

Open Data on the Web, London 2013

Position paper: Bill Roberts, Swirrl IT Limited

Know your audience: maximising use of Linked Data

Too many Linked Data publishing initiatives offer the end user a text box in which to enter SPARQL queries and a few words of introductory information – then the providers of the data wonder why the data is not used.

Linked Data is an ideal technology to support a wide range of data presentation approaches and services backed by data, but its value could be enhanced greatly by paying sufficient attention to the types of user interested in the data, what they would like to achieve with it and their preferred ways of interacting with the data.

While recognizing that any simple categorization of people and their purposes is necessarily crude, the following four groups represent some typical points on the spectrum of user types:

- people who want to know something about the topic: their preferred form of interaction is a well-designed graphic, possibly backed up by a table of underlying data
- researchers and analysts – want in depth information, with details of methodology and provenance, but their tool of choice is usually Excel
- mainstream web developers looking to create visualisations or to re-use data in other services or applications: their tool of choice is a JSON API.
- Linked Data experts who, if sufficiently interested, can do the forensic work to understand the scope and structure of your data: tool of choice is the SPARQL endpoint.

Often only the Linked Data expert user group is supported by Linked Data publishing efforts: this is the least populous of the four user groups.

Maximising the value of data requires supporting all of these groups.

To some extent the linked data experts and web developers have a role in providing value added services on top of ‘raw’ linked data to support the other two user groups, but even in that case, there is an onus on the data provider to help the ‘data intermediary’ to get quickly up to speed with the data.

Developers are people too! They have programming skills, they may be well used to working with data, they may be SPARQL experts: but they still need to spend time understanding the structure of data, how it is interconnected, the data modeling patterns that have been employed and so on. Anything the data

publisher can do to make that process quicker and easier will pay itself back many times in the lifetime of the data.

It is another aspect of the general principle that data tends to get published once but used many times. It is a good investment for those who understand the data in detail to provide as much useful information as possible, rather than requiring each individual user to do that work themselves, with far less familiarity with the data to support them.

In some cases this comes down to old-fashioned documentation: not everything has to be machine readable. Examples that can developers can copy, paste and adapt are particularly valuable.

Also, experience shows that users in the researcher/analyst group, when encountering a linked data website, often come back with a question of where they can download a copy in a format that familiar tools can understand – where can they get a CSV file?

The strength of Linked Data is that it is easy to generate slices of data (possibly combining multiple sources) in a variety of formats. We should exploit that strength to provide information in a range of useful formats.

We are currently applying this philosophy to revise and enhance the data presentation features of PublishMyData, Swirrl's Linked Data publishing platform. I would welcome the chance to make a brief presentation at ODW on work in progress.