An Introduction to Linked Data

Ingrid Mason¹, Anne Cregan², Conal Tuohy³, Anna Gerber⁴

¹Intersect Australia Ltd, Sydney, Australia, ingrid.mason@intersect.org.au
²Intersect Australia Ltd, Sydney, Australia, anne.cregan@intersect.org.au
³VeRSI, Melbourne, Australia, conal.tuohy@edu.au
⁴University of Queensland, Brisbane, Australia, agerber@itee.uq.edu.au

GENERAL INFORMATION

- This workshop is a half-day (3.5 hours) and includes a half hour break.
- Ingrid Mason is the primary convener for the workshop.
- The workshop does not include a hands-on component.
- There are no constraints on the number of attendees.
- The presenters require the capacity to plug a laptop into a projector, as a means to present and to demonstrate some projects and tools that support the construction of linked data.
- Wireless is desirable for the attendees.

DESCRIPTION

The workshop will provide:

- a basic technical understanding of linked data and linked open data
- an explanation of how XML (format), RDF (model) and SPARQL (query language) are used in linked data
- some examples of how linked data is being used in eResearch projects nationally and internationally
- information about tools that support the construction of linked data
- an explanation of some of the challenges in integrating heterogeneous datasets using linked data
- an insight into a community of practice that is forming around the provision of linked data services
- an overall understanding about how using linked data as a means for integrating heterogeneous datasets has an impact on resource discovery and also generates a new dataset for researchers to interpret

OUTLINE

Please provide an outline of the workshop content using the following format.

1. **What is linked data?** An explanation of linked data and linked open data.
   15 minutes

2. **What do these technical acronyms mean?** An explanation of XML, RDF and SPARQL.
   15 minutes

3. **How is linked data being used in eResearch?** An introduction to a selection of eResearch projects that use linked data to integrate datasets.
   30 minutes

4. **What tools can be used to create linked data?** An introduction to the tools that can be used to integrate datasets and support the creation of linked data.
   30 minutes

5. **What are the challenges in using linked data?** A discussion of some areas where linking data can get tricky or problematic.
   30 minutes

6. **How are other communities using linked data?** A discussion about the W3C Linked Library Data Incubator Group and the Linked Open Data – Libraries, Archives, Museums summits.
   30 minutes

7. **Why use linked data?** An explanation of the benefits of using linked data to integrate scholarly datasets.
   30 minutes

WHO SHOULD ATTEND

This workshop is for non- or semi-technical professionals working in eResearch. Academics, project managers, business analysts, university data, library and information professionals that want an easy and quick insight into what linked data is and how this method of structuring data can support the integration of heterogeneous datasets and serve researchers' needs.

WHAT TO BRING

Attendees need to have an interest in learning about linked data and ideally have some professional involvement in eResearch or research data management projects or services.

ABOUT THE PRESENTERS

Ingrid Mason. Ingrid has twenty years experience in diverse roles (project manager, repository manager, business analyst, librarian) working with technology in the university, cultural and government sectors in New Zealand and Australia. She has a Bachelor of Arts (English Literature) and Masters of Library and Information Studies (Victoria University of Wellington). Ingrid is establishing a programme of activities targeting eResearch support for humanities, arts and social science researchers, and learning and development. She is also a liaison for ANDS for NSW universities.

Anne Cregan. Dr Anne Cregan has been an eResearch Analyst at Intersect for 4 years and has worked on a broad spectrum of eResearch projects from inception through to deployment. Prior to working at Intersect, she did a PhD in Semantic Web technologies at the University of UNSW, focusing on ontologies, conceptual modelling and knowledge representation. Her PhD was sponsored by NICTA, and involved working in W3C and ISO standards committees for knowledge modelling and extraction. Prior to her PhD, she also has a wide range of IT, Analyst, Data Mining and Management experience from the corporate world, and a Bachelor of Science from the University of Sydney with honours in Psychology. She is currently interested in exploiting cloud technologies for data management and real-time services.

Conal Tuohy. Conal is an eResearch Analyst, focusing on the digital humanities domain and optimising software architecture for reuse. Conal has a background as a software developer, particularly using Java, SQL, Topic Maps, XML and XSLT. He has a particular interest in text encoding with TEI , and served for 2 years on the Technical Council of the Text Encoding Initiative Consortium.

Anna Gerber. Anna is currently part of the Open Annotation Collaboration. Anna is developing annotation tools and services that can be used to evaluate and demonstrate the applicability of the OAC data model in the context of annotations supporting collaborative development of scholarly editions. Anna is also involved in the AustESE project, developing eResearch tools to support the collaborative authoring and management of electronic scholarly editions. From July 2008 - July 2011, Anna was the Senior Software Engineer for the Aus-e-Lit project, a NeAT-funded project that provided collaborative annotation services, federated searching, graphical empirical reporting and compound-object authoring and publishing services for teachers, researchers and scholars of Australian Literature. Aus-e-Lit was a collaboration between AustLit and the eResearch Group.