we emphasize that, with the exception of the platform draft made on Wordpress, no other activity has yet been done among the ones listed.

9.1. Activity 1 – Co-construction of the Web Site of the Science Shop (Already Done) by Inviting Each Members of the Scientific Community and Civil Society to Contribute Actively in Its Training Activities

It was a bit difficult to implement the platform without funding. To do this, I conceptualized and asked the assistants to do the editing. At the meeting, we discussed it and each member of the team brought their criticisms. After integrating the critics of the day, I shared the work with the EBAD teacher colleagues for further criticism. This platform is beneficial for UCAD because it will serve as a framework for sharing and exchanging between the components of society and academics. Thus, it will be an essential tool for sustainable development in Senegal.

9.2. Activity 2 - Presentation of the Science Shop Project, as a Teaching Tool at the Cheikh Anta Diop University's Research Ethics Committee for Evaluation, (Re) Orientation and Validation

UCAD is endowed with an ethics committee that controls all the pedagogical tools that must be used in teaching. This phase becomes almost a formality with this platform which received the approval of 36 colleagues out of 36 during the workshop of Saly where I made a demonstration as well as shared the Direction of the research of the UCAD. In general, when teacher-researchers judge and accept a tool, the committee only ratifies. So, when the time comes, the SOHA team of Dakar will make the presentation.

This activity is a necessity for the pedagogical tools that teachers use as part of their courses. This activity is an added value for the training dispensed at UCAD and it's an obligation.

9.3. Activity 3 - Administration of an Interview Guide and Distribution of the Questionnaire in Various Forms to All Project Targets (Government, Universities, Civil Society, NGOs

Students who are engaged to conduct these surveys after raising awareness of the targets in question will reinforce the Science Shop team.

9.4. Activity 4 - Identification of Participants (Students, Teachers, Researchers, Pats or Any Interested Person) Ready to Commit and Mobilize Them for Start of the Project

As mentioned above, the platform was presented to the participants at the Saly workshop last March and they all showed their willingness to be the focal point in their structures to start the project. A dozen have already filled in the online form and we will add the others as soon as time permits. This list is of great interest to the project because these pioneers will popularize open science in their establishments and it will be easier to gather them for future activities and symposia with the support of the pioneers.

9.5. Activity 5 - Presentation of This Pedagogical Tool (Science Shop) in a Series of Lectures in Order to Make Clear the Purpose and Operation (On the Site of the Science Shop as Well)

The use of the baobab symbol will allow the message to pass quickly. In addition, we had a foretaste in Saly and we will use more arguments in the workshops and symposia to organize if we have the necessary funding. The explanation will vary according to the public but, we will adapt whenever necessary. These workshops' aim is to popularize open science through the use of the platform and this will be of benefit to UCAD because its actors will be mainly trained in open science and will use local knowledge to better take Community concerns, work with the other actors to conduct projects together, find solutions collaboratively and co-build a sustainable development.

9.6. Activity 6 - Creation of an Open Database, Accessible Free of Charge and Reusable by All, Based on the Data Collected and the Results Achieved by the Students

The creation of an open database, accessible and reusable by all, based on the data collected and the results obtained by the students will allow a better sharing of the information for all actors. The results of the surveys and the works of the students must be widely disseminated to the stakeholder communities. This database is useful because it will allow free access to all the results of the project and even to others that will be communicated to us.

9.7. Activity 7 - Review of the Range of Training Strategies in the Documentary Research Methodology That Participants Have Experienced and How to Incorporate an Open Science Approach into the Activities of the Boutique

To better evaluate the platform and the activities carried out there, it will first be necessary to evaluate the methods used by the students to better help those to come. We have already begun to teach the literature search methodology course in advance of the project because it is one of the subjects I teach at EBAD. But, this evaluation will allow me to enrich my courses and to update it constantly. This course is beneficial for all the actors of the UCAD. It makes it possible to improve the presentation and the quality of scientific works because I add to it the copyright, quotation, plagiarism and of course, the presentation of the bibliographic reference with Zotero, EndNote, Mendeley, etc. We also teach them how to use Research4Life resources. To do this, we adapt the course according to the discipline and the learner specialties.

9.8. Activity 8 - Organize Physical and/or Virtual Meetings Between Local Leaders and Learners Every 3 Months to Inform About the Progress of the Project (Videoconferences, Skype)

Organize physical and/or virtual meetings between local leaders and learners every 3 months to inform them about the progress of the project by videoconferences and Skype (this may also include webinars). The five public universities of Senegal are located in distant cities and actors teach almost in all of them. Sometimes an actor may be at another city or elsewhere in the world. So these virtual encounters will be beneficial for teachers and all. To succeed in such challenges, we must equip ourselves accordingly.

9.9. Activity 9 - Development by the Steering Committee of a Roadmap for the Adoption of Open Science in Universities/Higher Education Institutions and Validation of This Roadmap During a Symposium Between the Project Stakeholders

The development by the Steering Committee of a roadmap for the adoption of open science in universities and higher education institutions for validation of this roadmap during a symposium between the project stakeholders is necessary. This roadmap will be deposited with the authorities of each university hoping to set up a Science Shop and we can even strive towards the creation of a national Science Shop network because we share the same department. This will be the generalization of open science in the higher education institutions of Senegal. If we succeed in creating an African network of Science shops, local communities from all member countries will be able to interact with academics and vice-versa.

9.10. Activity 10 - Evaluation of the Activities of the Science Shop and the Means Put in Place to Ensure Its Perpetuation (Purchase of Laptops, USB Keys and Creation of a Database to Disseminate This Tool Even in the Absence of an Internet Access During a Conference)

The evaluation of the activities of the scientific store and the means put in place to ensure its perpetuation (purchase of laptops, USB keys and creation of a database to disseminate this tool, even in the absence of Internet access during a conference) is a necessity for the team members. In some cities in the interior of the country, the stays are always present, that is why it will be necessary to surround oneself with guarantees. The team members will need materials to present the platform everywhere and whenever needed.

9.11. Activity 11 - Creation of Sponsors and Partnerships to Support Activities but Benefit the Science Shop's Results

As we started with a lot of financial difficulty, we thought it was necessary to look for sponsors but it is not possible at the moment because we have no products on the platform. Sponsors will help to realize many of the activities.

9.12. Activity 12 - Organization of a Discussion Group with Committed Participants in Each University to Collectively Develop a Maintenance Guide to Reflect on the Conditions for the Implementation and Use of This Collaborative Teaching Tool in All Public Universities of Senegal

The organization of a discussion group with committed participants in each university to collectively develop a guide to reflect on the conditions for the implementation and use of this collaborative teaching tool in all public universities of Senegal will have more benefit for all actors and institutions. This kind of networking is an important form of extension of the project.

10. Expected Results

Each of these results is related to one or many of activities we defined.

10.1. Result 1 - Universities and Society Work Collaboratively (Together) on Projects with a Territorial Social Impact

We are convinced that when the activities we have defined are carried out efficiently, universities and society will finally be able to work collaboratively on projects with a local, regional, territorial, subregional or even international social impact.

10.2. Result 2 - The State, Civil Society Organizations and Companies Find in Solidarity, Appropriate Solutions to Their Different Concerns, Thanks to the Intervention of Motivated Students and Teachers-Researchers

If students are trained in Open Science and have projects they understand, the State, civil society organizations and companies find in solidarity, appropriate solutions to their different concerns, thanks to the intervention of motivated students and teachers-researchers. The best outcome of all these results is the appropriation by each actor of the platform that will lead to the further interaction of more actors. Thus, all will become full actors, not tele-spectators. Each actor has a major part to play in order to make the project successful and useful for all. Hence our Science Shop name is "Knowledge of all for all" and all actors must keep this in mind while doing his activities.

10.3. Result 3 - Students Find Adequate Answers to Social Concerns Enabling Them Both to Implement the Knowledge Acquired in Courses, to Acquire and to Experiment New Notions Through Scientific Works

After solid training on Open Science and the use of platform, students find adequate answers to social concerns enabling them both to implement the knowledge acquired in courses, to acquire and to experiment new notions through scientific works. They experiment with what they learnt. It allows teachers-researchers to check. This makes it possible to verify the accuracy and relevance of the new knowledge and skills acquired in the project.

10.4. Result 4 - The Emergence of New Types of Citizen Actors in the African Space

The emergence of new types of citizen actors in the African space derives from the quality of the management and the training's actors in open science. If there are new types of citizen, the goals of the projects are achieved.

10.5. Result 5 - A Framework of Excellence Geared Towards Meeting the Needs of Society Through Applied Research Is Created

When a framework of excellence focused on meeting society's needs through applied research is created, it means that stakeholders have taken the platform and lessons on Open Science. Actors have assimilated trainings and lessons. Now, they interact to solve the problems and needs of society as a whole together, without discrimination and prejudices on local and scientific knowledge. This is the reason for this pleasant and beneficial environment for all.

10.6. Result 6 - Sustainable Local Development Is Co-Built by All the Actors in Society

Once all the activities set out in the project have been carried out, all actors feel at ease in collaboration and exchanges. They are all decompartmentalized, committed and motivated. They have become new types of citizens and succeed in the overall objective that the project had set in solving the starting hypothesis: the actors of society, together, co-constructing a sustainable local development.

11. Conclusion

The UCAD Science Shop is an important element in the fight against scientific compartmentalisation and helps to break down the barriers between universities and society. It also enables academics to know and accept other forms or types of local knowledge. Thus, academics interact with society in a spirit of service to the community, fulfilling the third societal mission of universities. Students acquire new knowledge from society that teachers value, as well as the organizations they serve by finding solutions to their needs. This type of innovative system puts universities at the heart of the community's activities and facilitates access to scientific knowledge for local actors, while developing the qualities of commitment. The UCAD Science Shop is thus perfectly in line with the sustainable development program of Senegal.

References

[1] GRANGET, Lucia., La responsabilité sociale des universités à l'heure du savoir comme marchandise, Communication et organisation, 26 (2005), 127-147. (En ligne) https://communicationorganisation.revues.org/3281

[2] GRAU, Francesc Xavier (et al.), Higher education in the world 6. Towards a socially responsible university: Balancing the Global with the Local, Guni Network, 2017, 516 p., ISBN: 978-84-617-5508-0 (Online) http://www.aqu.cat/doc/doc_39815142_1.pdf

[3] MICLEA, Mircea, Beyond rhetoric. Some obstacles for a responsible glocal university, Guni Network, 2016 (Online): http://www.guninetwork.org/articles/beyond-rhetoric-some-obstacles-responsible-glocal-university (Consulté le 22/04/2017)

[4] EUROPEAN COMMISSION, Connecting Universities to regional Growth : A practical guide, 2011. http://ec.europa.eu/regional_policy/en/information/publications/guides/2011/connecting-universities-to-regional-growth-a-practical-guide (Online) (Consulté le 20/02/2017)

[5] OKATCH, Moses, McCOWAN, Tristan & SCHENDEL, Rebacca, The impact of tertiary education on development: A rigorous literature review, London, Department for International Development, 2014. (Online): http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/Tertiary%20education%202014%20Oketch%20report.pdf?ver=2014-06-24-161044-887 (Consulté le 04/01/2016)

[6] Shabani, J. (2008). The role of key regional actors and programs. Higher education in Africa: The international dimension. Accra/Boston: AAU/CIHE, 464-489.

[7] DIOUF, Diéyi et PIRON Florence (2015). La science ouverte comme outil collectif de développement du pouvoir d'agir et de la justice cognitive en Haïti et en Afrique francophone: vers une feuille de route (projet SOHA, 2015, (En ligne): http://www.scienceetbiencommun.org/sites/default/files/projet_science_ouverte_en_afrique_et_haiti_-_version_finale_1.pdf (Multiple Consultations…)

[8] BOERE, E., HEIJMEN, W. J. M., The social value of science shops : a cost-benefic analysis. APSTRACT: Applied Studies in Agribusiness and Commerce, 06, 2012 (En ligne): http://EconPapers.repec.org/RePEc: ags: apstra: 138114 (Consulté le 31/03/2017)

[9] DeBOK, C., & STEINHAUS, N. Breaking Out of the Local: International dimensions of science shops, Gateways; International Journal of Community Research and Engagement, 1 (2008), 165-178. (Online). http://www.academia.edu/625033/Breaking_Out_of_the_Local_International_dimensions_of_science_shops (Consulté le 16/01/2017)

[10] HAWKINS, L., MULDER, H., & STEINHAUS, N., Building a Science or Research shop: Refining or Expanding Your Model, Présenté à CU Expo, 2013, Corner Brook, NL, Canada (Online) http://research.library.mun.ca/1806/ (Consulté le 23/01/2017)

[11] LEYDESDORFF, L., & WARD, J., Science shops: a kaleidoscope of science–society collaborations in Europe. Public Understanding of Science, 14(4), 2005, 353-372. (Online), https://doi.org/10.1177/0963662505056612 (Consulté le 27/02/2017)

[12] MULDER, H. A., & DeBOK, C., Science shops as university-community interfaces: an interactive approach in science communication. In : D. Cheng, J. Metcalfe, & B. Schiele, At the Human Scale – International Practices in Science Communication. Beijing, Science Press, 2006.

[13] ACHAFFERT, Hamissou Rhissa, La science ouverte juste et le projet SOHA au Niger: quelles pratiques pour quels avantages ?, Justice cognitive, libre accès et savoirs locaux. Éd. Science et bien commun, 2016 (En ligne): https://scienceetbiencommun.pressbooks.pub/justicecognitive (Consulté le 03 01/2017)

[14] CHARLIER, J.-E., & CROCHE, S, L'inéluctable ajustement des universités africaines au processus de Bologne. Revue française de pédagogie. Recherches en éducation, 172 (2010), 77-84. (En ligne): https://doi.org/10.4000/rfp.2276 (Consulté le 10/03/2017)

[15] MILLOT GLEN, Le tiers-secteur scientifique – Exemple des boutiques de sciences, Fondation Sciences Citoyenne, 2012 (EN LIGNE) http://sciencescitoyennes.org/le-tiers-secteur-scientifique-exemple-des-boutiques-de-sciences/ (Consulté le 03/12/2016)

[16] CAMARA, Mamadou Lamine. Les chiffres de la décadence: Université Cheikh Anta Diop. Interview avec Yankhoba Seydi, le 02/10/2014 (En ligne): http://www.seneplus.com/article/les-chiffres-de-la-décadence (Consulté le 30/12/2016)

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-283

The Challenge of Creating the Cyprus Academic Library Consortium (CALC): Impacts and Benefits

Marios ZERVAS[a,1] and Maria HARAKI[a]
[a] *Library and Information Services, Cyprus University of Technology*

Abstract. This paper describes the development of the Cyprus Academic Libraries Consortium (CALC) and the process of formulating its own agreements with Publishers, but not before explaining how the Public Universities of Cyprus managed to participate in the Hellenic Academic Libraries Link (HEAL-Link) Consortium. Insufficient funds, lack of organization, danger of overlapping information resources, create a slow growth rate and obstacles in the provision of information material by Academic Libraries to their users. Realizing these problems and trying to avoid them CALC was created – comprised of seven members, with its main action limited only to subscriptions to Electronic Resources. Furthermore, criteria are outlined describing reasons for choosing services and electronic resources from Publishers. Also, the dissemination of cost per member institution for each deal is described through a formula created by the library directors and agreed to by the senate of each member institution. Although there are benefits of academic libraries joining the CALC consortium, the fact that there are practical and financial problems as well as challenges is not ignored. Methods, such as an evaluation on the usage of the resources and promotion of the same, must be applied by the librarians of each university consortium member. Qualitative methods are discussed and quantitative methods are presented as essential, to better evaluate the impact of resources in the research area (Cost Benefit Analysis (CBA) and Return on Investment (ROI) formulas). With attention to situations like the overcoming of the Swets (CALC's main supplier) bankruptcy, the paper continues to show that success can be achieved through proper record keeping, hard work and good relations – which is how CALC managed to continue its subscriptions to the Publishers and avoided any interruption of access to the electronic journals. Although setting up a consortium can provide many benefits, CALC is also in the process of acquiring its Legal Entity in order to be better able to process its administrative and cost actions. Further development and future actions are mentioned, including discovering how researchers are using the library's resources to help with their teaching and research.

Keywords. consortium, subscriptions, collection development

1. Introduction

This paper will describe the development of the Cyprus Academic Libraries Consortium (CALC). Some background regarding the history of the public universities

[1] Corresponding author, Arch. Kiprianou 31, 3036, Limassol, Cyprus; Email: marios.zervas@cut.ac.cy.

and their involvement with the HEAL-Link consortium, an initiative that has contributed to the creation and operation of CALC.

The criteria for choosing services and electronic resources from Publishers are outlined which state important reasons for participating in a consortium. Benefits are discussed in order to stress the burdens of library budgets and the limitations that libraries would have if they were to act alone. Problems and challenges are presented through examples and real experiences. Cost Benefit Analysis and Return of Investments are applied in order to demonstrate the benefits of participating CALC members in a deal with a publisher.

Finally the importance of acquiring a Legal Entity is an important measure for the avoidance of managerial constraints and for CALC to be better able to process its administrative and cost actions. Future actions are outlined to emphasize that CALC is not a static organization and will continue to evolve and enrich its activities.

2. A Little History

Before proceeding to describe how CALC was created, it is important to explain how the Public Universities of Cyprus managed to participate in the Hellenic Academic Libraries Link HEAL-Link Consortium [1].

Insufficient funds, lack of organization, danger of overlapping information resources all contribute to a slow growth rate and obstacles in the provision of information material by Academic Libraries to their users.

Realizing these problems and trying to avoid them, collaboration strategies and actions in Cyprus were initiated by the public libraries sector and in particular the University of Cyprus.

The history of the universities in Cyprus is recent. The University of Cyprus (UCY), which is the first higher education institution in Cyprus, was established in 1992. It officially became a member of HEAL-Link in February of 2000.

This initiative paved the way for the Cyprus University of Technology (CUT) and the Open University of Cyprus (OUC) to follow and they also became official members of HEAL-Link in 2009. This was a big achievement considering that CUT and OUC only began to accept their first students in 2007 and 2006 respectively.

In parallel, the three public universities created the Cyprus Academic Libraries Consortium (CALC) with the purpose of subscribing to additional resources and services that could not be covered through the HEAL-Link membership. Also, as members of HEAL-Link, we did not have the opportunity to choose whether or not to participate in a deal. The Publisher deals were for all the HEAL-Link members without exception. The three (3) Public Universities of Cyprus paid a pre-agreed percentage for each deal.

Sadly and due to the world economic crisis, Greece was faced with economic uncertainty and consequently this created many challenges for HEAL-Link and its members. This is why, in 2013, the Cypriot Universities withdrew from HEAL-Link based on a mutual understanding.

As early as 2006, independently of HEAL-Link, the three public universities had already started to collaborate at a local level.

3. Creation and Operation of CALC

Since the three public universities of Cyprus (CUT, UCY & OUC) were no longer members of the HEAL-Link Consortium, something had to be done in order to maintain their electronic resources and to continue to be able to serve their respective academic communities. Despite the economic difficulties Cyprus was facing due to the restrictions and implementation measures imposed by Troika, CALC expanded in 2013 to include private universities, and the process of formulating its own agreements with existing and new Publishers was completed successfully. In addition, all payments were made through the supplier Swets Information Services BV.

When we began collaborating directly with Publishers we used:

a) The percentage participation rate that we had with HEAL-Link as the starting point for our negotiations and,
b) The 2012–2013 Cypriot financial crisis as an argument due to the extremity of the financial situation at the time (Haircut). "On 25 March 2013, a €10 billion international bailout by the Eurogroup, European Commission (EC), European Central Bank (ECB) and International Monetary Fund (IMF) was announced, in return for Cyprus agreeing to close the country's second-largest bank, the Cyprus Popular Bank (also known as Laiki Bank), imposing a one-time bank deposit levy on all uninsured deposits there, and possibly around 48% of uninsured deposits in the Bank of Cyprus (the island's largest commercial bank), many held by wealthy citizens of other countries (many of them from Russia) who were using Cyprus as a tax haven. No insured deposit of €100,000 or less would be affected" [2].

The CALC consortium is comprised of seven members, of which four belong to the private sector. Its main action, so far, is limited only to subscriptions to Electronic Resources (electronic journals, databases, ebooks).

Increasingly, electronic resources have become the mainstream format for academic libraries [3].

CALC moved to direct contracts, negotiations and final agreements with more than 20 recognized Publishers and Suppliers of scientific journals and electronic resources and continued to collaborate with Swets.

4. Criteria for Choosing Services and Electronic Resources from Publishers

The main criteria for choosing these services were:

- The ability to provide on-line access to electronic content
- Providing abstracts and full-text files of scientific journals in subject areas related to the research and educational needs and directions of CALC members
- Obtaining depth of coverage and perpetual access
- Obtaining appropriate equipment and software for the development of the ILL service

Additional criteria for the selection of electronic resources (journal packages and databases) were:

- The common interest of the CALC member institutions in subject categories, as recorded in their research programs and related activities
- The large number of published journals and subscriptions in these subject categories
- The great cost of these journal subscriptions
- The importance of early notification for the academic community on the latest developments in these sciences

Each library sets priorities and determines collection needs in relation to its budget which is allocated by the university board. The CUT is unofficially the coordinator in regards to correspondence and negotiations with the Publishers. The dissemination of cost per member institution for each deal is based on a formula created by the library directors and agreed to by the senate of each member institution. The formula takes into consideration the FTEs (no. of Undergraduate Students, no. of Post Graduate Students and number of Academic & research Staff) of the members who participate in a specific deal.

5. Benefits

In the framework of benefits, this paper presents some examples of cases where certain existing consortium members, participating in a deal, have benefited from additional members joining.

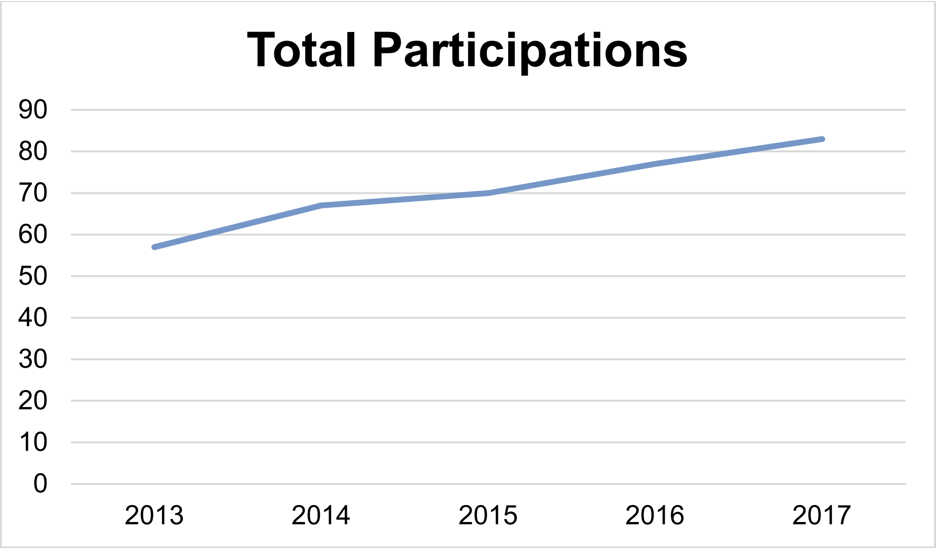


Figure 1. Total participations per year, 2013 -2016

According to Figure 1, it can be concluded that when a new member to an existing deal is added, it automatically implies more savings. Therefore, it is not obligatory to participate in any consortium deals and contribution is based on a library's ability to participate depending on budget availability and priorities.

Figure 2 presents an example of the benefits of participating in a consortium and a winning situation.

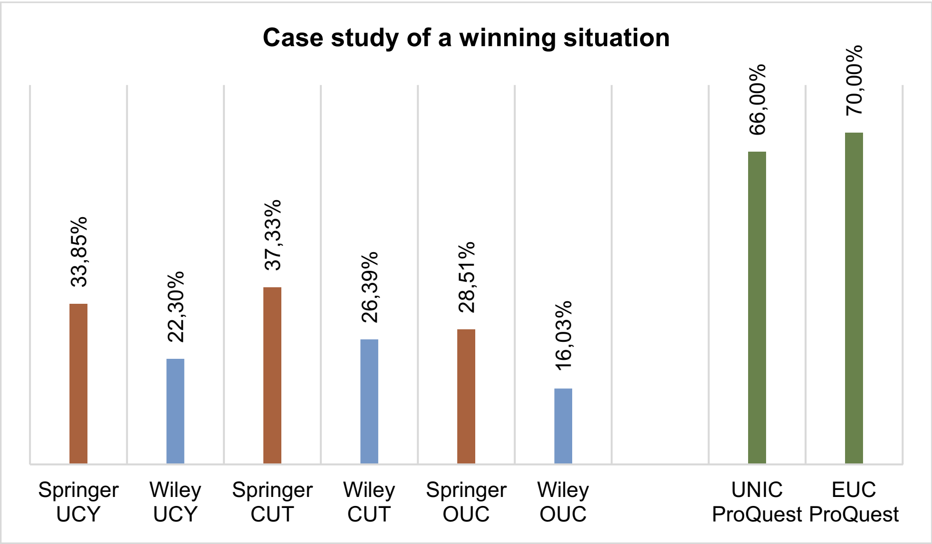


Figure 2. Case study of a winning situation

The combination of these offers from the Publishers, according to the above Chart, allowed CALC to have full participation, increased content and cost savings.

For example, in the case of Springer and Wiley, in 2013 the only participants were the three public universities. While researching the interest of the rest of the CALC members to participate in the renewal year 2014, only one private university expressed interest in Springer, namely FU.

Regarding the interest for Wiley, only UNIC was interested, because it already subscribed to Wiley independently and was paying a higher amount in comparison to the public universities.

Regarding ProQuest Central only two private universities had a subscription (UNIC & EUC). The public universities were not so interested because they already subscribed to EBSCO Academic and Business Complete which covered their needs.

Taking into consideration the above, the Publishers recognized that there was no interest on the part of the private universities to participate in the Springer and Wiley deals and so they decided to make an attractive offer to include all the CALC members.

As a result, Springer's offer included all the members with a 19% increase in the original price.

Wiley's offer included all the members with a 6.3% increase in the original price.

ProQuest's offer included all the members with an 11.4% increase in the original price.

The combination of these offers from the Publishers allowed CALC to have full participation and increased content.

Speaking on behalf of CUT, these deals benefited our University by saving approximately €20,000 yearly.

In 2013, CALC signed for the first time with Springer and Willey a multiyear “Big Deal” [4], where all the CALC members participated and benefited with access to a large portion of the publisher journals. In support, Kenneth Frazier noted “the Big Deal appears to be an especially good deal for smaller academic libraries that join with a consortium in purchasing a major comprehensive journal license” [4]. The creation of our consortium has benefited the participant libraries with more content and better pricing.

Since its creation, the CALC has benefited very much from the negotiation process with the Publishers. Everything to do with correspondence, negotiations, license agreements, invoicing, payments, access etc. has helped us acquire experience and knowledge which has proven to be invaluable towards our actions and future plans.

Such actions include, coordination and administration of CALC, invaluable learning experience, connections and development of good relations with Publishers/Suppliers and development of communication and cooperation among the CALC members.

6. Problems and Challenges

Although there are benefits for academic libraries joining the CALC consortium, we cannot ignore the fact that during its course there were also problems and challenges.
Swets Information Services BV was the Supplier who dealt with the invoicing and payments between the HEAL-Link Consortium and the Publishers. Once CALC was formed, it made sense for us to continue our cooperation with Swets and thus we had agreed for the invoicing to be done through Swets. The year 2013 was processed smoothly and all the payments to the Publishers were completed successfully. For the year 2014, unfortunately, CALC was faced with a very big problem due to the sudden news of the Swets bankruptcy!

Some payments were completed but some were not. In the case of four (4) publishers the money never reached them. As you can imagine, this problem created a lot correspondence between CALC and the Publishers. Proofs of payment had to be sent to the Publishers in order to show that, on CALC’s part, payments were indeed made. CALC also took the initiative and sought the representation of a lawyer in Holland.

On the other side the trustee in Bankruptcy on behalf of Swets in Holland, was sending statements claiming payments from some of the CALC members.

After dealing with the trustee in Bankruptcy and proving them otherwise about their wrongful claims, we still had to deal with the Publishers who never received their payment. Success in proving the Trustee in Bankruptcy wrong with regards to their claims came through proper record keeping and good filing systems.

Additionally, through hard work and good relations, CALC managed to continue its subscriptions to the Publishers and avoided any interruption of access to the electronic journals.

One example of good relations and mutual understanding was a case of a Publisher who was owed hundreds of thousands by CALC due to the Swets bankruptcy. The

Publisher ended up providing a waiver of claims towards CALC and in turn CALC had subscribed to additional material.

The Swets Bankruptcy may have caused problems for CALC, but it did not damage the negotiations section due to the fact that CALC, from the beginning, dealt with the Publishers directly.

Since then and until the present, CALC does not use a supplier for its deals with Publishers. The Swets bankruptcy has made our collaboration with Publishers more flexible. Publishers accept to invoice the CALC members separately. They also accept payment in installments and consequently CALC saves a lot on the commission that was paid to a supplier.

Another problem/challenge is the fact that some universities have had to cancel subscriptions due to the lack of sufficient funds. The increase of commonly subscribed resources raises the risk of the remaining members to be burdened economically should a member withdraws from a contract. "The collecting goals have to be balanced with collection evaluation or assessment activities at the local level" [5].

In order to overcome the above, an evaluation on the usage of the resources and promotion methods must be applied by the librarians of each university consortium member.

In the case of CUT, the evaluation of activities is performed at a local level by our library and our Tender Committee is then made aware of the cost per search and per download for each package of electronic resources. Our library follows the same method as explained by [6] and the collection policy document prepared is based on two primary goals: a) the curriculum and b) research support. The CUT Library's target in the long term is for the subject librarians to have an in-depth knowledge of their specific areas of the library collection and to help in collection development [5].

Many of our members must also set similar targets as it has become apparent, that they too need to make better use of the necessary personnel and apply their knowledge and expertise in order to evaluate their collections and know the usage of resources. This will help them to support their decisions and their according Committee to approve, with more certainty, the renewal of subscriptions.

7. Cost Benefit Analysis and Return of Investments

To quantify the above analyses, the CUT has adapted a combination of Cost Benefit Analysis (CBA) and Return on Investment (ROI) formulas and analyzed consortium deals with Publishers. That's why cost-benefit and return of investment analyses need to be done at the consortium level. As Judy Luther said, "a review of existing research identified several cost/benefit analyses based on user surveys and faculty productivity studies correlating citations and grants. However, there were no models for calculating a return on investment (ROI) in academic libraries" [7].

In our study, we are using the model by Betsy Kelly in order to calculate the CBA and ROI [8].

The formula for ROI is (a) "the total value of expense for the consortium to produce benefits" minus (b) "the total value of expense for the specific library to produce benefits" divided by (b) "the total value of expense for the specific library to produce benefits" and multiplied by one hundred.

The formula for CBA is (a) “the total value of expense for the consortium to produce benefits” divided by (b) “the total value of expense for the specific library to produce benefits”.

$$ROI\% = ((a - b)/b)*100 \quad \text{ROI is represented as a percentage} \tag{1}$$

$$CBA = a/b \text{ is represented as a ratio CBA: 1} \tag{2}$$

When the ROI is more than 1% or the CBA more than 1:1, they are considered positive thus indicating that the value of benefits surpasses the cost to provide access to the content.

Table 1. CBA and ROI

Library	Nb of Resources	Consortium Price	Library share	CBA	ROI
UCY	19	1933828	620486	3,11	212
CUT	18	1922087	305679	6,28	529
Neapolis	11	833461	47086	17,70	1670

By observing Table 1, it can be deduced that small libraries benefit more than larger libraries.

This conclusion is derived using the FTEs of each university consortium member and our formula, in order to calculate a library’s participation in each deal.

For example, for every Euro spent, CUT receives €6,28, UCY €3,11 and Neapolis €17,70.

A cost per use analysis is done every year and the data is presented to our Tenders Committee in order to justify the need for renewals.

A combination of quantitative and qualitative analyses is essential to better evaluate the impact of resources in the research area. Tenopir wrote, “ROI is only one method for measuring the value of a library’s collection and services. The benefit of multiple methods is that numbers in and of themselves rarely tell the full story” [9].

Further development and actions will include measuring the impact of the subscribed journals in relation with the articles cited by our researchers output. In the future, surveys and interviews will be conducted in order to discover how researchers are using the library’s resources to help with their teaching and research.

8. Legal Entity

Although setting up a consortium can provide many benefits, it demands a big commitment from all the members. In order to be better able to process its administrative and cost actions, CALC has already completed the drafting of the legal document in order to acquire its Legal Entity. This action aims to help the CALC to operate as an official nonprofit company.

The Legal Entity outlines the main terms and objectives thus making official all the actions of CALC in relation to the deals made between the members and the Publishers/Suppliers.

Without a Legal Entity, important managerial constraints are created such as: no autonomy, no dedicated budget, no management of budget and staff, etc.

9. Future Actions

CALC can go beyond the academic sector and can contribute to the public sector also, e.g. small government research institutions.

It is our belief that the below points relating to future actions can help in the evolution of libraries:

- The creation and function of a union catalog of bibliographic records and the development of common cataloguing policies can contribute to the continuous education of library staff and the sharing of expertise.
- The joint implementation of technical infrastructure, e.g. creation of electronic repositories, thematic portals, digitization and archive management etc.
- The systematic monitoring and adoption of international practices and standards in matters relating to the operation of libraries e.g. such as quality issues, copyright issues, and the adoption and use of creative commons licenses or other similar licenses.
- Providing seminars for library users to maximize the use of information resources.
- Provision of consulting services on library related issues.

10. Conclusions

It is our belief that Consortia has helped and contribute towards the development of libraries.

In the case of CALC, the collaboration among its members, has helped by establishing a common policy on electronic journal subscriptions, with the aim of promoting rational growth of journal collections among partners, and savings and access to a greater number of electronic sources in order to meet the educational and research needs of the users of the participating institutions. Proof of this is shown through the Cost Benefit Analysis (CBA) and Return on Investment (ROI) demonstrated above and especially for small Universities.

Also, CALC's experience maintaining good relations and strong collaborations with Publishers also helped with the overcoming of various problems and obstacles.

It is our belief that CALC can go beyond the academic sector and can contribute to the public sector also, e.g. small government research institutions.

A challenge that is of great concern to CALC, is when universities have to withdraw their participation from subscriptions due to the lack of sufficient funds. Evaluation of the usage of the electronic resources can help to prevent this from happening and minimize the risk.

References

[1] HEAL-Link consortium Retrieved April 10 2017, Available at http://www.HEAL-Link.gr/?lang=en
[2] 2012-2013 Cypriot financial crisis retrieved April 11, 2017, Available at https://en.wikipedia.org/wiki/2012%E2%80%9313_Cypriot_financial_crisis
[3] Pan, D. & Fong, Y., 2010. Return on Investment for Collaborative Collection Development : A Cost-Benefit Evaluation of Consortia Purchasing. *Collaborative Librarianship*, 2(4), pp.183–192. Available at: http://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1232&context=collaborativelibrarianship.
[4] Frazier, K., 2005. What's the Big Deal? *The Serials Librarian*, 48(1–2), pp.49–59. Available at: http://www.tandfonline.com/doi/abs/10.1300/J123v48n01_06.
[5] Kinner, L. & Crosetto, A., 2009. Balancing Act for the Future: How the Academic Library Engages in Collection Development at the Local and Consortial Levels. , 49(4 OP-Journal of Library Administration. Jun2009, Vol. 49 Issue 4, p419-437. 19p. 5 Charts.), p.419. Available at: http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=ehh&AN=39254211.
[6] Evans, G.E., 2002. Management issues of co-operative ventures and consortia in the USA. Part one. *Library Management*, 23(4/5), pp.213–226. Available at: http://dx.doi.org/10.1108/01435120210429943 [Accessed January 3, 2017].
[7] Luther, J., 2008. University investment in the library: What's the return? A case study at the University of Illinois at Urbana–Champaign. *Library Connect white paper*, (1), p.18.
[8] Kelly, B., Hamasu, C. & Jones, B., 2012. Applying Return on Investment (ROI) in Libraries. *Journal of Library Administration*, 52(8), pp.656–671.
[9] Tenopir, C., 2010. Measuring the Value of the Academic Library: Return on Investment and Other Value Measures. *The Serials Librarian*, 58(1–4), pp.39–48.

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-293

Co-Constructing an Open and Collaborative Manifesto to Reclaim the Open Science Narrative

Denisse ALBORNOZ[1,a], Alejandro POSADA[a], Angela OKUNE[b], Rebecca HILLYER[c] and Leslie CHAN[a]
[a] *University of Toronto Scarborough*
[b] *University of California, Irvine*
[c] *University of Stellenbosch*

Abstract. The OCSDNet Manifesto is a result of one year of participatory consultations and debates amongst members of the 'Open and Collaborative Science in Development Network' (OCSDNet), a network of 12 research-practitioner teams from Latin America, Africa, the Middle East and Asia. Through research projects grounded in diverse regions and disciplines, OCSDNet members explore the scope of Open Science as a transformative tool for development thinking and practice and offer the 'Open and Collaborative Science Manifesto' as a foundation upon which to reclaim the mainstream narrative about what Open Science means and how it can realise a more inclusive science in development. This paper describes the mechanisms used for collaboration and consensus building, and explores the ways in which the process of building this document serves as a case study for the opportunities and limitations of integrating collaboration, opportunities for participation and openness into research activities.

Keywords. Open science, inclusive science, right to research, cognitive justice, collaboration

1. Introduction

This paper describes the process by which the OCSDNet arrived at the Open and Collaborative Science Manifesto and the opportunities and limitations of integrating openness, collaboration and opportunities into network research activities. The Open and Collaborative Science in Development Network (OCSDNet) is composed of twelve researcher-practitioner teams from the Global South interested in understanding the role of openness and collaboration in science as a transformative tool for development thinking and practice. Research teams are supported by a team of four external advisors and a network Coordination Team. The project is funded by the International Development Research Centre (IDRC) in Canada and the Department for International Development (DFID) in the UK. Throughout this paper we will make a distinction between the OCSDNet coordination team and the 12 research teams of sub-grantees selected to conduct research and implement projects in their respective countries and regions. The OCSDNet coordination team is comprised of the Principal

[1] Corresponding Author; E-mail: albornoz.denisse@gmail.com.

Investigator Leslie Chan, and four research associates: two associates from Latin America based in Toronto, an associate from Canada based in South Africa, and an associate from US based in Kenya. The coordination team, on top of managing the network, is also in charge of collecting data, reporting and consolidating findings, and as such, is also a 'research team'. The term has been left out as to avoid confusion between this group and the sub-grantees.

The development of the OCSDNet Manifesto is a response to the lack of transformative and critical approaches to Open Science in the global scientific and development community. Most mainstream narratives about Open Science, emerging particularly from Europe and North America, envision open science as a system of technology-driven tools and processes [1] [2] [3] that, when utilised, are assumed to accelerate scientific discoveries, improve transparency and reproducibility of research, increase research uptake, and improve accountability to the scientific community as well as to the public [4] [5] [6]. While we recognize a great deal of progress has been made through technology-enabled collaboration, we also note that the established voices in the Open Science community have failed to address how the current approach to "open" exacerbates and amplifies disparities in knowledge production and circulation [7] [8] [9] [10] [11].

OCSDNet imagines open science in a different way. We envision Open Science as an intrinsically inclusive and collaborative practice that is constantly striving to be reflective about power and privilege within structures of knowledge creation. With this in mind, the 'Open and Collaborative Science Manifesto' invited network members to collectively question and discuss the knowledge ecosystem in their contexts – asking to whom does knowledge belong?; are benefits of science disproportionately concentrated to some privileged groups over others?; who gets to participate in knowledge production processes?; and in what ways can technology be used to increase the agency of more people over scientific knowledge production?. We recognize that these questions have not been adequately raised and debated in conversations about Open Science or deliberated at the intersections of Open Science and Development [11] [12].

Using these questions as a starting point, the Manifesto has evolved to emcompass seven common principles as seen in Figure 1. The seven principles arrived at by the network pose that Open and Collaborative Science in Development: 1) enables a **knowledge commons** where every individual has the means to decide how their knowledge is *governed and managed* to address their needs; 2) recognizes **cognitive justice** and the need for *diverse* understandings of knowledge making to co-exist in scientific production; 3) practices **situated openness** by addressing the ways in which *context, power* and *inequality* condition scientific research**;** 4) advocates for every individual's **right to research** and enables different forms of *participation* at all stages of the research process; 5) fosters **equitable collaboration** between scientists and social actors and cultivates *co-creation* and social innovation in society; 6) incentivizes **inclusive infrastructures** that empower people of *all abilities* to make, and use accessible open-source technologies and; 7) uses knowledge as a pathway to **sustainable development,** equipping every individual to improve the *well-being* of our society and planet.

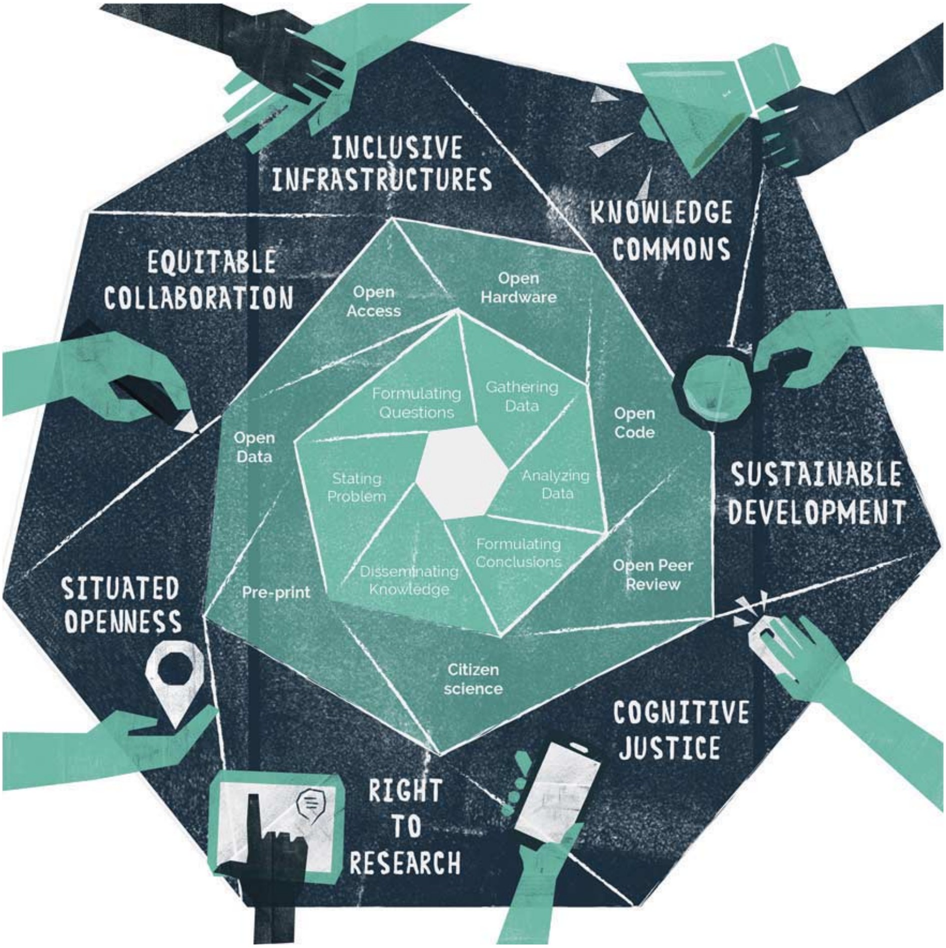


Figure 1. OCSDNet Manifesto Infographic

2. Co-Constructing a Manifesto

The idea of constructing a Manifesto was born in May 2015, after several members of the network met in Singapore to present at the ICTD Conference and realized the network needed to produce a document that outlined our position in the Open Science debate, calling for a more inclusive, collaborative and just approach to knowledge production. While network members came from different disciplinary, cultural and ethnolinguistic backgrounds, we shared the notion that the story about Open Science needed to be reclaimed and refocused – from the technocentric rhetoric dominating the debate, back to the values and a vision for a world that Open movements propose.

From June 2015 onwards, the OCSDNet coordination team conducted a series of participatory, collaborative and horizontal consultative mechanisms, which took place over the course of one year to tap into the synergies and divergences in our vision for Open Science. These included formal project reports and position papers, as well as more informal group calls, workshops and collaborative editing sessions in which network participants shared and debated their views about what Open Science means for them and their communities. The coordination team specifically looked for common keywords, themes, and ideas that encapsulated the principles and processes guiding the research practice of the 12 research teams. In addition, we also carried out feedback sessions to improve the content of our document and to develop a tone, language and dissemination format that reflects the inclusive and collaborative spirit of the scientific model it proposed. The result was an optimistic, reflective and critical Manifesto that we hope promotes conversation in the scientific community and beyond, about the rationales we use to advocate for and propose a redefinition of Open Science.

It is important to acknowledge that the process of consultation and the production of this Manifesto was informed by the many scholarly traditions that have historically challenged the hegemony of positivism and a market-driven scholarly communication system. As such, the principles comprising the Manifesto are not new and have been central to fields such as critical theory, postcolonial, feminist and indigenous epistemologies among others (Figure 2). As part of our process, we gathered these various ideas and documented the ways in which they informed the principles of the Manifesto in a collaborative annotated bibliography and reading list.[2] Through this open resource, we aim to pay homage to the work of these authors, but also to further make visible the intersections between Open Science and the many streams of social justice scholarship. We also hope that its users will continue to make suggestions and contribute to it as the understanding of Open Science and the field continues to expand.

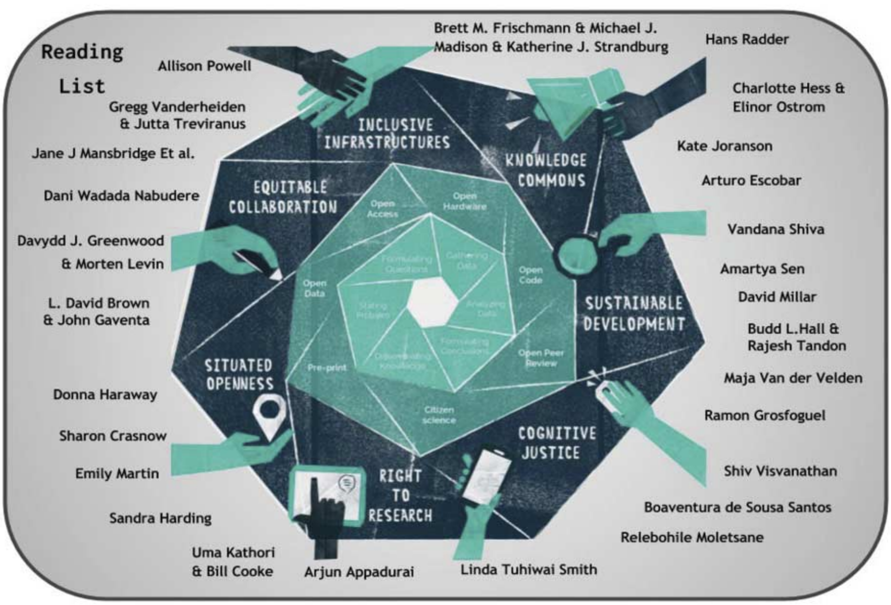


Figure 2. OCSDnet Manifesto Principles and Reading List of Key Authors

[2] The annotated bibliography and collaborative reading list is available here.

Finally, to increase the impact and reach of the Manifesto, it has been translated into several languages (French, Spanish and Afrikaans thus far) and multi-modal formats that invite diverse audiences to join the conversation including an online visual infographic, a 3-minute animated video, posters and pamphlets, blog posts, an open syllabus/reading list, among others, in collaboration with the Cooperativa de Diseno from Buenos Aires, to be posted on the OCSDNet website. Through these efforts, we aim to measure the impact and influence this document can have in sparking conversations about Open Science in local and global scientific communities.

3. Manifesto as a Methodology

While the participatory nature of the Manifesto was an extremely valuable data collection activity (as explained in 3.1.), it allowed us to do much more. The Manifesto became an overarching methodology in our practice – a system of methods that enabled us to constantly pay attention, monitor, and evaluate the myriad of research practices, working styles, interactions, relationships and power dynamics that are taking place across the network. More specifically, it allowed us to address the four pillars driving our research:

3.1. Learning Analysis

Firstly, the Manifesto was a tool to gain a deeper understanding of the values, findings and lessons that drive the thinking and practice of the network. We asked "what have we learned through network case studies about what is required for open and collaborative science in development?". Through participatory consultations, the 12 research teams collectively brainstormed, discussed and synthesized the contextual conditions, practices and normative values required to facilitate openness and collaboration in science as per the experience of the communities with whom they were working. This data was consolidated into the seven principles outlined in the Manifesto, producing a document that tells the story of Open Science from a situated point of view, informed by the diverse cultures, disciplines and identities that make up their practice.

The dialogue facilitated questions such as "what principles should be included in the Manifesto?", or "indicate if you agree or do not agree with the following principles". These questions created the conditions conducive for productive debate regarding the assumptions we were making around each principle, and the extent to which they reflected the contextualized nature of their practice. For example, mid-way through the process, the second principle read "scientific knowledge and infrastructure should be open, accessible and freely available to all". This statement generated discussion particularly between groups who defended open hardware and DIY technologies in citizen science practices. They tended to advocate for a type of openness that empowered the general public to take ownership of the technologies and knowledge(s) that affect them: "We want 'truly open' to be something more diverse; more dynamic, involving new actors and groups". Meanwhile, the team from South Africa, working closely with indigenous communities who have historically been marginalized and dispossessed in large part due to policies of openness, eloquently argued that we needed to work towards a language that demonstrates a more situated approach to openness that takes histories, power and inequality into account.

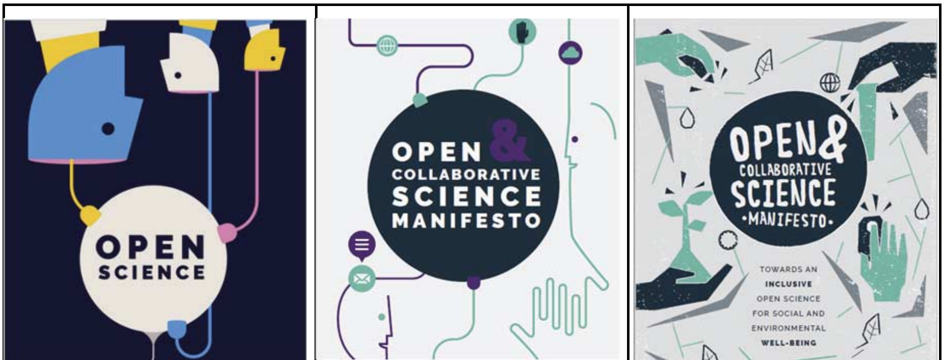
Ultimately the second principle evolved into what today is principle 3, abandoning overarching generalizations about openness as being positive and beneficial for all communities and arriving at a more nuanced and self-aware approach that extended to the rest of the concepts. The network's use of the concept 'open' evolved from a debate about the state of open vs. closed, towards a conceptualization that is embedded in the socio-political dimensions of knowledge. We also began to acknowledge the risks of open knowledge sharing, while aiming to redistribute opportunities to overcome barriers impeding participation in scientific research. Reflections around our use of language to explore contextualized openness also led to a decisive turning point in our objectives in which we set out to question the power of discourse and narratives of Open Science as a first step in questioning the norms, attitudes, behaviours, practices and ultimately policies that exacerbate and amplify disparities in knowledge production.

3.2. Network Building

Drafting a common vision also means consensus has to be reached via dialogue and debate. As posed by our second research question: "what have we learned about the tools, research processes, styles of governance and leadership required to drive a network on open and collaborative science in development?", we are also interested in the infrastructure required to support a network; particularly one that is working towards generating knowledge and experiences aimed at informing their collective understanding of Open and Collaborative Science.

As a result, our process of building and consolidating the OCSDNet network has required constant and iterative reflection on the ways in which a networked and distributed social and technical infrastructure facilitates knowledge exchange, collaboration and equitable governance. In this respect, the Manifesto, as a consultation process and an exercise in collaboration, gave network members and the coordination team the opportunity to interact with one another, have difficult conversations and in doing so, get to know the interests, motivations and thought process driving the work of their peers.

This was also the case for the process of 'knowledge translation', which required several iterations and rounds of discussions between the graphic designers and the OCSDNet coordination team, to make sure the designers capture the nuances and the contextual nature of the concepts presented in the Manifesto into a graphic form. For example, the first graphic proposals were more techno-centric, given how the graphic design team thought about science, and evolved into graphics that conveyed the more community-led and participatory nature of a more open and collaborative science. (See Figure 3).



Graphic proposal July - 2016	Graphic proposal September - 2016	Graphic Proposal October - 2016

Figure 3. Evolution of OCSDnet Manifesto Graphic identity as visualized by Design Cooperative

The experience of conducting a consultation was also instrumental for our understanding of the nature of transnational collaborative working. As a general observation, the method allowed us to redistribute the narrative power in terms of defining the story we would tell as a collective, from the central OCSDNet coordination team to the network members who are implementing the projects on the ground and thus have a deeper understanding of the different layers of openness. However, the participatory nature of the consultation did not include everyone to the same degree, and failed to break down many of the barriers for participation present in the network, such as language and cultural barriers, lack of incentives to share and the various levels of power and influence between members situated in different institutional contexts.

Mitigating these challenges required a considerable and unanticipated amount of involvement and facilitation from the coordination team. The OCSDNet coordination team attempted to minimize interference in content creation, but rather focused on designing consultation methods that maximized the participation of network members and created opportunities to accommodate those who were not as “loud” or active in the activities. This included creating ample opportunities for participation, via one-on-one interviews, e-mail, conference calls and collaborative writing, so as to bypass these barriers and ensure that most members had the chance to contribute to the final document, including those who expressed some discomfort during offline and public consultations. This experience highlighted that collaboration cannot always be institutionalized or systematized, as networks still rely on organic community building.

In spite of these efforts, participation in the process was not always welcomed nor perceived as positive, and was rather received with mixed responses by some members of the network. In some cases, participants acknowledged that the consultation was taking place but did not fully participate in the process, unless it was part of a mandatory report or as a result of a one-on-one request from the coordination team. Other participants often did not feel compelled to think normatively about Open Science nor did they see how such a document could advance the objectives of their projects. However, the case for most members was active participation in offline and online consultations, going beyond the brainstorming stage, and taking ownership of the editing process and working closely with one another to refine the direction of the final product.

3.3. Field Building

The third question and pillar driving our research looked at the extent to which we are contributing to the creation and expansion of the OCS in Development field. Field building, as defined by Lynn's work on advocacy and evaluation [13], involves a) defining a common advocacy goal, b) sharing the values driving these advocacy efforts, c) and using a language that expresses these goals and values, as well as inspiring collective action. As the nature of a Manifesto suggests, the principles are a public declaration and articulation of the values driving the thinking and practice of our network. Producing such a document allows us to collectively contribute to the field of Open and Collaborative science as it develops, evolves, interacts with related fields exploring openness (such as Open Access, Open Data, Open Education, etc.), and responds to emerging critiques. However, it is important to note that our network has intentionally positioned itself and its research at the intersections of Open Science and Development scholarship as well, which thus far remains largely under-researched. In doing so, we are attempting to build and contribute to a new field of study and practice.

However to get there, as noted by Lynn [13], on top of a common vision and a collective, field building also involves *action* and developing common strategies to advance the group's objectives into decision-making, agenda-setting, and change-making at different levels. Under this definition, the process of building the Manifesto was an instrumental process in developing and negotiating a common language and vocabulary. Yet as a document, it has not yet proven its potential to be a useful advocacy tool to challenge the narrative around Open Science. Through the dissemination of the final product, we would like to continue raising questions of influence and power with audiences beyond our network, questioning *who has the power when important decisions are being made*, and probing how we can use this tool to actively and strategically engage other actors in advocacy around Open Science in Development. We intend for the document to ultimately provide an initial roadmap on how to collectively realize a more open and collaborative science, while considering that the steps to realize this vision will also need to be contextualized and situated within their contexts and realities.

3.4. Reflective and Adaptive Learning

Finally, the Manifesto was a key activity for reflection and iterative learning – providing us with evidence to answer the fourth and final question driving our research: "to what extent have we effectively engaged the sub-projects throughout the learning process to promote a culture of reflexive and adaptive learning in the network?". While other data collection activities such as the collection of monthly reports or reflection papers enabled us to interact with the project teams on a regular basis, the Manifesto was one of the most consistent spaces for dialogue, debate and exchange between network members, allowing us to receive feedback from network members to help us improve and adapt the ways in which we were driving the consultation.

While the coordination team attempted to negotiate or mediate some of the conflicts, the exercise also brought to light some of the Global North-South power dynamics embedded in collaborative network relations. Even though the Manifesto is very critical about the dominance of 'western' science, in some instances, researchers from North American institutions dominated the conversation and thus, drafted a significant bulk of the content of the Manifesto. One of the core challenges in negotiating this dilemma was bypassing the language barriers. For example, members from the West African team who speak French and members from the Kyrgyz team who speak Russian were considerably less active in their contributions. This was a lesson for the coordination team to account for the implications of ethnolinguistic and cultural barriers in the way we structure participatory collaborative exercises, and iteratively adjust our mechanisms to emerging dynamics.

In this respect, the Manifesto activity also provided the space and opportunity for collective and self-reflection, proving to be a tool that creates checks and balances to mitigate imbalances between the coordination team and the network participants. At one stage of the consultation process, a project team from the Middle East constructively suggested that the coordination team was overemphasizing its critique of Open Science as a practice that reproduces and amplifies inequalities in the draft, while not paying enough attention to other understandings and perceptions present in the network. As articulated by one of the project leads from Lebanon during the Bangkok network meeting:

> "Constantly emphasizing inequalities in power relations, implies that one side is less powerful. And we are not. We *are* powerful. We are practicing a new, better model of science."

Since the coordination team is affiliated with a North American university and has the most direct access to resources and funding, our dominance over the narrative becomes problematic in a context where we are fostering and exploring more horizontal and collaborative approaches to knowledge making. The Manifesto was a process that allowed us to collectively work on the tone, language and arguments used by OCSDNet to build a narrative of Open Science that represents a majority who were involved in the process.

4. Next Steps

Moving forward, OCSDNet has produced a Monitoring and Evaluation framework to track the impact and uptake of the Manifesto, by strategically identifying the actors we would like to engage in discussion, and the different ways in which we could introduce this conversation in diverse regions, fields and areas of work. This strategy has been designed by identifying different types of audiences at the community, institutional and policy level, identified through a network-wide mapping exercise, and a gradient of outcomes in relation to three levels of possible effects of interacting with the Manifesto (See Figure 4)

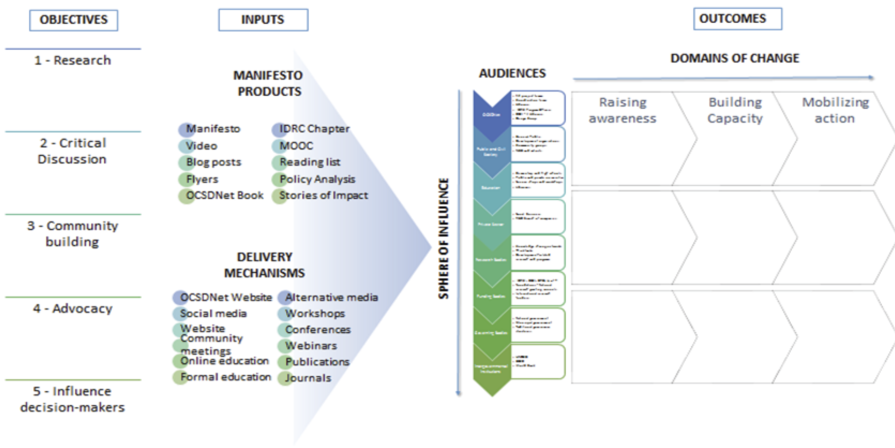


Figure 4. Outline of Monitoring and Evaluation framework for dissemination of Manifesto

At an initial level we *expect* to see how ideas and concepts introduced in the manifesto are generally recognized, remembered and understood. At an intermediate level we would *like* to see if there is growing public support for the Manifesto and increased capacity of key actors to mobilize and advocate for an inclusive open science in their context. Finally, at an ultimate level, we would *love* to see how specific actions, programs and practices are designed to realize the vision raised by the Manifesto. As part of methodology, we have also recognized the need to undertake a discourse analysis of the Open Science field as one of the ways through which to strengthen our argument regarding the need to reclaim the Open Science narrative, as well as to better understand the impact that a collaborative manifesto and its nuanced take on language, can have on reshaping the mainstream discourse.

5. Conclusion

Through the process of constructing this Manifesto, we aimed to amplify a narrative of what Open Science can mean for development, the language we use to discuss it, and the core reasons why we are advocating for it. In the process, we achieved a series of objectives that align with our theory of change as a research network. First, we consolidated the core values that ground our thinking and practice as a network and were able to articulate a shared understanding of what a more inclusive take on Open Science can offer to scientific research and development. The processes of Open Science take place in a wide array of contexts by diverse individuals who face and negotiate a unique set of barriers, challenges, and opportunities in their regions, and this Manifesto was designed and constructed to capture these experiences. Second, we were able to develop a common language among the OCSDNet community to discuss openness, collaboration and development, and include concepts that are relevant and appropriate across multiple contexts and perspectives. Third, we created an output that allows us to introduce a reflexive and critical framework into global and local conversations about Open Science that recognizes the role of power relations in knowledge production, which we hope, will ignite future debate and discussions; and finally, we produced an accessible advocacy tool to disseminate and communicate OCSDNet's vision to relevant and diverse audiences, including policy-makers, scholars, educators and the general public, in multiple languages and formats.

At a more meta level, we hope the Manifesto, both as an output and consultation process, ultimately mirrors the opportunities and challenges of embodying the principles it communicates. As a document, we desire for the Manifesto to stand as an invitation and tool to initiate a broader conversation in the Open Science community about the way in which these structural inequalities continue to shape global scientific knowledge production; and as a process, to offer a roadmap and a case study on what happens when you integrate openness, collaboration and opportunities of participation into various research processes. In the case of our network, doing so gave us tremendous hope in terms of the importance and feasibility of introducing a value-based framework into collaborative research processes. However, the process of co-creating a collective document also showed us that even amongst researchers who align with a more inclusive and equitable scientific practices, collaboration still requires constant reflection, brokering and facilitation. This allowed us to stay creative in terms of the mechanisms we set in place to foster openness, collaboration and participation between members, but more importantly to remain critical of their limitations, particularly in terms of the type of 'participation' we were able to enable.

We remain aware that the vision articulated in this document is just a snapshot of the current thinking and state of the debates of our network. The principles put forth are by no means sufficient or conclusive, and the reclaiming and reshaping of the Open Science narrative needs to be an ongoing process. We view this product and process as non-static and expect it to evolve as each team furthers their experience and understanding of openness and collaboration. On the same account, the monitoring and evaluation strategy will also allow us to continue reflecting on how our identity as a network and our willingness to collaborate with one another changes after this phase of the research is over. Can a document and a participatory consultation such as a Manifesto bind us as a collective and set a foundation for future partnerships and collaborations? Will actors take ownership of the product and continue to disseminate this vision among their networks? Or will it reveal we are far from consolidating a common agenda around a more inclusive Open Science? Either way, this is only an initial step of a larger effort to better understand the development implications and outcomes of co-constructing and diversifying the narratives and arguments for an open and collaborative science.

References

[1] OECD. (2015). *Making Open Science a Reality* (OECD Science, Technology and Industry Policy Papers). Paris: Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd-ilibrary.org/content/workingpaper/5jrs2f963zs1-en

[2] Grigorov I., Elbæk M., Rettberg J., & Davidson J. (2015). Winning Horizon 2020 with Open Science. *Zenodo*. https://doi.org/10.5281/zenodo.12247

[3] Orth, A., Pontika, N., & Ball, D. (2016, June). FOSTER's Open Science Training Tools and Best Practices. In Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing (p. 135). IOS Press.

[4] Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

[5] Leonelli, S., Spichtinger, D., & Prainsack, B. (2015). Sticks and carrots: encouraging open science at its source. *Geo: Geography and Environment*, *2*(1), 12–16. https://doi.org/10.1002/geo2.2

[6] McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... & Spies, J. R. (2016). How open science helps researchers succeed. *Elife*, *5*, e16800.

[7] Nyamnjoh, Francis. (2009/2010). Open Access and open knowledge production processes: Lessons from CODESRIA. The African Journal of Information and Communication, 10, 67–72. http://link.wits.ac.za/journal/AJIC10-Nyamnjoh.pdf

[8] Tkacz N. From open source to open government: A critique of open politics | ephemera. ephemera: theory & politics in organization. 2012;12(4):386–405.

[9] Tyfield, D. (2013). Transition to Science 2.0: “Remoralizing” the Economy of Science. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, *7*(1), 29–48. https://doi.org/10.4245/sponge.v7i1.19664

[10] Kansa, E. (2014). It’s the Neoliberalism, Stupid: Why instrumentalist arguments for Open Access, Open Data, and Open Science are not enough. Retrieved from http://blogs.lse.ac.uk.myaccess.library.utoronto.ca/impactofsocialsciences/2014/01/27/its-the-neoliberalism-stupid-kansa/

[11] Okune, A., Hillyer, B., Albornoz, D., Sambuli, N., & Chan, L. (2016). Tackling Inequities in Global Scientific Power Structures. The African Technopolitan Vol. 4, Issue 1.Retrieved from https://tspace.library.utoronto.ca/handle/1807/71107

[12] Chan L, Okune A. & Sambuli N., 2015. “What is open and collaborative science and what roles could it play in development?” In: Sarita Albagli, Maria Lucia Maciel, and Alexanfre Hannud Abdo (editors). Open science, open issues. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), pp. 87–112, and at http://livroaberto.ibict.br/bitstream/1/1061/1/Open%20Science%20open%20issues_Digital.pdf

[13] Lynn, J. (2014). Assessing and evaluating change in advocacy fields. Centre for Evaluation Innovation. http://www.evaluationinnovation.org/sites/default/files/Spark-Evaluating_Change_In_Advocacy_Fields.pdf

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-305

EKT ePublishing: An Innovative Service to Support Open Access Scholarly Publishing in Greece

Marina ANGELAKI[a,1], Dimitris EFSTATHIOU[a] and Irakleitos SOUYIOULTZOGLOU[a]
[a] *National Documentation Centre/National Hellenic Research Foundation*

Abstract. The poster presents EKT's ePublishing platform, an innovative service to support open access publishing in Greece, which enables the research community of the country to transition from a print-only mode of work to online working environments and enhance the visibility and impact of their research outputs. Emerging within an ecosystem with no prior experience or open access oriented culture, it enables the cultural shift towards open and collaborative scientific practices and the open science/open access paradigm. Focusing on the Open Book Press, EKT through its participation at the HIRMEOS project is expected to enhance the technical capacities and services of the platform that enable identification, authentication and interoperability as well as tools that enrich information and entity extraction.

Keywords. Electronic publishing, electronic books, open access, Greece

1. Introduction

The paper presents the electronic publishing platform of the National Documentation Centre (EKT) in Greece. The National Documentation Centre (EKT) – part of the National Hellenic Research Foundation (NHRF) – supports research and technology by providing infrastructure and services for the use and dissemination of Greek scientific and cultural content, while placing emphasis on open availability and the reuse of content [1]. EKT has developed an ePublishing platform, an innovative service to support open access publishing in Greece. The platform enables the research community of the country to transition from a print-only mode of work to online working environments and enhance the visibility and impact of their research outputs.

The paper focuses on the development of EKT ePublishing platform, its impact on the Greek research community and the challenges faced in the adoption of this experiment. The paper will then place particular focus on the foreseen improvements of the book platform (Open Book Press) undertaken in the context of the European funded project HIRMEOS.

[1] Corresponding Author, National Documentation Centre/National Hellenic Research Foundation, 48 Vas. Konstantinou Avenue, GR-11635 Athens; E-mail: mangelaki@ekt.gr.

2. EKT ePublishing: The History and Challenges in Developing the Service

The ePublishing platform is based on EKT's successful collaboration with non-profit research organisations and scientific societies focusing primarily on – but not limited to – the Social Sciences and Humanities. EKT ePublishing platform was launched in 2007 with the transition of three print journals in the humanities to an online and print format [2]. The development of the service was co-financed by Greek and European funds under the "Digital Convergence" operational program and currently hosts three distinct platforms (for journals, books and conference proceedings) providing services to 27 publishers and hosting close to 7.000 articles, 600 conference papers and 6 monographs.

EKT ePublishing platforms use open source software that support in full the digital editorial process (OJS for the ejournals and eProceedings platform and OMP for the Open Book Press which hosts monographs). EKT provides a wide range of services to publishers including among others web hosting, online management of the publishing process, OJS training, technical support, helpdesk service, consulting services in producing guidelines and policies aligned with current international developments, and usage statistics. In addition, it provides persistent identifiers and indexing services, which significantly increase online availability and visibility of high quality Greek content and enhance the impact of research published in Greece. Visitors have direct and full open access to downloadable peer-reviewed content available in multiple formats (html, pdf, epub) through the different platforms with the use of advanced search tools and retrievable metadata information. This is achieved through a single-access point giving full access to online content, and allowing navigation by scientific discipline, metadata information. Those wishing to submit their manuscript to a selected journal can do so by using the electronic submission process [3].

EKT ePublishing was developed within an ecosystem with no prior experience or open access oriented culture. Despite the existence of necessary infrastructure (mainly repositories) there is still low awareness on the benefits associated with the transition to an open access environment and low engagement on the part of key stakeholders [4]. Furthermore, the findings of a previous study showing the relatively low uptake of complete online management of the editorial process by some journals hosted in the platform seem to be still valid [5]. Nevertheless, EKT ePublishing plays a key role in enabling the cultural shift towards open and collaborative scientific practices and the open science/ open access paradigm. A further challenge relates to the long-term financial sustainability of the project. EKT's participation in a number of EU-funded projects allows the smooth operation of the platform and the implementation of a series of technical improvements.

3. The eBooks Platform

The Open Book Press is one of the three epublishing platforms dedicated to the support of electronic monographs [6]. The importance of a platform dedicated to monographs is justified by the fact that monographs are still an essential part of the scholarly communication in the social sciences and humanities. The platform was launched in 2015 and currently hosts 6 monographs. EKT through its participation at HIRMEOS, a Horizon 2020 project, will enhance the technical capacities and services of the platform that enable identification, authentication and interoperability as well as tools that enrich information and entity extraction. These improvements are expected to render technologies and content interoperable and allow the platform to be embedded into the European Open Science Cloud.

4. Conclusions

EKT ePublishing has become a unique service for the Greek research and academic community in alignment with EKT's strategy of providing horizontal infrastructure and services to stakeholders in Greece. EKT's participation in European projects and networks allows it align the services provided with international standards. Future plans focus on expanding collaborations and strengthening the technical capacities of the service. In relation to the ebooks platform ensuring a steady source of financing is a key issue.

Acknowledgments

The improvements in the technical capacities of EKT's eBooks platform are undertaken in the context of the Horizon 2020 HIRMEOS project (Grant Agreement No 731102).

References

[1] National Documentation Centre www.ekt.gr
[2] Nafpliotis, A., Tsoukala, V., Houssos, N., Kalaitzis, A. and Sachini, E. EKT ePublishing: Developing an open access publishing service for the Greek research community,18th International Conference on Electronic Publishing. Let's put data to use: digital scholarship for the next generation, *Proceedings of the 18th International Conference on Electronic Publishing (Ebook)*. IOS Press BV, Netherlands, pp. 112-118, 2014 http://helios-eie.ekt.gr/EIE/handle/10442/14317
[3] ePublishing, www.epublishing.ekt.gr
[4] Picarra, M. Regional Challenges in Achieving Open Access and Proposed Recommendations, PASTEUR4OA, August 2015, https://zenodo.org/badge/DOI/10.5281/zenodo.44327.svg
[5] Sachini, E., Tsoukala, V., Houssos, N., Stathopoulou, I-O., Paschou, Ch-E., Paraskevopoulou, A. Open Access in the Humanities: a case study of developing three open access electronic journals in Greece, 13th Conference on Electronic Publishing: Innovation and Communication Paradigms and Technologies, 10-12 June 2009, Milan, Italy, pp. 543-556.
[6] eBooks, http://ebooks.epublishing.ekt.gr/

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

doi:10.3233/978-1-61499-769-6-308

Developing an Academic Publishing Service Continuum

Matthias AMMON[1,a] and Danny KINGSLEY[a]
[a] *Office of Scholarly Communication, Cambridge University Library*

Abstract. This paper describes the establishment of a continuum of publishing and preservation services for the academic community at the University of Cambridge, particularly in relation to grey literature. It sets out the initial identification of the need for this service and the process of establishing a variety of options. As the project is at an early stage, the paper discusses the particular issues such an initiative faces in a research university consisting of a large number of institutions with significant levels of autonomy.

Keywords. Library publishing, academic led publishing, grey literature, repositories, academic publishing

1. Introduction

The University of Cambridge is formed from a variety of institutions, including over 100 academic departments organised into six schools. As the starting point for an investigation into academic-led publishing initiatives at the University of Cambridge and the potential need to offer services in this area through the Office of Scholarly Communication, a search was undertaken for publications emanating from the departments of the University that did not come under the usual output category of peer-reviewed journal articles or monographs. The exact amount of material falling under this category is extremely difficult to quantify, as it is not always clear whether publication series are ongoing or defunct, what the association with the University is, or who the contact person might be. Faculties and departments at the University enjoy a large amount of autonomy in administration. It is therefore difficult to get an overview of publication undertakings that happen across the university, as there is no central place where such information is collated. Such initiatives may indeed be undertaken by members of the academic community on a personal basis without necessarily having an official affiliation with a department.

An analysis of the material that was discovered established that the publications comprised a wide variety of outputs and formats. In terms of publication method this ranged from photocopied sheets of paper to professionally produced and indexed publications, both in print and online. Discussions with some of the originators of the content indicated a lack of knowledge about, and understanding of, the need for unique identifiers, indexing services and preservation. In some cases the same material was

[1] Corresponding Author, Office of Scholarly Communication, Cambridge University Library, West Road, Cambridge, United Kingdom, CB3 9DR; E-mail: mra29@cam.ac.uk.

being published across multiple platforms – each with their own unstable links, causing identification, citation and long term preservation issues.

In addition, the lower quality publications posed a potential branding challenge for both Cambridge University Press and the University of Cambridge itself. Discussion with both organisations resulted in the decision to provide solutions to increase the quality and stability of academic-led publishing outputs to a standard that is expected of the University of Cambridge. In addition it was clear that a solution needed to be found for capturing, preserving and disseminating these outputs as part of the research undertaken by the academic community at Cambridge.

2. Context

Institutional repositories have grown in importance within UK universities due to funding mandates requiring the research output of funded work be made Open Access. In addition to a place of deposit for peer-reviewed literature under these terms, institutional repositories also offer a place for the capture and dissemination of grey literature. For our purposes, grey literature refers to outputs produced by or in association with university departments that have not gone through the peer-review process and/or are published by academics or departments themselves, rather than a commercial publisher. This includes but is not limited to full academic journals, working papers, technical reports, student journals, lecture series, conference proceedings, one-off publications and other ephemera. In the first instance, the focus was on the capture of already existing material, such as back issues of journals, for preservation, with a view to channelling future publications through the infrastructure hosted by the University Library.

Cambridge University established a DSpace repository in 2005. The Office of Scholarly Communication (OSC) was established in 2015 and took responsibility for the management of the repository. In 2016 the OSC upgraded the repository with an improved interface and name - Apollo. In addition to Open Access versions of peer-reviewed articles, the repository also holds research data, such as 175,000 chemical molecules, and is home to several thousand theses.

Clearly Apollo is a simple solution to many of the use cases we uncovered in our original audit of academic-led publishing in Cambridge. One of the first activities in this project was to address the vulnerability of the *Cambridge Journal of China Studies*. This journal had been in print publication for a decade with an online version of the articles available on an unstable website. The OSC undertook a bulk upload of all the past articles for the journal to Apollo[1]. This meant the articles now had digital object identifiers (DOIs), were properly indexed and discoverable through search engines and their long-term preservation was being addressed.

However, the inadequacy of the repository interface for the purposes of displaying a journal's table of contents became apparent. The repository is organised into collections, which can provide statistics for and pull related material together, however the standard interface with a linear list of content is not appropriate for, say, the display of images, or conference proceedings.

As work has progressed over the past year, the OSC has had more in-depth discussions with departments about other potential use cases for publication. Several research areas have a long established tradition of publishing working paper series. While these were usually initially paper publications, mostly they are now available online on a department's website. In one instance, some investigation identified that the papers were being uploaded into several websites as well as in the repository, meaning there were several different links to the same material, only one of which (the repository version) had a DOI attached or displayed. This is clearly far from ideal.

Other use cases that have arisen include the desire to publish conference proceedings, and a platform for publishing typeset and copyedited monographs.

While the details of the needs of the academic community may be specific to Cambridge, the recent rise of new university presses or academic-led presses, based in or originating from the university library, in the UK indicates a desire on the behalf of university-centred academic communities and the academic libraries associated with their institutions to provide publishing capabilities and services that establish alternatives to the traditional publisher model [2].

In this regard, Cambridge is at an advantage because it already has a long established University Press, with all the infrastructure and know-how that a large academic publisher is able to provide. On the other hand, Cambridge University Press is precisely that, a large academic publisher with commercial considerations and established products, costs, tools and policies. It is a considerable bonus to be able to leverage the infrastructure of an established publisher and to collaborate with CUP. However, the relationship between Cambridge University Press and the University affects the extent to which the Library can offer publishing services that may be perceived to be in competition with CUP.

There are opportunities for offering a range of publishing services to reflect the multifarious nature of use cases and user needs discussed above, allowing as much flexibility and independence as desired on the part of the publishing departments or academics. There is a distinction between 'Library-led' publishing and 'Academic-led' publishing. The former, according to the Library Publishing Coalition, is a set of activities led by college and university libraries to support the creation, dissemination and curation of scholarly, creative or educational works [3]. Academic led publishing, on the other hand, can be defined as set up and run by academics, usually not for profit and providing alternative publishing options to commercial publishers.

Discussions with our academic community have demonstrated that the need for a solution – a publishing service in this instance – is often only perceived when a concrete example of the proposed tool can be shown. Conversations so far have indicated that outreach to academics and development of the various options therefore need to run in parallel, so that needs can be established while at the same time attracting interest by being able to offer a ready-made service.

The OSC has identified a clear gap in service provision for immediate action and is targeting the material that falls outside the usual peer-reviewed publication model. These research outputs have an urgent need for capture and preservation. Strategically, the value of the assistance the library would be able to provide for this material can be most readily perceived in these cases.

3. Future Developments

The development of this service is still in its early stages. We ultimately envisage a continuum of services that we will be able to offer the academic community for the capture and preservation of any kind of research output. Based on the variety of user needs and discussions that have been held with members of the academic community, departmental administrators and representatives of graduate student journals, we are planning to launch the full service with a multi-stage continuum. The stages are:

1. The repository 'as is': individual items or collections can be deposited in the repository with no additional visual or editorial additions.
2. Repository with enhanced display: a repository collection presented with a visually improved user interface customised to reflect the logo or colours of the originating author, department, organisation or event.
3. Publication overlay module: a customisable module that can be integrated into a department's website via its web design software, with links to metadata and individual items in the repository. The module automatically checks for new content added to the repository and updates the page on the department's website. This allows for customised images and texts such as editorial introductions while keeping the hosting and preservation in the hands of the repository management team. This is in part modelled on the journal *Discrete Analysis* which uses a similar system for preprints published on arXiv.org [4].
4. Facilitating academic-led publishing, possibly by providing publishing and hosting options for the community to develop their own open access outlets.
5. Working with a professional publisher to capitalise on their infrastructure to provide publishing services for non-peer-reviewed literature.

Work is underway on all stages of the service and we are hoping to have a full offering sometime in 2017.

References

[1] *Cambridge Journal of China Studies* Collection in Apollo https://www.repository.cam.ac.uk/handle/1810/255394

[2] Lockett A, Speicher, L. (2016), New university presses in the UK: Accessing a mission. Learned Publishing 2016, 29: 320–329.

[3] Library Publishing Coalition, https://www.librarypublishing.org/

[4] *Discrete Analysis*, http://discreteanalysisjournal.com/

L. Chan and F. Loizides (Eds.)

Subject Index

Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices
L. Chan and F. Loizides (Eds.)

Author Index

This page intentionally left blank

This page intentionally left blank

This page intentionally left blank