

The Derilinx Data Publishing Pipeline

Deirdre Lee & Richard Cyganiak

INSIGHT-NUIG

Contributor names and short CVs

Deirdre Lee (deirdre.lee@deri.org) is a Research Associate in the eGovernment group at INSIGHT Centre for Data Analytics, NUI Galway. Her main research interests include Linked Open Data, Open Government, eParticipation, and Semantic Web technologies. Lee is participating in many EC projects, including co-ordinating the EC CIP-PSP project Puzzled by Policy (www.puzzledbypolicy.eu) and is the technical coordinator of the EC FP7 Research for SME associations project Linked2Media. She has published many papers in scientific journals and conferences and has served as a program committee member for conferences and a technical committee member on standardisation bodies, including W3C. Lee received a B.A. in ICT at Trinity College Dublin and a M.Sc. Computer Science (by research) in Trinity in 2008. Lee worked as a research assistant in the IBM Zurich Research Lab from 2004-06

Richard Cyganiak (richard.cyganiak@deri.org) is a Research Associate in the eGovernment group at INSIGHT Centre for Data Analytics, NUI Galway. He works in the Linked Data Research Centre with an interest in web-scale information integration. He is a co-founder of the Linked Open Data (<http://lod-cloud.net/>) initiative, and is creator or main contributor to a number of Linked Data related software projects, including D2RQ, Pubby, Neologism, Sigma, and Sindice. Richard is a fellow of the Web Science Research Initiative.

Type of the presentation proposed

In-use contribution

Title of the presentation

The Derilinx Data Publishing Pipeline

Summary of the presentation (100 words)

In this presentation, we will present the Derilinx Data Publishing Pipeline. The Derilinx Data Publishing Pipeline provides a complete and easy-to-use solution for sharing and integrating data, based on a number of Open Source technologies. Data publishers can use the pipeline to map existing datasets to open Web standards, export the raw data as 5-star Linked Data and publish the data in accessible data catalogues. As a case study, we will present how the Derilinx Data Publishing Pipeline was used to relieve the burden of the discovery, access and reuse of planning application data for multiple local authorities.

Extended abstract of the presentation (1 - 4 pages in 11pt A4 format)

Any query to a government department, be it from within the department, from another organisation, or from a member of the general public, triggers the following response: (1) Where does the data I need to respond to the query reside? (2) How can I access the data I need? (3) Is the data in a format that I can use? This execution of this discover-access-reuse process is currently handled manually, requiring a lot of time and effort from civil servants. For example, let's say a query lands on a civil servant's desk, enquiring about how many square metres of new residential housing have been built in the past year.

The Derilinx Data Publishing Pipeline automates the process of data discovery-access-reuse, to alleviate the burden on civil servants of manually investing many man-hours for each and every query received. The discovery-access-reuse process described above is facilitated through the publication of raw data as 5-star Linked Data via the Derilinx Data Publishing Pipeline. The pipeline provides a complete and easy-to-use solution for sharing and integrating data, based on a number of Open Source technologies. No prior knowledge of Linked Data technologies is required, as the Data Publishing Pipeline provides intuitive interfaces for handling the mapping and publication process.

The first stage is to map existing data-schemas to open Web standards using the ODPP Mapper. The Mapper will assist in the selection of internationally recognised concepts from standardised vocabularies, such as those recommended by W3C, Oasis and the European Commission. Once the schemas are mapped, the original data can be exported as 5-star Linked Data. To ensure the data is discoverable, a Linked Data metadata description is associated with the dataset. Security of the data is paramount, and so the security constraints of the dataset is defined. The data may be restricted to particular roles within an organisation, to everyone within an organisation, or it may be open to the public as Open Data. Finally, the Linked Data is published to a data catalogue, a federation of all accessible data. Publication of the data may be once-off, or, if the underlying dataset is dynamic, it may be configured to update periodically. The data is now ready to be discovered, accessed and reused, as described above.

As a case study, we will present how the Derilinx Data Publishing Pipeline was used to relieve the burden of discovery-access-reuse of planning application data for multiple local authorities. The planning domain was chosen because, while the data is rich, dynamic, frequently used and popular with the general public, there are also not as many privacy concerns as with other data, as much of the data is already out in the public domain. It is also a domain where local authorities carry out a lot of processing, analysis and visualisation, so one that could directly benefit through the Derilinx Data Publishing Pipeline.

The main areas that planning departments deal with are: Planning Control, Forward Planning, Planning Enforcement, and Building Control. Many public services are provided around planning, including submitting a planning application, requesting the status of an application, enquiring about fees, and appealing a decision. Each local authority currently provides a unique set of contact points and proprietary tools to address these services. A person can directly contact an authority via face-to-face, phone and email, or they can utilise the online forms and GIS tools that are available. However the original, raw data is never published.

From our work, we identified the following limitations for planning departments in local authorities:

- lack of interoperability of planning data with data from other departments
- lack of interoperability of planning data across local authorities
- lack of availability of reusable planning data for third-parties wishing to innovate around the data

In our presentation, we will demonstrate how these limitations are increasing workload, preventing innovation, and preventing effective decision-making. However, we also demonstrate how the use of the Derilinx Data Publishing Pipeline can directly address these issues for greater socio-economic impact.