<table>
<thead>
<tr>
<th>Title</th>
<th>Response to Ireland's Open Data Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors(s)</td>
<td>Shankar, Kalpana; O'Neill, Brittney</td>
</tr>
<tr>
<td>Publication date</td>
<td>2014-09-19</td>
</tr>
<tr>
<td>Publisher</td>
<td>Deptartment of Public Expenditure &amp; Reform</td>
</tr>
<tr>
<td>Item record/more information</td>
<td><a href="http://hdl.handle.net/10197/6148">http://hdl.handle.net/10197/6148</a></td>
</tr>
</tbody>
</table>
Comments on the Open Data Strategy for Ireland
19 September 19, 14

Dr. Kalpana Shankar
Lecturer, School of Information and Library Studies
University College Dublin
Kalpana.shankar@gmail.com
1-716-7704

Brittney O’Neill
brittney.o-neill@ucdconnect.ie

We would like to commend Insight and DPER for their hard work and inclusive process on furthering Ireland’s Open Data initiatives. Although we have many concerns and thoughts about the initiatives to date, we will limit ourselves to two categories in which we feel that we have the greatest depth of expertise: the digital curation component and education/outreach of data creators. Some general comments on data audit are also included at the end.

Kalpana Shankar is a lecturer at University College Dublin in the School of Information and Library Studies and Brittney O’Neill is a recent MLIS student. Kalpana has been studying the social and organizational dimensions of data use and sharing for over a decade, most recently from the perspective of repositories and business models for institutional and data sustainability in repositories (funded by the Alfred P. Sloan Foundation in New York). Many of these thoughts are also derived from early findings of her research with Brittney on open data in Ireland (currently being funded by the IRC New Foundations Scheme).

Data life cycle and digital curation
Many of the components of implementation are addressed in the Best Practices and other reports, but the data life cycle developed by the Digital Curation Centre is internationally acknowledged as an easy to understand starting point for conceptualizing and then implementing a data curation strategy and should be consulted and referenced to be current with the international state of the art. In short, the life cycle should be invoked as international Best Practice when developing a data strategy ([http://www.dcc.ac.uk/resources/curation-lifecycle-model](http://www.dcc.ac.uk/resources/curation-lifecycle-model)). Digital curation as a set of key activities in open should be addressed. Preservation planning, appraisal/selection, and disposal need further discussion and implementation. It bears repeating: digitization does not equal preservation. Many other activities are involved.
Capacity building/Outreach and education

The life cycle clearly illustrates an important lacuna in the Open Data Best Practices and Evaluation documents: insufficient capacity in the Irish public sector for understanding or implementing trustworthy data repositories or the higher education sector for training professionals who can work with data. The reports currently suggests that Insight and DPER do not know what they do not know. While Insight has done excellent work in surfacing issues and engaging interest in the private and public sectors, there is still a question of “who will bell the cat?” that needs to be asked and there needs to be more concrete discussion and implementation of capacity building if Open Data Ireland is to be institutionally sustainable.

The tasks needed to make open data useful and usable at hand go well beyond the expertise of computer scientists and engineers. DPER and Insight need to partner with and engage the skills of librarians, archivists, cultural policy and heritage specialists, and academics (most of whom are mentioned in the reports only in passing if at all and many of whom are quite knowledgeable about relevant policies and practices based on work in their institutions) to teach both those in the public sector about open data and develop technical capacity for implementation. DRI, UCD SILS, ISSDA, IQDA, and others have been doing this kind of work and networking with international colleagues to learn and share best practices. In short, these are not just sources of data (in the case of ISSDA and IQDA), but assemblages of people who have skills.
Much of the language of the reports still focuses on a “pipeline” approach where “high quality data” would be funneled to those who “need” it (primarily the private sector). This approach ignores the very real work of making data open – skills which are not simply technical ones, but organisational and social ones. Research suggests that many people already involved with open data (especially in the public sector) still believe that just making data sets available online in CSV format constitutes a trustworthy data repository. Our interviews with open data project managers, however, routinely identify the “human” issues as the most critical in making data trustworthy. Organisational culture and work flow integration are regularly identified as the largest inhibitors to data sharing especially when the data is held within government departments.

Data audit (general observations)

Sheer numbers of data sets are not adequate representatives of the quality/quantity of actual data. For example, qualitative data vs. quantitative, rich vs minimal, dynamic vs. static, regularity of updates are all far more critical than the numbers of data points.

A strict focus on machine-readable data from government ministries only leaves out a substantial part of the Irish Open Data story. The academic and local government sectors are producing/presenting important data, much of which is relevant to the central government.

In the methods section, the report indicates it will only include Government agencies, and yet in the data list, the audit includes The Health Well, ISSDA, and ISDE but leaves out MIDA, IQDA, and all the DRI projects. How was this dividing line concocted? Regardless, it paints a false picture of the Open Data environment, either implying the absence of the MIDA, IQDA, DRI, etc. or falsely labelling ISDE and ISSDA as government bodies rather than academic projects. Ultimately, this highlights the problem that, in Ireland, Open Data is not just government data or just academic data. Much of the data is derived by academicians in academia but using government funding and through government projects. While this structure is certainly not unique to Ireland, it plays an important role in understanding the landscape of Open Data here.

Our interviewees indicated that end-users want live, real-time data from the departments of government. If all available government data is ostensively on StatCentral already, what is gained by this additional crawling? Would it not be better to leverage the work already done by the CSO or mandate better behaviour from departments in terms of data sharing?

Suggestions
• Provide data lifecycle and basic digital curation training to public sector employees. UCD SULS has the practical expertise to offer workshops for groups from the public sector on the digital life cycle and in 2015, is anticipating beginning the first Taught MSc in digital curation in Ireland,
pending approvals. However, workshops could easily be coordinated either on- or off-site to discuss these specifics.

- Go beyond closely held individuals and groups (Insight, DPER) to engage specialists from some of the arenas mentioned above: university librarians, DRI, UCD SILS, geospatial specialists at UCD and NUwe Maynooth all have expertise in areas of open data not represented in Insight or DPER. Usability experts are also needed.

- For future iterations of comments, an open source platform that allows both anonymous/pseudonymous and signed comments would truly allow dialogue among stakeholders. The email process does not allow for anonymous commentary. Although there would be a lot of chaff to weed through in an open forum, the comments would provide a rich treasure trove of qualitative data, and allow movement towards the construction of a “data community” that has been regularly identified as a necessary component of Ireland’s open data strategy.

Summary

We would suggest that these concerns are part of a larger concern with sustainability – not just of digital data, but the institutions that are charged with making data open and available and usable. The former would be best addressed by more attention to the digital curation process; the latter through ongoing training of those who are charged with making data available as well as those who would be interested in using it. These are not just technical/computing skills, but also entail a whole suite of professional activities that should be tapped into. Otherwise, open data will devolve into stale data sets on outdated websites – the digital equivalent of tumbleweeds.