

Why should we publish Linked Data?

Jon Blower (1), Maik Riechert (1), Manolis Koubarakis (2), and Nino Pace (3)

(1) Department of Meteorology, University of Reading, United Kingdom, (2) National and Kapodistrian University of Athens, Greece, (3) Advanced Computer Systems, Rome, Italy

We use the Web every day to access information from all kinds of different sources. But the complexity and diversity of scientific data mean that discovering, accessing and interpreting data remains a large challenge to researchers, decision-makers and other users. Different sources of useful information on data, algorithms, instruments and publications are scattered around the Web. How can we link all these things together to help users to better understand and exploit earth science data? How can we combine scientific data with other relevant data sources, when standards for describing and sharing data vary so widely between communities?

“Linked Data” is a term that describes a set of standards and “best practices” for sharing data on the Web (<http://www.w3.org/standards/semanticweb/data>). These principles can be summarised as follows:

1. Create unique and persistent identifiers for the important “things” in a community (e.g. datasets, publications, algorithms, instruments).
2. Allow users to “look up” these identifiers on the web to find out more information about them.
3. Make this information machine-readable in a community-neutral format (such as RDF, Resource Description Framework).
4. Within this information, embed links to other things and concepts and say how these are related.
5. Optionally, provide web service interfaces to allow the user to perform sophisticated queries over this information (using a language such as SPARQL).

The promise of Linked Data is that, through these techniques, data will be more discoverable, more comprehensible and more usable by different communities, not just the community that produced the data. As a result, many data providers (particularly public-sector institutions) are now publishing data in this way. However, this area is still in its infancy in terms of real-world applications. Data users need guidance and tools to help them use Linked Data. Data providers need reassurance that the investments they are making in publishing Linked Data will result in tangible user benefits.

This presentation will address a number of these issues, using real-world experience gathered from four recent European projects: MELODIES (<http://melodiesproject.eu>), LEO (<http://linkeddeodata.eu>), CHARMe (<http://linkeddeodata.eu>) and TELEIOS (<http://www.earthobservatory.eu>). These projects have all applied Linked Data techniques in practical, real-world situations involving the use of diverse data (including earth science data) by both industrial and academic users.

Specifically, we will:

- Identify a set of practical and valuable uses for Linked Data, focusing on areas where Linked Data fills gaps left by other technologies. These uses include: enabling the discovery of earth science data using mass-market search engines, helping users to understand data and its uses, combining data from multiple sources and enabling the annotation of data by users.
- Enumerate some common challenges faced by developers of data-driven services who wish to use Linked Data in their applications.
- Describe a new suite of tools for managing, processing and visualising Linked Data in earth science applications (including geospatial Linked Data).