

## Constructions that Fascinate: Ireland, Open Data, and the Data Imaginary

Kalpna Shankar, School of Information and Communication Studies, University College Dublin

### Abstract

In large part thanks to the energetic activism of Tim Berners-Lee, 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. The premises of the Open Government 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. We also know from studies of technology and society that it is likely that culture and society will play a significant role in how open data (and other large-scale digital initiatives) develop and evolve over time. Complex technologies also embed within them conceptions and communications of visions of the future: what Jasanoff (2015) call the “sociotechnical imaginary”. Actors ‘order the world’ through their constructions of technologies: platforms, standards, and policies.

### Introduction

The rhetoric of the Information Society, a societal transformation argued for by economists, sociologists, and others, is one in which the immaterial - labor, products, and movements - are privileged over their industrial counterparts in knowledge production and economic activity (Webster, 2007), with emancipatory effects to follow (Srinivasan, 2006). Several decades of critique suggest that this approach to understanding the interaction of digital information and social transformation is simplistic at best. Instead, what this approach represents is what might be termed the "digital imaginary". The imaginary, first articulated as a concept in anthropology (Anderson, 1991) but with multiple and shifting meanings, provides an overarching theoretical framework to examine digital data and their meanings. Nardi and Kow (2010) theorize the digital imaginary as "social constructions consisting of a set of cultural notions, predicaments, and anxieties expressed in, and circulated by, digital media". They suggest the use of the term "imaginary" suggests a sense of wonder, optimism, and perhaps illusion.

The ephemerality, volume, and pervasiveness of the digital have demanded new mechanisms and philosophies of curation to encompass new technologies, institutions, and social formations (Havila, 2008; Anderson and Allen, 2009). Decisions must be made about which information is to be preserved and in what formats, where it will be kept, how it will be preserved and presented – anticipating what will be wanted in the future (Churchill and Ubois, 2008). What is archived must be retrievable, so indexing and other representational practices must be in place. These purportedly technical decisions are at their heart seemingly objective but historically and culturally situated decisions and determine how our memories are structured, represented, and therefore, what we remember and how (van House and Churchill, 2008).

Democracies around the world have embraced the large-scale information systems that make public administration data collection possible and reusable through what has come to be termed the Open Government Data (OGD) movement. The idea motivating this movement is that certain publicly funded and acquired data - administrative, government, publicly funded research, to name a few - should be made broadly available. With respect to government data in particular, the rationale is two-fold. The first is that making the administrative work of the public sector accountable and transparent over time; the other is the potential for leveraging public data for the development of new applications and tools by both the private and public sectors.

Civil society organizations have been working both in partnerships with and in competition with the private sector to enact their own imaginary. The idea of OGD draws in

part on the perspective that citizens not only have access to information, documents, and proceedings, but can also become participants in a meaningful way (Harrison et al., 2011). It is a contested imaginary with respect to who is empowered (Gurstein, 2011) and how “accountable” and transparent open data truly is (Peixoto, 2013). Citizen groups, civil society, and “data journalists” are now extensively involved not just in creating portals to Open Government Data, but also pressuring governments to release data and promoting use and re-use. To give one concrete example, in 2011 the Open Knowledge Foundation ran the Open Data Challenge, which was Europe's biggest open data competition to date, attracting 430 entries from 24 European Member States.

This paper, based on an IRC funded New Foundations project, examines the Open Government Data movement in Ireland through what might be called a “data imaginary” lens: the co-construction of data and society through the rhetorics of emergence, contestation, and stabilization.

### **Ireland and Open Government Data**

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), an international initiative which the Irish government implemented in 2014.

Ireland’s Open Government Data initiatives originate from the Department of Public Expenditure and Revenue (DPER). 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 et al., 2014) and a national government data portal was launched (data.gov.ie). The portal received little public discussion aside from one national newspaper, the Irish Times, with 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. As of this date, public engagement and civil society is primarily driven through the OGP to implement the National Action Plan and the self-assessment report has been made open for public commentary.

### **Methodology**

In order to orient the research project within the Irish context, the study began in 2013 with the mapping of stakeholders in Open Data within Ireland. The researchers opted to use manual techniques rather than crawlers to identify relevant open data projects. 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. 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.

Potential interviewees were identified through contacts gleaned from repository websites. 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 were government-funded researchers or employees of local or national government and administrative entities.

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.

### **Emergence of a Vision**

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. 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 neither 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 government data or commercial ventures. Though Piwovar, 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 or government 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. 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.

### **Stabilization**

Invariably, the question of “how do you intend to make this project sustainable and stable?” 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.

## **Contestation**

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.

## **Conclusion**

Our faith in data is daily constructing our data-driven reality. Gitelman (2013) for one argues that our “zeal for more and more data can become a faith in their neutrality and autonomy, their objectivity” (p.3). The ramifications of these ways of talking about data are not trivial and have widespread import. The European Union has become increasingly and overtly conscious of the potential roles of the new information and communication technologies: for financial, social, and political gain, but even in defining a sense of what it means, in this day and age, to be ‘European’. As Morley and Robbins (2013) put it, “Its policy increasingly recognizes that culture is at the heart of the European project (or more crudely put, that questions of culture lie beneath the ‘bottom line’ of the potential profits of pan- European markets)” (p.2).

As Kow and Nardi (2010) write, the imaginary takes place in a long history of understudied and undertheorized constructing and making. Further, constructing the imaginary diverts our collective attention from intervention in shaping our future. One can argue that “the ship has sailed”, that it is too late to intervene or critique in the data-driven future implemented by our institutions. However, new technologies that allow individuals and groups to intervene and critique are increasingly inexpensive and ubiquitous (and add to the quantity of available digital data). The time for an “anticipatory ethics” (Shilton, 2014) of data is now as societies confront the sheer quantity of data being created by governments, companies, individuals, and even technologies that operate and interact on their own (such as sensor networks and automated high-frequency financial trading).

A confluence of factors provides us with the opportunity to investigate how societies are being imagined and transformed. The sheer quantity of data being created by governments, companies, individuals, and even technologies that operate on their own

(algorithmic financial trading software, sensor networks, and closed-circuit televisions, to name a few). While the availability of more data is often touted as an emancipatory good, the rhetoric of the "data-driven society" is also troubling because our recent experiences - with the work of Julian Assange, Aaron Swartz, and Edward Snowden - also suggests that the same data can become an omnipresent, omniscient instrument of control.

OGD, for all its language of participation (of all citizens) and abundance (because data is everywhere), obscures the power relationships, historical antecedents, and anxieties that become embedded in the mundane interactions of data standards, user interfaces, institutions, terms of use, and even what counts as a "successful" deployment. OGD provides us with a rich, textured, multifaceted arena for integrating these perspectives and unpacking them at the same time: to examine how openness and other values are simultaneously universalized and localized, to examine how digital data practices evolve over time and space, how they become contested entities, and indeed, how their form and use encode our very understandings of culture, society, and history. In other words, Open Government Data as a unique laboratory in which to understand historical processes, contemporary narratives, and current and future values across cultures and societies and their implications for constructing the future.

## References

- Anderson, B R. O'G. (1991). *Imagined communities: reflections on the origin and spread of nationalism* (Revised and extended. ed.). London: Verso.
- Anderson, S. R., & Allen, R. B. (2009). Envisioning the archival commons. *American Archivist*, 72(2), 383-400.
- Churchill, E., & Ubois, J. (2008). Designing for digital archives. *Interactions-New York*. 15(2), 10-13.
- Cygniak, R., Kamdar, M., Maali, F., Lee, D., & Decker, S. (2014). *Open Data Ireland: Data audit report*. Galway: Insight Centre for Data Analytics.
- Gitelman, L. (Ed.). (2013). *Raw data is an oxymoron*. Cambridge, MA: The MIT Press.
- Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2).
- Harrison, T. M., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., ... & Pardo, T. (2011, June). Open government and e-government: Democratic challenges from a public value perspective. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times* (pp. 245-253). ACM.
- Havila, I. (2008). Participatory archive: towards decentralized curation, radical user orientation, and broader contextualisation of records management. *Archival Science*, 8(1), 15-36.
- Henneken, E. A., & Accomazzi, A. (2011). Linking to data: Effect on citation rates in astronomy. *ASP Conference Series*, 1-5.
- Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*, 1-47.
- Nardi, B., & Kow, Y. M. (2010). Digital imaginaries: How we know what we (think we) know about Chinese gold farming. *First Monday*, 15(6).
- Morley, D., & Robins, K. (2013). *Spaces of identity: Global media, electronic landscapes and cultural boundaries*. Routledge.
- Peixoto, T. (2013). The uncertain relationship between open data and accountability: A response to Yu and Robinson's "The New Ambiguity of Open Government". *UCLA Law Review Disc.*, 60, 200-214.
- Piwozar, 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](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>
- Shilton, K. (2014). Anticipatory ethics for a future Internet: Analyzing values during the design of an Internet infrastructure. *Science and engineering ethics*, 1-18.

Srinivasan, R. (2006). Where information society and community voice intersect. *The Information Society*, 22(5), 355-365. Webster, F. (2007). *Theories of the information society*. Routledge.

Van House, N., & Churchill, E. F. (2008). Technologies of memory: Key issues and critical perspectives. *Memory Studies*, 1(3), 295-310.

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>

DRAFT