Discovering Open Data Standards

Deirdre Lee, Derilinx

(based on discussion during the Open Data Standards session at IODC 2016[1])

Open Data Standards

More and more Open Data portals are popping up around the world. National governments are publishing Open Data, as are states, regions and municipalities. Even private organisations are starting to release information in Open Data formats. This is hugely beneficial for making governments and businesses more transparent and enabling new and innovative solutions for society.

No dataset is an island. While digging into one dataset can be interesting, the real value of Open Data is when it is brought together - compared, contrasted, and analysed. Information from different sources can be aggregated to find new insights and trends. This may be data from different domains within an organisation, for example, combining footfall data with the position of transport links in a city. Or it might be data from the same domain, but sourced from different organisations, regions or countries, for example, comparing public contract data across EU states. Or it might be a combination of both – data from different domains and different sources.

This makes the task of aggregating and analysing data complex. Data standards can help facilitate the interoperability of data by providing a common understanding of what is being described and how. The W3C Data on the Web Best Practices recommends to 'reuse vocabularies, preferably standardized ones - use terms from shared vocabularies, preferably standardized ones, to encode data and metadata.'[2]

Building on data standards, using *open* data standards lowers the barrier to entry for data publishers and data users. Rachel Bloom succinctly describes what is meant by an open standard in a podcast on the Geothink project[3].

> *"So when we talk about (open requiring contributions from) multi-stakeholders we're talking about people who contribute that are from different facets of society. So the private sector, the public sector, civil societies, and also the obvious which is that open implies that there should be no royalties or fees associated with using the standard. It should be repurposable, they should be able to extend it how they wish, it should have a license that is open so that there is legal ramification for using the standard as you please."*

However, even in this context, open data standards can be interpreted in a number of different ways. In 'Evaluating domain specific standards' from the International Open Data Charter Working Group[4], Tim Davies and James McKinney point out that when people talk about open data standards they might be referring to:

---

[1] http://opendatacon.org/agenda/pre-events/open-data-standards-day/
[2] https://www.w3.org/TR/dwbp/#ReuseVocabularies
[3] http://geothink.ca/geothoughts-9-geothink-project-measures-open-data-standards-for-consumer-and-publisher-uses/
[4] https://docs.google.com/document/d/1vwgCPg75O_WZX5wLHhxdeDY1mB6KuPA53typ5GYjbRE/edit#

- Principles and policy standards

- The Open Definition

- File formats for data

- Domain specific policy standards

- Generic technical standards

- Domain specific technical standards

For the purpose of this paper, I am referring to open data technical standards, such as DCAT, Schema.org and GTFS.

Problem Statement

Publishing data using open data standards has clear benefits for users. However, the challenge of implementing open data standards lie with the data publisher. The first, and often most difficult question is often which open data standard should I use? This can be broken down further into (i) which open data standards are available and (ii) which open data standards are suitable for my needs?

There is an increasing number of open data standards available, some dealing with general concepts such as people, addresses or organisations, some niche standards for specific domains such as transport, health or science, and some standards that are unique to a country or region. Open data standards are created using different approaches. Standards can be produced from grassroots communities, who have a particular need around data integration and collaborate to establish and grow a standard. This was the case for standards such as OpenSpending[5] and Open311[6]. Another approach to establishing a standard is through a formal process of a standardisation body, such as W3C, ISO or IETF. This is how many web standards, such as HTML5 and RDF, and international standards, such as ISO 3166 Country Codes were formed. Standardisation bodies have the advantage of being able to provide a sustainable framework for establishing and maintaining standards. However, there can be some barriers for citizens, CSOs or SMEs to join. Standardisation bodies may also not have the agility or specialisation to address new or changing needs of different communities. Often a combination of both approaches can be used to create a standard. For example, the Water Point Data Exchange (WPDx) Standard uses widely adopted geospatial and ISO standards as its foundation, and extends this with concepts based on feedback from a niche community[7].

Discoverability of open data standards

While there is a large number or open data standards available, there is not a simple discovery mechanism to find them. Standardisation bodies have dedicated websites and sometimes catalogues that lists their standards. Open data standards often have dedicated websites, for example the OpenReferral[8], the OpenContracting Data Standard[9] and the Humanitarian Exchange Standard[10]. However, there is no one stop shop for finding open data standards if you are a data publisher wishing to publish standardised open data.

---

[5] http://next.openspending.org/
[6] http://www.open311.org/
[7] https://www.waterpointdata.org/about-wpdx-standard
[8] https://openreferral.org/
[9] http://standard.open-contracting.org/
[10] http://hxlstandard.org/

Profiling of open data standards

The second challenge is how to choose which open data standard is appropriate for your needs. There are a number of factors that might influence your decision – is the standard applicable to my domain/case, who else is using the standard, what level of granularity does the standard address, is the standard actively maintained, etc. It is therefore important that this context, or metadata, is available along with the open data standard. It is also important that this metadata is described in a common way across all standards, so it is easy for a potential adopter to search and compare. Yes, a metadata standard for describing standards!

Ongoing Initiatives

There is a growing number of initiatives that are exploring potential solutions for these challenges, including, but not limited to the following:

Geothink, a Canadian geospatial and open data research partnership, are running an open data standards project that is assisting municipal open data publishers in standardising their datasets. As part of this project, Geothink have developed a catalogue for open data standards[11]. Additionally, Rachel Bloom created two spreadsheets looking at 'high-value' open datasets published by Canadian municipalities and an inventory of open data standards applicable to these types of datasets[12]. An accompanying white paper explains the two spreadsheets and the primary objective of Geothink's Open Data Standards Project[13].

GovEx, the Centre for Government Excellence, is gathering a list of civic data standards; a launching point for using interoperability to broaden the reach of data for public good[14, 15]. The aim is to build a library for suppliers and consumers of government data to learn about existing standards, as well as the skills and resources to create new ones. GovEx are also working on a catalogue for Open Data Standards[16].

The International Open Data Charter Technical Working Group are developing a method for identifying and evaluating available data standards relevant to a particular data category, dataset or domain[17]. They have drafted a template that sets out the information to be identified for each standard.

James McKinney and Philip Ashlock worked on a standards' dashboard project with the aim of creating a dashboard for consulting the quality of domain-specific standards at a glance, similar in design to the Open Data Index[18]. This initiative has now been superseded by the standard assessment goals and standard assessment methodologies[19]

The U.S. Data Federation has created a register of initiatives to support government-wide data standardization and data federation initiatives across both Federal agencies and local governments[20].

---

[11] http://www.geothink.ca/opendatastandards/
[12] https://docs.google.com/spreadsheets/d/12wcUhE6waDz0RPT81E5aebcJf58AH92FstMZQKy5kRc/
[13] https://docs.google.com/document/d/1T2-fVOb_l3QD8sBTpEpubFSzWi2qAkwEdPIz1JhJrWU/
[14] https://govex.jhu.edu/enabling-civic-data-standards/
[15] http://labs.centerforgov.org/open-data/civic-data-standards/
[16] https://chriscardoz.github.io/jkan-1/
[17] https://docs.google.com/document/d/1vwgCPg75O_WZX5wLHhxdeDY1mB6KuPA53typ5GYjbRE/
[18] https://docs.google.com/document/d/1ez3UXtBs_YpIZKEmmnDaeMHJ3oX3D0uahAdG8HKU_sw/
[19] https://docs.google.com/spreadsheets/d/1bfSLcqgrJC1bE2kgPRRlrF7Z-A1HQ_EqxL1IG
[20] http://federation.data.gov/

Development Initiatives manage the Joined-up Data Standards Thesaurus to allow users to build, store and map data standards for sustainable development[21].

Data.gov.uk have created a Register of Registers to help service teams across government to build better digital services[22]. One such register is the Local Authority Register, detailing all 354 local authorities in England[23].

And as highlighted in the Call for Participation, the UK's Digital Curation Centre - jointly with the RDA's Metadata Standards Catalog group - manages an extensive catalogue of metadata standards used in different scientific disciplines, including biology, earth science, and social sciences & humanities[24].

Goal for SDSVoc

At the recent International Open Data Conference 2016, Bill Anderson and James McKinney hosted a session on Open Data Standards[1]. A number of those involved in the aforementioned initiatives attended the session, and many of the issues and ideas presented in this position paper were discussed. A full overview of the discussion is available online[25].

The aims of the Smart Descriptions & Smarter Vocabularies (SDSVoc) Workshop are well aligned with the challenges of discovery and profiling of open data standards. Therefore, I propose to bring this discussion to SDSVoc, exploring possible solutions, including an open data standards catalogue and an open data standard metadata standard.

[21] http://juds.joinedupdata.org/thesaurus/
[22] https://www.gov.uk/government/publications/registers/registers
[23] http://local-authority-eng.alpha.openregister.org/
[24] http://www.dcc.ac.uk/resources/metadata-standards/list
[25] https://docs.google.com/document/d/1XkfQrNhixjdrnFun0D4w1N7KRDwa_51nTd9ANCuyPJQ/