Building our houses on rock

A Position Paper for the Workshop about Open Data on the Web by John Sheridan, Head of Legislation Services at The National Archives

"Therefore whosoever heareth these sayings of mine, and doeth them, I will liken him unto a wise man, which built his house upon a rock: And the rain descended, and the floods came, and the winds blew, and beat upon that house; and it fell not: for it was founded upon a rock. And every one that heareth these sayings of mine, and doeth them not, shall be likened unto a foolish man, which built his house upon the sand: And the rain descended, and the floods came, and the winds blew, and beat upon that house; and it fell: and great was the fall of it" [1]

These are still early days for open data on the web. We have seen rapid progress with various initatives by national governments and cities around the world, an emerging start-up culture and the growing use of open data by the established commercial sector. As with the rise of the first towns in the industrial revolution, much of what has been created so far has been done at pace, with little, viable, long term supporting infrastructure. There is now an urgent need to think about how we solidify the progress that has been made to date, to create more robust open data - high quality, maintained, open data that can be relied upon - that we can use and build on with far greater confidence and trust.

The workshop about Open Data on the Web is an important and timely opportunity to explore three related priority areas:

- Discovering: how do open data users know where there is just sand (ephemeral, unmaintain, unsupported open data), and where there is rock, on which to confidently build? How is this conveyed, given many data catalogues have mixed sets of largely unsupported data with high quality data that is fully maintained and supported? What standards do we have for conveying the solidity of open data and are those standards sufficient?

- Cementing: what are the processes or steps that data publishers need to go through to solidify their open data publishing, to cement it, so others can trust and rely on the data we are invited to use? What feedback mechanisms have proven successful, where data users have been able to encourage data publishers along the road of making greater commitments?

- Sustaining: what experience do we have as a community of various strategies for creating and sustaining reliable and trusted open data:

  a) distributed data curation, management and publishing, where responsibility is shared, and the availability of the data is ensured by a network or community;

  b) institutional commitments and guarentees by trusted organisations to specific open datasets;

---

[1] King James Bible, Matthew 7, 24-27

c) new statutory obligations (eg INSPIRE Directive[2]) to provide and maintain essential reference data.

The National Archives role is to collect and secure the record of government, to preserve it for generations to come, and to make it as accessible as possible. We have a number of official publishing responsibilities, for the UK Government's official legislation website at legislation.gov.uk and also the UK Official Journals, the London, Belfast and Edinburgh Gazettes. We also operate the UK Government's Web Archive[3] - one of the most actively used and relied upon web archives in the world, containing many thousands of archived open datasets from government.

At the workshop we are keen to share our experience of:

- The social, technical and legal steps involved in moving our open data from early experiments to becoming core national data infrastructure, and the role that open standards have played in supporting that journey.

- Making commitments to our open data, through Memoranda of Understanding with partners, and also by publishing contracts (that address, for example, the persistence of URIs) to convey guarentees, so others can rely on our open data and are therefore willing not just to use it, but to contribute to its curation and management, for example our Expert Participation programme for legislation.gov.uk[4].

- Shaping the wider environment, so we can start to link our data with trusted third parties with confidence. For example, we have recently been involved in the initiative to establish a European Legislation Identifier scheme, agreed as a conclusion of the European Union Council of Ministers in October 2012[5], which is now enabling us to use and share relevant, high quality Linked Open Data about legislation with partners in Europe.

- Developing a new proposition (and letting a new contract) for the UK Government's Official Gazette, a service where information can be put permanently on the public record, in the public domain, designed for use as conclusive evidence in legal processes, where the data is open, where the origin is transparent and the provenance is clear, so the data can be trusted.

- High end official publishing, where provenance and authenticity are essential aspects of the value of the open data we publish, how we are using W3C standards in an official publishing context, what has proven easy and difficult, and where there are still gaps.

[3 March 2013]
John Sheridan, Head of Legislation Services, The National Archives.

---

[2] Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:en:PDF
[3] The UK Government Web Archive http://www.nationalarchives.gov.uk/webarchive/
[4] The National Archives Press Release about Expert Participation
http://www.nationalarchives.gov.uk/documents/expert-participation-press-release.pdf
[5] Council conclusions inviting the introduction of the European Legislation Identifier (ELI) OJEU 2012 C325/02
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2012:325:0003:0011:EN:PDF