"Hurling Against a Haystack"[1]: The Incentives and Challenges of Open Data in the Republic of Ireland

Brittney O'Neill, University College Dublin
Kalpana Shankar, University College Dublin

Abstract
Comparative research on data practices is hampered by the difficulty of studying across cases. The Republic of Ireland, with its small geographical size and population, an economy that is highly leveraged in information technology investment, and a centralized funding model provides a unique "laboratory" for examining data infrastructures in social and cultural contexts. This project reports on a preliminary study of the Open Data movement in the Irish public sector with an eye to surfacing themes for understanding data practices and challenges across different sectors. Using semi-structured interviews with individuals (n=11) involved in open data administration across the Republic, the researchers discuss current implementation and ongoing practices. Initial findings with respect to difficulties in measuring success, the sustainability of data, and valuating data are discussed. Future work on understanding how culture may play a role in open data infrastructures, stated and implicit values and biases, and creating and measuring need and impact are briefly discussed.

1   Introduction

A wide variety of institutions around the world have embraced the large-scale information systems that make digital data discoverable, shareable, and reusable. The availability of extremely large datasets (and the tools to analyze them) to tackle previously intractable problems - predicting flu outbreaks using Google searches (Ortiz et al., 2011), natural disaster mitigation (Zook et al., 2011), and automatic financial trading, to name a few – have been greeted with enthusiasm, but have proven in practice to have mixed effects. Such initiatives increasingly include government and publicly funded research data. Part of the "Open Data" movement, these initiatives take numerous forms and implementations, but the core idea motivating many of these projects is that certain publicly funded and acquired data should be made available in forms that can be reused and analyzed by anyone with the tools and expertise to do so. With respect to government data in particular, the rationale is two-fold. The first is making the administrative work of the public sector accountable and transparent over time. The other is leveraging public data for the development of new applications and tools (for economic growth) by both the private and public sectors.

While petabytes of data are abstract, the wires, servers, machines, and networks upon which they depend still are not. We have an imperative to understand how and what values are being inscribed in both the ephemeral (the data sets) and the obdurate (the physical infrastructures) - because the data will disappear or become "black boxed" without our scrutiny or will force choices upon us that may be too difficult to turn back from (Winner, 1993). Furthermore, abstraction of data without attention to the institutions and networks underpinning them obscures underlying but far-reaching questions of what is remembered, what is forgotten, and who decides (Blanchette, 2011). In a recent call for proposals for research on open data, the Berkeley Center for Law and Technology at the University of California and Microsoft noted that, "While open government is often presented as an unqualified good, sometimes Open Data can identify individuals or groups, leading to a more transparent citizenry... Hence, open data sets may have a disparate impact on certain groups."

---

[1] Hurling is an ancient Gaelic team sport that still remains extraordinarily popular in Ireland today. It involves using a "hurley", a wooden stick, to hit a small ball, a "sliotar", between the opponents' goalposts. One of the interviewees for the study expressed his frustration with the challenges of open data in Ireland as "hurling against a haystack" (a futile gesture). This evocative phrase expressed for the researchers the highly cultural and situated nature of data practices that is difficult to articulate.

Research on the institutional, cultural, and ethical dimensions of open data is clearly in its infancy. While individual open data initiatives can be examined, one of the most-often expressed frustrations of the field of information studies is the inability to conduct rich comparative studies of data practices and infrastructures (Beaulieu, Scharnhorst, & Wouters, 2007). Comparative study, for reasons of resources, access, methodologies, and even good "questions" has been difficult. As a result, most studies of data projects have been deep rather than wide, focusing on one or two specific projects or initiatives.

This paper describes a research project in progress: an attempt to study ACROSS open data projects and initiatives to explore how the specifics of the public service, economic conditions, and other social and cultural specifics play out in open data. This work in progress is set in the Republic of Ireland to surface questions, themes, and potential methods that can be used for comparative work. Going beyond the technical implementations, the project is exploring how stakeholders perceive open data, incentives and barriers to implementation, and concerns and challenges about sustainability and reuse. Ireland provides an unusual laboratory for such a study. Like many other countries, the Republic of Ireland has been looking to open data and large-scale data infrastructures in health, science, cultural heritage, and government data to address research and policy questions and spur technological innovation. Additionally, its small geographical size and population, centralized public funding sources, membership in the European Union, and significant investments in digital data (and a plethora of information technology-oriented multinationals, small/medium sized enterprises, and startups) as economic drivers make it ideal for a cross-disciplinary study.

In this paper, we discuss some emergent themes that cut across open data projects in Ireland. Based significantly on interviews with researchers and administrators and policy and public-facing documents, we discuss the following:

a)   How are success and failure in Open Data defined?
b)   How do Open Data initiatives in Ireland maintain themselves over time?
c)   How do stakeholders assign value to Open Data?

2   Background

The premises of the Open Data movement are that the widespread availability of appropriate platforms, data standards, and tools (many, if not all, open source) can facilitate the sharing of digital data of all forms. In large part thanks to the energetic activism of Tim Berners-Lee (2009; 2010), open data has become a focal point for public sector policy around the globe. The arguments underpinning the need for open data rest on transparency, accountability, and economic and other activity that can be enhanced by the use and reuse of expensively collected data, much of it publicly funded. Within the European Union, the Horizon 2020 research initiative promotes open access to publications and data as way to ensure public return on public investment as well as a form of economic stimulus. However, the Horizon 2020 initiative offers little advice or standards for the implementation of open access systems on the ground level.

The uneven distribution of emancipatory rhetoric versus ground level research is also borne out in the global literature on open data. Much of this work, like Tim Berners-Lee's TED talks (Berners-Lee, 2009; 2010), are couched as "calls to action" (e.g. Borgman's (2009) The Digital Future is Now: A Call to Action for the Humanities), essays motivating commitment to open data (e.g. Molloy, 2011; Taylor, 2014). These papers discuss potential issues including motivating researchers to contribute data (e.g. Beck & Neylon, 2012), effective licensing (e.g. Hrynaszkiewicz & Cockerill, 2012; Hrynaszkiewicz, Busch, & Cockerill, 2013; Bloom, Ganley, & Winker, 2014) and ways to make data reusable (e.g. Beale, 2012; Molloy, 2011) but their focus is on presenting the anticipated benefits of open data such as community involvement (e.g. Beale, 2012, Davies & Edwards, 2012), economic stimulation and innovation (e.g. Wolkovich, Regetz, & O'Connor, 2012), and accountability to public funders (e.g. Beale, 2012; Beck & Neylon, 2012).

Beyond the global assessments of the Open Data Barometer (Web Foundation, 2013) comparatively few researchers have assessed the impact and results of open data once it has been implemented. For example, Moyes et al. (2013) built and assessed an open data infrastructure for providing open malarial research data. Ultimately, however, they provide only a descriptive assessment of their interface without any indication of actual use or impact. In contrast, by exploring the data.gov interface in the USA, Peled (2011) explores inter-departmental dynamics in conjunction with download and use statistics to find that though departments complied with the minimal requirements of the project, the data produced was not optimal quality, much data was coveted for use in political trades, and despite use figures, departments were not motivated to continue uploading data or to release any data beyond

the minimal requirements. This conflict between meeting minimal compliance and actually fully committing to the initiative is likely present in most similar initiatives such as the United States' data.gov or the European INSPIRE directive, which mandate minimal compliance rather than measured impact (Commission Decision of 5 June 2009 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards monitoring and reporting [2001] OJ L148/18). As of now, few interfaces or data interfaces are ready to be thus assessed, and no critical research seems forthcoming on the topic.

In efforts to better understand the continued struggles to create a research culture of open data, research has been conducted to assess the state of open data through interviews with active stakeholders in the open access/open data environment (particularly scholarly research, where the bulk of research has been conducted). As early as 2007, Fry, Schroeder, and den Besten (2009) interviewed relevant stakeholders in the UK and found that the greatest barrier to open data adoption was fragmentation within the larger research environment rather than issues of intellectual property rights as had been previously assumed. More recently in the UK, the RECODE project has conducted interviews, surveys, and focus groups to assess both stakeholder values and ecosystems (RECODE, 2013) and infrastructure and technology challenges (RECODE, 2014). This research, though finding strong philosophical commitment to openness as a way of supporting innovation and providing public accountability, also found concerns about different values and needs across research fields, dangers of scooping and lack of intellectual property rights, financial, legal and cultural challenges to adoption. In the Netherlands, Janssen, Charalabidis, and Zuiderwijk (2012), conducted similar research in group sessions where they found strong support for open data as an economic driver, but also found representation of a range of myths including the widely held myth inherent to much prospective research that, "if you build it, they will come". Gurstein (2011) and others (e.g. Janssen et al., 2012) have cautioned that these mythologies need to be examined more closely: increased transparency of the public sector is a hoped-for and positive outcome of computing and open source technologies, but open data has the potential to engender new kinds of "digital divides".

2.1   Open Data in Ireland

As noted earlier, the European Union's Horizon 2020 research initiative promotes open access to publications and data as way to ensure public return on public investment as well as a form of economic stimulus. The EU commission has also produced the INSPIRE directive that provides guidelines for the achievement of a more concrete federated network of open environmental data across the EU. Though not mandating the specific model of enactment within each country, the INSPIRE directive provides standards for interoperability, technological infrastructures & interfaces, and metadata with associated deadlines for compliance.

Though both EU initiatives apply to all member countries, within the member countries conditions vary widely, from the global leader, the UK, to comparatively nascent initiatives such as in Greece (Open Data Institute & World Wide Web Foundation, 2013). Ranked only ten places above Greece in 34th place, Ireland is reported to have a high level of Readiness in Civil Society but lacks in other measures of readiness, has low scores for datasets, and is lacking in open data impacts particularly in the social sphere (ibid.). Though Ireland has made efforts to move from policy (readiness) to practice (uploaded datasets and assessing impact) in the past year since the publication of the 2013 Open Data Barometer, most action seems to still be at the level of planning and policy in terms of both the Open Knowledge Foundation's activism (openknowledge.ie), and the Open Government Partnership (OGP) program (ogpireland.ie) which the Irish government is currently undertaking.

Ireland's involvement with OGP had until mid-2014 appeared to be limited to policy and planning, but as of May 2014 an audit on open data was produced (Cygniak, 2014) and a national government data portal was launched. The portal received little public discussion aside from one national newspaper, the Irish Times', scathing accusations of "screen scraping" and unusable data (Worrall, 2014). Though the guiding National Action Plan was developed by committee, the results of the new data portal's launch suggest that, as predicted by the Open Data Barometer, Ireland's civil society is ready for open data, but business and government are lagging behind in readiness, and the actual infrastructures are still somewhat lacking. In fact, most current policy in Ireland has been informed not by consultation with actors in open data or thorough empirical research, but instead by the discussion of committees composed of government departmental representatives and self-nominated volunteers (e.g. National Action Plan for Open Government, Open Data Ireland: A Briefing Paper, Open Data Ireland: Open Data Audit). As such there is a notable gap in the Irish context between those delivering prospective policy and those actually producing and managing the data on the ground, made painfully obvious by the failure of the committee-driven data.gov.ie vis-à-vis the more successful Irish Social Science Data Archive

(ISSDA), the Irish Spatial Data Exchange (ISDE), and Statcentral, all of which were developed by individuals already engaging with data as researchers or statisticians.

3   Methodology

In order to orient the research project within the Irish context, the study began with the mapping of stakeholders in Open Data within Ireland. In contrast to the Open Data Ireland Audit (2014) discussed earlier, the researchers opted to use manual techniques rather than crawlers to identify relevant bodies.

The first phase of mapping began with broad search queries such as "open data in Ireland" and then refining those queries in response to search results. Emails and phone calls to relevant individuals refined the searches. Though all relevant bodies, including policy makers and government ministries were included, emphasis was placed on stakeholders that were directly interacting with open data, either as researchers or data managers. Other stakeholders were also identified by following links to open data partners and through word of mouth discussion with other academics and professionals interested in data initiatives.

Basic summaries were derived for all policy bodies and repositories and data publishers were studied to map the following information about relevant repositories:

- a) State of "openness"
- b) Metadata standards
- c) Form of data presentation (including file type if relevant)
- d) Sources of data
- e) Funders/supporters/partners of the interface
- f) Stated stance on reuse/citation/licensing
- g) Access policy/restrictions
- h) Information for data producers

Stakeholders were then mapped by sector as seen in Figure 1 and a network map as shown in Figure 2 was created by associating Irish stakeholders with those that they support or partner with both within Ireland and the EU more generally. In the creation of the network map, only those with more than one relationship were mapped.
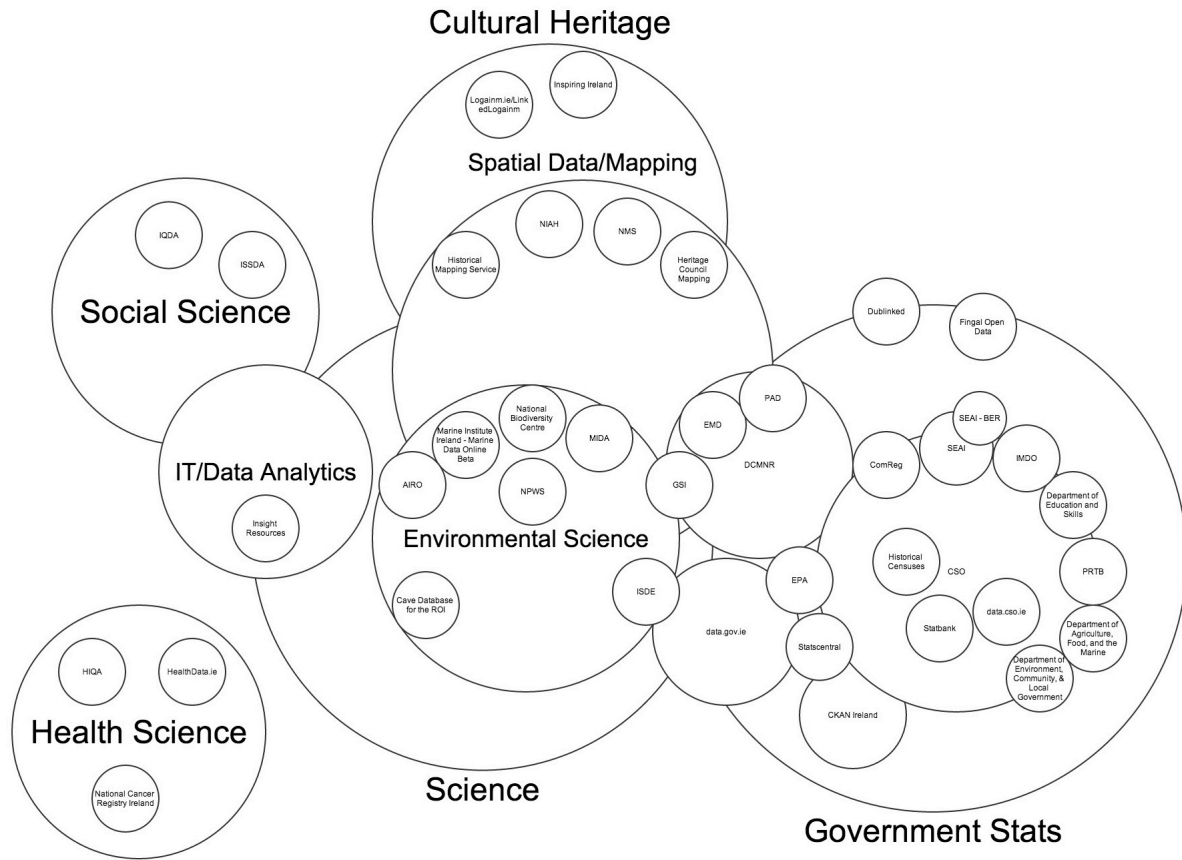
Cultural Heritage

Logainm.ie/Link edLogainm

Inspiring Ireland

Spatial Data/Mapping

NIAH

NMS

Historical Mapping Service

Heritage Council Mapping

IQDA

ISSDA

Social Science

Dublinked

Fingal Open Data

IT/Data Analytics

National Biodiversity Centre

Marine Institute Ireland - Marine Data Online Beta

MIDA

PAD

EMD

SEAI - BER

SEAI

IMDO

AIRO

NPWS

GSI

DCMNR

ComReg

Department of Education and Skills

Insight Resources

Environmental Science

Historical Censuses

CSO

PRTB

ISDE

EPA

Statbank

data.cso.ie

Department of Agriculture, Food, and the Marine

Cave Database for the ROI

data.gov.ie

Statscentral

Department of Environment, Community, & Local Government

HIQA

HealthData.ie

Health Science

CKAN Ireland

National Cancer Registry Ireland

Science

Government Stats

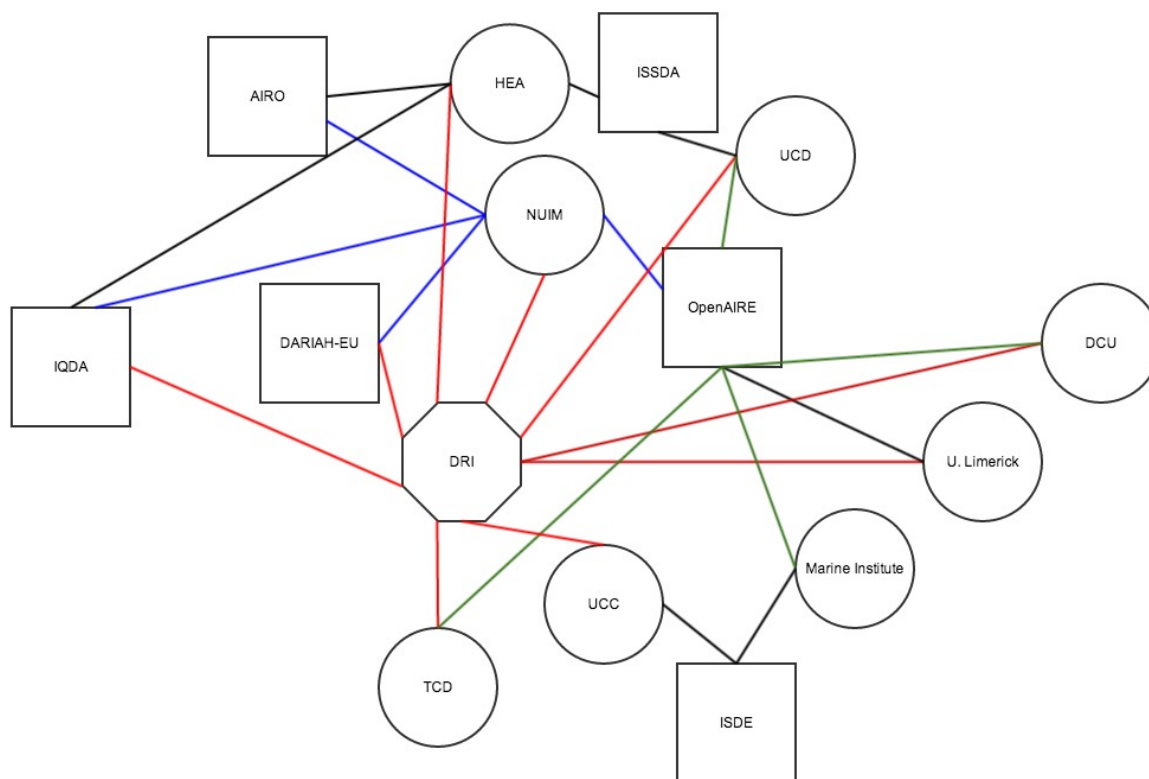Figure 1. Irish open data repository/discipline map.

5

Figure 2. Open data in Ireland institutional network map.

Following this process of mapping, stakeholder websites were further investigated to find relevant contact information. Contacts were sought across sectors from the most active, largest, or most elaborated data repositories in Health, Social Science and the Humanities, Science, and Government.

Interviews
The next stage was to conduct semi-structured interviews to explore the facilitators, barriers, and general ethos of open data in Ireland. Participants were identified as described above through contacts gleaned from repository websites. The mapping suggested that in the Irish context, scientific research is often conducted under the aegis of government agencies and we found in practice that a significant number of individuals often had multiple roles: as a researcher in a university, for example, as well as a founder of an open data project or founder of a start-up company that used open data in some way. To date, eleven individuals from eleven distinct open data projects have been interviewed. Five of these individuals are academics with direct responsibility for data management projects; six are government researchers or employees of local or national government and administrative entities.

The project researchers contacted all individuals through an email in which the research project was described, the researchers' reasons for contacting that individual were stated, and an interview was requested. In some cases, as roles had evolved, those contacted suggested other relevant individuals and we followed through with those individuals as well. Interviews were scheduled as per the interviewee's preference: in person or by Skype. A list of interview themes and the informed consent form was emailed several days prior to the interview. Each interview was recorded; one researcher took notes during the interview while the other conducted the interview. All interviews took approximately 45 minutes to one hour. The researchers debriefed with each other after each interview to discuss any themes that emerged or any issues that could be resolved by an email.

The interview questions were adapted from instruments developed by the European Union funded RECODE project (http://recodeproject.eu/research/). Questions were drawn from both RECODE (2013) *Deliverable D1: Stakeholder Values and Ecosystem* and RECODE (2014) *Deliverable D2.1: Infrastructure and Technology Challenges* and, depending upon the participant, questions for either managers or senior academics were used. Since measures of success were of interest to the researchers and few Irish repositories have published clear goals or statement of how they planned to evaluate their

success, several additional questions were also included to explore goal orientation and measures of success. The key themes of the interview were as follows:

    a) Open Data Policies
    b) Open Data Program Goals & Motivations
    c) Barriers to Open Data
    d) Enablers of Open Data
    e) The Value of Open Data
    f) The State of Open Data in Your Field
    g) Practicalities of Data Access
    h) Funding
    i) Measures of Success
    j) Data Management Plans
    k) Reliability & Validity
    l) Accessing and Using Open Data

4 Preliminary Analysis

The researchers have used this general list of themes to classify the answers resulting from the interviews with researchers and administrators of open data (we noted that in Ireland, these individuals were often the same, with research responsibilities that were somewhat separate from their open data commitments). The interview questions and themes have provided a straightforward top-down coding scheme to address some of the initial questions, but emergent coding of the interviews, terms of use and policy documents, and secondary literature is underway as well. In the next sections, we discuss some of our emergent themes of interest.

4.1 Measures of Success

As the INSPIRE directive notes, initial "success" in open data is generally measured by compliance with a specific set of infrastructural and data contribution goals in getting data into a system. However, based on our interviews, few organizations have developed other concrete goals to strive for. When scientific researchers were interviewed, many noted that that journals and conferences, particularly those in computer science, are beginning to demand data publication alongside participation and article publication. Some funding bodies are also reported to favor those who share data or demand the development of data management plans, but for those outside of the domain of the INSPIRE directive, measures of input success do not seem to be of great importance. Most interviewees cited the acquisition of data as a significant challenge to "input success". They cited technical reasons such as incompatible data formats, uncertainty about the utility of data, and the mismatch of data description standards. However, organizational challenges were also cited. These included data depositors' concerns about the labor needed to make data useful, worries about the potential for misuse or unintended use (what one interviewee called "dual use weaponry": data collected for one purpose and used for a different purpose), and general lack of knowledge of open data.

    Once data is acquired, interviewees identified the reuse of open data as the significant primary goal of open data whether for purposes of research replicability, economic innovation, or evidence-based planning and policy-making. Reuse, however, still lacks a metric. Many bodies collect analytics such as visit and download statistics from their repository, but interviewees frequently stated that these were not adequate nor representative measures of "real world" reuse. All were hard-pressed to indicate what would be representative or useful since user views or downloads do not mean that users have employed data in any further policy, economic, or research activity.

    Citation tracking was also discussed in the academic research environment as an important metric. However, citation tracking suffers both from the lack of a systematic method of tracking citations across wide ranging publications, and from the lack of citation norms in commercial ventures. Though Piwowar, Day, and Fridsma (2007) and Henneken and Accomazzi (2011) both found that sharing data along with open access publications increases rates of citation, outside of formal research projects, citation tracking is time-consuming and impractical. Academics include citing in their daily practice, but for commercial entities, there is no established practice of citing sources of free and open data or information. With the rare exception of anecdotes from entrepreneurs who chose to report back to data providers on data reuse, interviewees were unable to provide clear evidence of reuse.

    One interviewee discussed the notion of "brand impact" as a measure of success for data initiatives, citing international recognition and positive reputation within the field of open data as evidence of success. This interviewee referred to the ongoing development of his particular open data portal, a

highly publicized data release initiative for economic impact, as an important "brand" for Ireland in open data communities. Again, however, "brand recognition" lacks a clear and repeatable metric and appears to be employed "by feel" rather than by any other measure.

Thus, there are few qualitative measure of success and even fewer quantitative metrics by which one can measure or demonstrate the success of open data projects; open data advocates and managers tend to be limited to anecdotal evidence and invocations of the potential power of open data to insure ongoing support for their programs. Several organizations' administrators, however, argued that the lack of metrics are not necessarily relevant at this point in Ireland's open data movement. Many of the repositories are still in their formative stages and so, as one interviewee put it, there is "an element of just needing to agree with it and measure afterwards" without necessarily defining the metric beforehand. This notion may be related to the close connection between open data and computer science more generally where an ethos, of developing the system and then waiting to see what happens has been the driving force behind significant web and other technology innovations.

4.2   Sustainability

Invariably, the question of "how do you intend to make this project sustainable?" elicited a response focused on funding. Most Irish and European funding programs emphasize short-term research without funding for long-term data management and curation; as a result, long-term funding was identified as a major challenge for open data repositories. One repository manager spoke of patching together small grants to pay interns to keep his data updated and live after EU funding ended. Two interviewees noted that they had access to reliable funding by making their open data projects part of the regular workflow of their organizations, but still expressed concerns. Directly related to issues of funding, orphaned data sets also emerged as victims of an unsustainable system. Interviewees in the research sector discussed the data sets left without maintenance or support at the end of funding cycles, while those from governmental bodies encountered datasets that were orphaned within the governmental system and so were difficult to reuse within an open repository as no one knew who could "sign off" on the data. The presence of orphaned data, especially within government suggests a lack of long term data management planning even before data is opened to the public. One manager of a local government authority data project noted that his agency had integrated open data into their knowledge platform workflow, but admitted that if a new executive decided that open data was no longer a top priority, the flow of data would cease (and the data would be orphaned).

In all but two interviews, issues of technical data sustainability (servers, standards, etc.) did not emerge but insuring data availability over time did arise as a significant challenge for open data. Two interviewees, both with professional and educational backgrounds in library and information studies, discussed the issues of file compatibility and the importance of metadata and coding to the longevity of data sets without being asked about the topics. Several interviewees also connected sustainability to the timeliness and up-to-dateness of data sets by arguing that making high value data available was likely to result in more sustainable repositories. Static data was generally perceived by most of the interviewees to have lower potential for reuse than dynamic data (even if dynamic data was only updated annually), but static data was accepted as being "low hanging fruit" that was easy to obtain, easy to curate, and requiring less maintenance. At the other end of the spectrum, highly dynamic data that is changing minute to minute (such as marine research data) provides challenges not only in terms of the cost of providing and presenting the data (it is often not archived) but also in terms of maintaining that level of dynamism. Interviewees indicated that data is "as good as useless if it's out of date" and thus, maintaining the accuracy and timeliness of the very data points themselves is a key element of open data sustainability. Out of date data will likely not get reused.

4.3   The Value of Data

That data is somehow valuable underpins the very premise of the open data movement, but the exact nature of that value is unclear. Depending upon domain, data is perceived of as a public or commercial good and its value seems to depend highly upon both its potential for economic and social returns and the cost of its production. All data are not created equal.

Though much of the official open data rhetoric hinges on the perception of open data as a public good, several interviewees from the academic research environment discussed the tendency of individual researchers to perceive their research data as an extension of their own personal value as researchers and argued that the data was their stock in trade – an extension of a "brand" as a researcher. Others from both research and governmental organizations discussed the most valuable data sets as those held and leveraged by private corporations (because of scarcity, volume, and lack of general availability). This

discussion of privately held (and highly coveted) data suggests an inherent value to data that is not necessarily revealed only through mass reuse. The question of whether data ought to be released publicly by default or leveraged for commercial returns was answered in many different ways depending upon the interviewee. For some, publicly funded data should be freely shared because it was paid for by the public and therefore should be made public whereas for other interviewees, the expense of public funding meant that the data ought to be marketed commercially in order to recoup on the initial investment.

Of course, most of the interviewees argued for open data on the grounds that it was just the "right" thing to do. Somewhat surprisingly, transparency arose as a motivator less often in government data than in government funded research where open data was seen as a tangible return on public investment in research. In the academic arena, transparency as a motivator is also evident in several interviewees' mention of research replicability. By opening up data, it was argued that the research process becomes more transparent, so it is easier for other researchers to verify results and hold the initial researcher accountable for the validity and accuracy of his/her data. One repository manager however, countered this notion by arguing that transparency implies calling people out on their errors and so discourages participation (both in research and government administration) while a purely economic argument (that data could make money) provides much more successful motivation to make data open.

Within an academic research context, transparency and economics are closely linked as motivators. Making data open suggests that research monies are being well used and provide a tangible return on the public's economic investment. But economics are not only prominent in terms of showing return on investment. Multiple interviewees both from government and environmental data sectors expressed faith that open data would someday yield economic benefits in terms of commercial innovation. Though unable to provide specific measures of economic stimulus driven by open data, interviewees were confident that the benefit was real and expressed the benefit as a powerful motivator not only for their own work but also for eliciting contributions from researchers, research organizations and government bodies.

Economic factors were also cited in discussions of preventing costly duplication of research work and in calculations of success as reduction in service costs. Several interviewees working with government agencies discussed measurable decreases in telephone inquiries over recent years. One interviewee reported savings of up to 26 person days per year in the decrease of statistical inquiries since the launching of an open data repository where users could request and receive data on their own.

5   Conclusion

We began this project by focusing on whether issues around the implementation and use of open data could be examined closely by investigating extant projects and initiatives in one country, Ireland, across different sectors.   We acknowledge that the range of projects we chose to study (cultural heritage, health, sciences, etc.) limits generalizations with respect to content and users. Instead, our goal was to explore the commonalities and differences of open data discourse that are present in a single cultural and social moment, one in which the term "open data" is being used across disciplines.

Our initial results, suggest that in spite of enthusiasm and funding, whether a "data community" exists in Ireland (or could be developed) remains to be seen. Several interviewees admitted that the availability of open data to relevant stakeholders had not actually increased the uptake of such data for relevant policymaking or planning (although the natural science repositories reported a more enthusiastic user base). Because Ireland has a smaller population that most European countries of similar economic and technological status, perhaps there are simply fewer people interested in using open data to create applications, or to provide goods and services. The perceived lack of expertise in the government sector with respect to reusing open data for policy-making and service evaluation may argue that external consultants can be hired to implement open data systems, but since these consultants are not the target users, open government data may not be reused by those who are creating it. Though perhaps less relevant in the academic research environment, where data sharing is becoming more commonly accepted as part of the research and publication process (if not always executed in practice), the lack of users of government data was identified as a potential threat to the sustainability of repositories that rely on promises of economic benefit to motivate funders. Even if open data is being reused by entrepreneurs or other start-ups, it is hard to show direct evidence of its role as an economic stimulus.

If web analytics such as download and visit statistics are not accurate representations of data reuse and citations are unreliably employed (especially outside of academia), then how can stakeholders track the impact of their open data programs? One repository requires formal requests for data access which allows them to track precisely how their data is used, but most other interviewees felt that any

barrier at all to accessing data would lead to an overall reduction in use. In fact one interviewee even expressed a belief that demanding that users report on use of data was fundamentally counter to open data efforts.

Traditional results-based research and project planning tend to rely on specific measurable results, but open data (as with so many massive digital phenomena) may in fact prove too elusive for concrete impact measurement. Nonetheless, the arena warrants further exploration to justify the extensive outflow of time and resources required to develop and maintain open data. If there is no reliable way to measure reuse of data, then perhaps there are other useful metrics that can be developed or available, or perhaps, as one interviewee pointed out, we just have to "agree with it and measure afterwards."

As we suggested above, the value of data remains currently somewhat under-defined. It remains unclear whether data should be assigned value on the basis of cost or of potential social benefit and it seems to differ substantially from individual to individual, even within projects. With the current proliferation of digital data and limited resources for curation, it is critical to identify the most valuable datasets to prioritize for collection, processing, and access provision – but how? Additionally, the manner in which value is constructed in open data will determine how it expands and is managed into the future which will in turn impact the overall trajectory of open data in society, research, governance, and business. By understanding how data creates value researchers and policy makers can shape the future of data in society and drive comparatively more or less successful data systems that deliver on their promises (with the caveat that we still don't know what "success" means).

Lastly, we know from deep studies of technology and society that it is likely that here, too, culture and society will play a significant role in how open data (and other large-scale digital initiatives) develop and evolve over time. Interviewees regularly expressed an opinion that needs further unpacking: that the structure of Irish research funding, its public sector, its civil service, and most telling, a generic appeal to "the way things are done in Ireland" created negative repercussions for open sharing of information, the use of evidence for effective policy making and planning, and the basic understanding of open government data. This in turn will raise many questions about what counts as "open" and for whom and what unexpected negative repercussions may redound to particular groups or citizens (and how open data repositories "embed" these biases or counter them). The potential for open data to reconfigure the nature of participation in governance and citizenship, with the potential for impact to numerous civil liberties, requires further exploration. As Davies and Bawa (2012) put it, "[W]hat we find in society today is not only various practices relating to open data, but also an active shift in paradigms about access and use of information and data, and notions of 'openness' and 'information/data'.

References

Beale, N. (2012). How community archaeology can make use of open data to achieve further its objectives. *World Archeology, 44*(4), 612–633.

Beaulieu, A., Scharnhorst, A., & Wouters, P. (2007). Not another case study: A middle-range interrogation of ethnographic case studies in the exploration of e-science. *Science, Technology & Human Values*, *32*(6), 672-692.

Beck, A., & Neylon, C. (2012). A vision for Open Archaeology. *World Archaeology, 44*(4), 479–497.

Bloom, T., Ganley, E., & Winker, M. (2014). Data access for the open access literature: PLOS's data policy. *PLoS Biology*, *12*(2), e1001797. doi:10.1371/journal.pbio.1001797

Berners-Lee, T. (2009, February). Tim Berners-Lee: The next Web [Video file]. Retrieved from http://www.ted.com/talks/tim_berners_lee_on_the_next_web

Berners-Lee, T. (2010, March). Tim Berners-Lee: The year open data went worldwide [Video file]. Retrieved from http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide

Blanchette, J. F. (2011). A material history of bits. *Journal of the American Society for Information Science and Technology, 62*(6), 1042-1057.

Borgman, C. L. (2009). The digital future is now: A call to action for the digital humanities. *Digital Humanities Quarterly, 3*(4). Retrieved from http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Cygniak, R., Kamdar, M., Maali, F., Lee, D., & Decker, S. (2014). *Open Data Ireland: Data audit report.* Galway: Insight Centre for Data Analytics.

Davies, T., & Edwards, D. (2012). Emerging implications of open and linked data for knowledge sharing in development. *IDS Bulletin*, *43*(5), 117–127. doi:10.1111/j.1759-5436.2012.00372.x

Davies, T. G., & Bawa, Z. A. (2012). The promises and perils of open government data (OGD). *The Journal of Community Informatics*, *8*(2).

Fry, J., Schroeder, R., & Besten, M. Den. (2009). Open science in e-science: contingency or policy? *Journal of Documentation*, *65*(1), 6–32. doi:10.1108/00220410910926103

Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone?. *First Monday*, *16*(2).

Henneken, E. A., & Accomazzi, A. (2011). Linking to data: Effect on citation Rates in astronomy. *ASP Conference Series*, 1–5.

Hrynaszkiewicz, I., Busch, S., & Cockerill, M. J. (2013). Licensing the future: Report on BioMed Central's public consultation on open data in peer-reviewed journals. *BMC Research Notes*, *6*, 318. doi:10.1186/1756-0500-6-318

Hrynaszkiewicz, I., & Cockerill, M. J. (2012). Open by default: A proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals. *BMC Research Notes*, *5*, 494. doi:10.1186/1756-0500-5-494

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. doi:10.1080/10580530.2012.716740

Molloy, J. C. (2011). The Open Knowledge Foundation: Open data means better science. *PLoS Biology*, *9*(12), e1001195. doi:10.1371/journal.pbio.1001195

Moyes, C. L., Temperley, W. H., Henry, A. J., Burgert, C. R., & Hay, S. I. (2013). Providing open access data online to advance malaria research and control. *Malaria Journal*, *12*, 161. doi:10.1186/1475-2875-12-161

Open Data Institute & World Wide Web Foundation. (2013). *Open data barometer: 2013 global report.* Retrieved from http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf

Ortiz, J. R., Zhou, H., Shay, D. K., Neuzil, K. M., Fowlkes, A. L., & Goss, C. H. (2011). Monitoring influenza activity in the United States: A comparison of traditional surveillance systems with Google Flu Trends. *PloS one*, *6*(4), e18687.

Peled, A. (2011). When transparency and collaboration collide : The USA open data program. *Journal of the American Society for Information Science and Technology, 62*(11), 2085–2094. doi:10.1002/asi

Piwowar, H. a, Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PloS One*, *2*(3), e308. doi:10.1371/journal.pone.0000308

RECODE. (2013). *Deliverable D1: Stakeholder values and ecosystems.* Retrieved from http://recodeproject.eu/wp-content/uploads/2013/10/RECODE_D1-Stakeholder-values-and-ecosystems_Sept2013.pdf

RECODE. (2014). *Deliverable D2.1: Infrastructure and technology challenges.* Retrieved from http://recodeproject.eu/wp-content/uploads/2014/04/D2.1-Infrastructure-and-technology-challenges.pdf

Taylor, P. (2014, February 25). Government transparency means all-ish data all the time. *Govtech.* Retrieved from http://www.govtech.com/data/Government-Transparency-Means-All-ish-Data-All-the-Time.html

Winner, L. (1993). Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology. *Science, Technology, and Human Values,* 362-378.

Wolkovich, E. M., Regetz, J., & O'Connor, M. I. (2012). Advances in global change research require open science by individual researchers. *Global Change Biology*, *18*(7), 2102–2110. doi:10.1111/j.1365-2486.2012.02693.x

Worrall, J. J. (2014, July 25). Government's open data portal at risk of becoming a data dump. *Irish Times.* Retrieved from http://www.irishtimes.com/business/sectors/technology/government-s-open-data-portal-at-risk-of-becoming-a-data-dump-1.1877026

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical & Health Policy*, *2*(2), 7-33.

Table of Figures