Key Aspects and Approaches of Open Access, Open Data and Open Science

Iryna Kuchma

eIFL, c/o ADN Kronos, Piazza Mastai 9, 00153 Rome, Italy
iryna.kuchma@eifl.net

Abstract. Overview of the key aspects and approaches to open access, open data and open science, emphasizing on sharing scientific knowledge for sustainable progress and development.

1 Open Access

Open access (OA) is free, immediate, online access to the results of research, coupled with the right to use those results in new and innovative ways. There are two complementary strategies to achieve open access to scholarly literature: open access journals and open access repositories.

Open access journals use a funding model that does not charge readers or their institutions for access (subscription or access fees). Users can read, download, copy, distribute, print, search, or link to the full texts of the journal articles. These journals do no longer invoke copyright to restrict access to and use of the material they publish. Instead they use copyright and other tools to ensure permanent open access to all the articles they publish. For the journal publishers, open access brings increased readership and, with that, increased citations, and maximum visibility and impact for a journal's contents. And it means that the best possible dissemination service is being provided for research.

Directory of Open Access Journals (DOAJ, http://doaj.org/) lists 9,931 peer-reviewed scholarly journals from 134 countries. 5,827 journals are searchable at article level. And a total of 1,695,228 articles are freely and openly available. There are 44 open access journals from Bulgaria listed in DOAJ:

- Cybernetics and Information Technologies, and General and Applied Plant Physiology (Publisher: Bulgarian Academy of Science);
- International Journal of Myriapodology, BioRisk, Subterranean Biology, NeoBiota, Nature Conservation, PhytoKeys, Comparative Cytogenetics, ZooKeys, Journal of Hymenoptera Research, and MycoKeys (Publisher: Pensoft Publishers);
- Mathematica Aeterna (Publisher: Hilaris);
- Ecologia Balkanica (Publisher: Union of Scientists in Bulgaria);
- Journal of Theoretical and Applied Mechanics, and Folia Medica (Publisher: Versita);

Digital Presentation and Preservation of Cultural and Scientific Heritage, Vol. 4, 2014, ISSN: 1314-4006
Autumn Training School. Development and Promotion of Open Access to Scientific Information and Research - Lecture Notes

- Bulgarian e-Journal of Archaeology (Publisher: Association of Bulgarian Archaeologist);
- Journal of BioScience and Biotechnology (Publisher: Plovdiv University Press) and ZooNotes (Publisher: Plovdiv University);
- Novo Znanie (Publisher: University of agribusiness and rural development);
- International Journal Bioautomation (Publisher: Academic Publishing House);
- Trakia Journal of Sciences, and Bulgarian Journal of Veterinary Medicine (Publisher: Trakia University);
- Journal of IMAB: Annual Proceeding (Scientific Papers) (International Medical Association Bulgaria);
- Bulgarian Journal of Science and Education Policy (Publisher: University of Sofia);
- Acta Linguistica (Publisher: Eurasia Academic Publishers);
- Journal of Culture Collections (Publisher: National Bank for Industrial Microorganisms and Cell Cultures Bulgaria);
- Bibliosphere (Publisher: NBU Library);
- Venets : The Belogradchik Journal for Local History, Cultural Heritage and Folk Studies (Publisher: SCS Consulting Ltd.);
- Information & Security: An International Journal (Publisher: Procon Ltd.);
- Advances in Theoretical and Applied Mechanics, Advanced Studies in Biology, Nonlinear Analysis and Differential Equations, Applied Mathematical Sciences, International Journal of Algebra, Pure Mathematical Sciences, Advanced Studies in Medical Sciences, International Journal of Contemporary Mathematical Sciences, Contemporary Engineering Sciences, Advanced Studies in Theoretical Physics, International Mathematical Forum, International Journal of Mathematical Analysis, Advances in Applied Physics, and Environmental Sciences (Publisher: Hikari Ltd).

Providing open access to scholarly monographs is a growing movement and Directory of Open Access Books (DOAB, http://www.doabooks.org) lists 2151 academic peer-reviewed books from 70 publishers.

Providing open access to scholarly monographs is a growing movement and Directory of Open Access Books (DOAB, http://www.doabooks.org) lists 2151 academic peer-reviewed books from 70 publishers.

Open access repositories (or open access archives or open access digital repositories) contain research output, not only refereed journal articles, but also theses and dissertations, unpublished reports and working papers, conference and workshop papers, books, chapters and sections, multi-media and audio-visual material, learning objects, datasets, software, patents, etc. They might be institutional or thematic. When these repositories conform to standards created by the Open Archives Initiative they are interoperable, forming a global research facility. Common metadata protocol allows other web applications, such as data mining. Scholars and students deposit their research outputs in open repositories – a practice commonly called self-archiving.

Directory of open access repositories (OpenDOAR, http://opendoar.org/), lists 2706 repositories. It includes six repositories from Bulgaria: Academic Research Repository at the Burgas Free University (Научен портал на Бургаския свободен

университет, http://research.bfu.bg:8080/jspui/), DSpace at IMI (Bulgarian Digital Mathematics Library at IMI-BAS, http://sci-gems.math.bas.bg/jspui/), Bulgarian OpenAIRE Repository (IMI-BAS, http://www.bg-openaire.eu/), Electronic Repository – Central Medical Library – MU, New Bulgarian University Scholar Electronic Repository (http://eprints.nbu.bg/) and Research at Sofia University (http://research.uni-sofia.bg/).

Free and open source software (FOSS) is usually used to set up open access repositories and they benefit from free technical support provided by the large community of FOSS users. There are low installation and maintenance costs. Repositories could be set up quickly and institutions can immediately start gaining benefits. Institutions can also mandate open access, speeding development. ROARMAP: Registry of Open Access Repositories Mandatory Archiving Policies (http://roarmap.eprints.org/) lists 220 institutional open access mandates, 44 sub-institutional open access mandates, nine multi-institutional open access mandates, 90 funder open access mandates and 115 open access thesis mandates.

Open access publicizes institution's research strength, provides complete record of the research output in easily accessible form and new tools to manage institution's impact.

There are the following benefits of open access for researchers:

- **Immediate access** to research output for everyone upon official publication;
- More **visibility, usage and impact;**
- **Immediate impact** of the work;
- **Intensification of research** through fast dissemination and use of research;
- Possibly a **citation advantage** as well;
- **Monitoring** of research output;
- **Preservation** of research output by the library;
- **Keeping the rights** instead of signing them away;
- **New contacts and research partnerships**.

Open access also contributes to economic growth. Over 80% of the private sector is classified as small and medium size entrepreneurs (SMEs) in Europe and they play a key role in innovation. John Houghton, Alma Swan and Sheridan Brown provided quantifiable evidence to how much lack of open access costs SMEs in their study Access to research and technical information in Denmark (2011, http://eprints.ecs.soton.ac.uk/22603):

- "More than two-thirds reported **having difficulties accessing** market survey research and reports and Doctoral or Masters theses, 62% reported difficulties accessing technical reports from government agencies and 55% reported difficulties accessing research articles. Comparing responses on importance and ease of access, suggests that research articles and market survey research and reports are seen to be both important and difficult to access."
- "Respondents **wanted improved access** to research articles, market surveys and reports, patent information and scientific and technical standards. Fifty-nine per

cent of researchers wanted improved access to research articles, as did 47% of all respondents."

- "**Use of open access materials is widespread.** More than 50% used free institutional or subject repositories and open access journals monthly or more regularly, and among researchers 72% reported using free institutional or subject repositories and 56% open access journals monthly or more regularly."
- "**Both access and access difficulties involve costs.** The average time spent trying to access the last research article they had difficulty accessing was 51 minutes. Among researchers, the average time was 63 minutes. If around 60 minutes were characteristic for researchers, then in the current environment the time spent dealing with research article access difficulties might be costing around DKK 540 million (EUR 72 million) per year among specialist researchers in Denmark alone."
- "**Access to academic research brings substantial benefits for firms.** Twenty-seven per cent of the products and 19% of the processes developed or introduced during the last three years would have been delayed or abandoned without access to academic research. These new products contribute an average 46% of annual sales. Hence, the value of academic research to sales was equivalent to DKK 16 million (EUR 2.1 million) per firm per year, and the average value of cost savings was DKK 490 000 per firm per year."
- "**Access barriers and delays involve costs.** It would have taken an average of 2.2 years longer to develop or introduce the new products or processes in the absence of contributing academic research. For new products, a 2.2 years delay would cost around DKK 36 million (EUR 4.8 million) per firm in lost sales, and for new processes it would cost around DKK 211 000 per firm."

"Open Access to science and data = cash and economic bonanza" is the title of the speech that Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda, delivered at the Berlin 11 Open Access conference on November 19, 2013 (available at http://europa.eu/rapid/press-release_SPEECH-13-941_en.htm). And she continues: "But in fact it is only partly true. Because in reality, science has always been inherently open. Right from the start, even before the first learned societies or the earliest academic journals, the scientific community realised that it is through openness that they examine, compare and learn...Embracing this change is good for all of us: avoiding duplication while facilitating replication, accelerating discovery, and driving innovation."

2 Open Research Data

Ross Wilkinson, Research data enhancement manager at Australian National Data Service (ANDS) and Research Data Alliance (RDA) suggests that the distinction between open access publication and open research data should disappear, they are research findings".

And there are tremendous examples, such as Hubble telescope, that has an open archive for data, which led to significant increase in publications.

Policies ensuring that research data are openly available have been increasingly implemented by governments, funding agencies and journals. These policies are based on the idea that authors are poor stewards of their data, particularly over the long term. Timothy Vines, Arianne Albert, Rose Andrew, Florence Debarré, Dan Bock, Michelle Franklin, Kimberley Gilbert, Jean-Sébastien Moore, Sébastien Renaut and Diana J. Rennison in their article "The availability of research data declines rapidly with article age" (Current Biology , Volume 24 , Issue 1 , 94 – 97, http://dx.doi.org/10.1016/j.cub.2013.11.014) estimated how the availability of research data changes with time since publication. They "requested datasets from a relatively homogenous set of 516 articles published between 2 and 22 years ago, and found that availability of the data was strongly affected by article age. For papers where the authors gave the status of their data, the odds of a dataset being extant fell by 17% per year." In addition, the odds that they "could find a working email address for the first, last or corresponding author fell by 7% per year. Their results reinforce the notion that, in the long term, research data cannot be reliably preserved by individual researchers, and further demonstrate the urgent need for policies mandating data sharing via open repositories/archives." (http://arxiv.org/abs/1312.5670).

Ideally, all collected data should be available for further analysis. And about 80% of all data has been collecting dust in drawers or has been dying on hard disks.

The findings of a study, carried out by DAMVAD Norway, show that most Norwegian researchers are in favour of making research data accessible to other scientists. Nearly three of four researchers are willing to share their data. Eight of ten believe that open access to research data enhances research. The report can be downloaded here (in English with a summary in Norwegian).

Sarah Callaghan, a Senior Researcher and Project Manager for the British Atmospheric Data Centre, provided an overview of the reasons for and against formal data citation. Data citation and sharing is good for science and is also good for scientists. She lists the following reasons for data sharing:

1. "**Science is all about reproducibility** – if someone else can't reproduce your results, then your conclusions are invalid, and therefore the science doesn't work. **For a lot of scientific domains, reproducing results means using the original data collected, which means having access to it in the first place, which means sharing.**
2. **Data sharing cuts down on academic fraud.** It's hard work fabricating datasets (I know this from personal experience, having spent most of my PhD trying to simulate synthetic rain fields that looked anything like the real ones…), and having other people using your data means that they're more likely to notice if something seems a bit wrong (which is also useful for error corrections).
3. **Data sharing saves time and money.** If a dataset already exists to test your hypothesis, why spend the effort and the money to collect an entirely new one?
4. **Data sharing improves the transparency of the research process.** If the data's available to anyone who wants it, then you can't be accused of hiding evidence about a controversial topic (like climate change)." (Data Citation and Sharing: What's in it for me?:

http://blogs.lse.ac.uk/impactofsocialsciences/2014/01/07/data-citation-and-sharing-whats-in-it-for-me/)

Seth Long critically examines the practice of re-purposing data and finds data in the digital humanities beg to be re-purposed, taken from one context and transferred to another, opening up a wealth of opportunities for research ("Re-purposing" data in the Digital Humanities: Data beg to be taken from one context and transferred to another: http://blogs.lse.ac.uk/impactofsocialsciences/2014/04/02/re-purposing-data-in-the-digital-humanities/).

Christopher W. Belter examines three data sets archived at the National Oceanographic Data Center and his findings "suggest that all three data sets are highly cited, with estimated citation counts in most cases higher than 99% of all the journal articles published in Oceanography during the same years." (Belter CW (2014) Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets. PLoS ONE 9(3): e92590. doi:10.1371/journal.pone.0092590).

Public health crises such as the spread of drug-resistant tuberculosis highlight the need for improved sharing of data. For humanitarian organizations, there is a lack of guidance on the practical aspects of making such data available. In 2012 Médecins Sans Frontières (MSF) decided to adopt a data sharing policy for routinely collected clinical and research data. The MSF policy builds on the principles of ethical, equitable, and efficient data sharing to include aspects relevant for an international humanitarian organization, in particular concerning highly sensitive data (non-maleficence), benefit sharing (social benefit), and intellectual property (open access). Read more on how this policy was developed, the principles underlying it, and the practical measures taken to facilitate data sharing: Karunakara U (2013) Data Sharing in a Humanitarian Organization: The Experience of Médecins Sans Frontières. PLoS Med 10(12): e1001562. doi:10.1371/journal.pmed.1001562.

Sharing of data has leads to progress on Alzheimer's disease (see Gina Kolata's article in the New York Times: http://nyti.ms/I0JicA): "It was unbelievable," said Dr. John Q. Trojanowski, an Alzheimer's researcher at the University of Pennsylvania. "It's not science the way most of us have practiced it in our careers. But we all realized that we would never get biomarkers unless all of us parked our egos and intellectual-property noses outside the door and agreed that all of our data would be public immediately."

There is evidence that studies that make their data available do indeed receive more citations than similar studies that do not (Gleditsch, Metelits & Strand, 2003; Piwowar, Day & Fridsma, 2007; Ioannidis et al., 2009; Pienta, Alter & Lyle, 2010; Henneken & Accomazzi, 2011; Sears, 2011; Dorch, 2012). Heather A. Piwowar and Todd J. Vision in their article "Data reuse and the open data citation advantage" (2013, https://peerj.com/articles/175/) indicate a 10%–30% citation benefit that may be an effective motivator for data deposit.

It's a high time for us to stop treating data as comodities and start changing the behaviour of data owners who act as free traders on weekends and protectionists on weekdays. As Marco Fahmi, Queensland University of Technology, says "a lot of helicopter parenting is going on with data; let your data grow wings and fly away."

3 Open Science

The Polymath Project (http://polymathprojects.org/) – a collaboration among mathematicians to solve mathematical problems by coordinating many mathematicians to communicate with each other on finding the best route to the solution – is one of the most frequently mentioned examples of open science in action. The project began in January 2009 on Tim Gowers' blog when he posted a problem and asked his readers to post ideas for a solution. This experiment resulted in a new answer to a difficult problem, and since then the Polymath Project has grown to describe a particular process of using an online collaboration to solve any math problem (https://en.wikipedia.org/wiki/Polymath_Project).

In 2013 EIFL joined 26 organizations, including major sponsors Public Library of Science (PLOS), Wellcome Trust and Google, in the launch of the Accelerating Science Award Program (ASAP), which recognizes the use of scientific research, published through open access, that has led to innovations in any field that benefit society. Six teams of scientists whose innovative reuse of existing research enabled important advances in medical treatment and detection, ecology and science education have been shortlisted and three of them received the awards. These examples demonstrate how the reuse of open access research can accelerate scientific progress and benefit society as a whole.

For example, Dr. Nitika Pant Pai and colleagues (Caroline Vadnais, Roni Deli-Houssein and Sushmita Shivkumarare) have been widely recognized for their work on rapid point-of-care diagnostics for HIV and coinfections to curb the spread of HIV, particularly for marginalized populations in low- and middle-income countries, as well as high-income countries. Her team created a tailored smartphone application and HIV self-test that can save time and increase screening, counseling and treatment rates. “There is so much energy in giving knowledge and sharing knowledge,” explained Dr. Nitika Pant Pai. “Knowledge only grows by sharing. That is the beauty of open access. It will grow, it has grown and there is great momentum," she continued. “Being an award recipient will help shine light on the fact that open access acts like a catalyst – by allowing unrestricted knowledge sharing – it exponentiates the power of knowledge to transform and impact lives beyond borders, boundaries, languages, and regions; facilitates creation of novel innovations, improved practices and policies." To learn more about this project, visit: HIV Self-Test Empowers Patients.

Another award recipient, Matthew Todd, PhD, from the University of Sydney, highlighted how open access can bring scientists and researchers together to achieve a common goal. Todd created a global, collaborative Open Source Malaria Consortium that uses a crowdsourcing model to accelerate drug discovery to fight deadly malaria, by sharing and discussing open data with scientists around the world, in real time. Todd gave the audience an example of a student at the University of Edinburgh, Patrick Thompson, whom he had never met, but was working on his malaria project from Scotland. “I knew he was working on these molecules, and he was trying to make some candidate molecules for malaria," said Todd. "And I woke up one morning, and there was a picture of him on Twitter holding up this compound that he was about to send off to be tested to see if it would kill a parasite. It was just fantastic. It made me

excited about science in the way that I was as a kid." To learn more about this project, visit: Global Collaboration to Fight Malaria.

Many aspects critical to understanding science, experiments and the natural world are hard to convey using only words and diagrams. Good quality multimedia can help make that understanding easier. Daniel Mietchen, PhD, and his group (Raphael Wimmer and Nils Dagsson Moskopp) created the Open Access Media Importer (OAMI), a bot that can find and download supplementary multimedia files from reusably licensed open access research articles deposited in PubMed Central and uploads them to Wikimedia Commons, the media repository used by the Wikipedias and their sister projects.

To date, the bot has uploaded more than 17,219 files that are being used in more than 200 English Wikipedia articles and many more in other languages that together garner about three million monthly views. "We want people to play around with scientific materials and to engage with scientific processes," explained Dr. Daniel Mietchen. "Scientific research should play a more public role in our society, and open licenses greatly facilitate that. We are glad that the award highlights the value of reusing, revising, remixing and redistributing open access materials." To learn more about this project, visit: Visualizing Complex Science.

And you can also see more details about the ASAP Honorable Mentions: Calculating Ecotourism Impact, Measuring and Understanding the Sea, and Smartphone Becomes Microscope.

In 2014 the Finnish Ministry of Education and Culture established Open Science and Research Initiative to incorporate open science and research to the whole research process. This will help to improve the visibility and impact of science and research in the innovation system and society at large; and to foster the research system in Finland towards better competitiveness and higher quality, transparent, collaborative and inspirational research process. This national initiative promotes open access publications, open research data, open research methods and tools, as well as new skills and support services in open science domain (read the policy document here).

And I would like to end up with a quote from What, exactly, is Open Science? by Dan Gezelter: "Michael Faraday's advice to his junior colleague to: "Work. Finish. Publish." needs to be revised. It shouldn't be enough to publish a paper anymore. If we want open science to flourish, we should raise our expectations to: "Work. Finish. Publish. Release." That is, your research shouldn't be considered complete until the data and meta-data is put up on the web for other people to use, until the code is documented and released, and until the comments start coming in to your blog post announcing the paper. If our general expectations of what it means to complete a project are raised to this level, the scientific community will start doing these activities as a matter of course." (http://www.openscience.org/blog/?p=269)