DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2017

A Data Model for Lifecycle Management of Natural Hazards Engineering Data

Presentation

Maria Esteva
University of Texas at Austin, USA
maria@tacc.utexas.edu

Ashley Adair
University of Texas at Austin, USA
a.adair@austin.utexas.edu

Craig Jansen
University of Texas at Austin, USA
cjansen@tacc.utexas.edu

Josue Balandrano Coronel
University of Texas at Austin, USA
jcoronel@tacc.utexas.edu

Sivakumar Ayeegoundanpalay Kulasekaran, University of Texas at Austin, USA
siva@tacc.utexas.edu

Keywords: research data lifecycle; Fedora 4 repository; multi-structured metadata

Abstract

Natural Hazards engineering data derives from sophisticated experimental design and contains a complex array of relationships. Representing and publishing these data is challenging, as the domain lacks a metadata schema and specialized vocabulary. To build the functionalities required to curate and publish datasets within the DesignSafe-CI, an end-to-end research data lifecycle platform (https://www.designsafe-ci.org), the curation team took a multi-step approach.

First, the team undertook modeling of the research processes of seven kinds of experimental projects and corresponding hazard types that are supported by the CI. Researchers in the space were asked to draw and describe their research workflows, noting the equipment, the processes involved and their output data, the software used to analyze the data, and the documentation that are indispensable for proper data interpretation and reuse. To derive a generic experimental data model, the team analyzed these workflows and identified common processes as well as the relationships between those. The activity arrived at core metadata elements that represent the steps and methods involved in Natural Hazards projects, as well as sets of user-suggested vocabularies specific per experiment type. The resultant data model emphasizes the datasets structure and the provenance of the multiple data outputs obtained from different configurations within an experimental project. Definitions for the core metadata and vocabularies are maintained in the community meta-dictionary YAMZ (www.yamz.net).

In the DesignSafe-CI portal the data model was implemented as interactive functions that allow users to progressively tell the story of their project by categorizing, describing, and relating data from an experiment. Using the metadata and the vocabularies users can start and stop working with their data at any time: selecting and deselecting files, adding or removing categories, and editing descriptions. As users go about curating their files in the interface, the network of relations of the experiments is formed in the back-end through a middleware metadata API. This allows rendering the experiments graphically as a tree showing the links between processes, data, and descriptive tags and narratives; helping users arrive to the decision of publishing.

At the point of publication, the final data and metadata transitions to a Fedora 4 repository. For this, we mapped each of the elements and terms from the data model to three metadata schemes: Dublin Core, to describe the experimental project; PROV to represent provenancial relationships among processes and their outputs; and DataCite for metadata that will be passed in the minting of Digital Object Identifiers (DOIs). Mapping across these schemes results in multi-structured metadata that standardizes elements and vocabularies. Beyond description and contextual

https://doi.org/10.23106/dcmi.952138030

information as minimum requirements to publish scientific data, this data model emphasizes the structure of the experiments and uses terms familiar to users in the domain to facilitate data reuse. Its mapping to standard schemas enable proper publication, exchange, and web exposure of the data, and allows queries that relate the components. Friendly user evaluations conducted for the preliminary release of the curation and publishing pipelines suggest that they are intuitive and will be complemented with a larger study in the Fall.

https://doi.org/10.23106/dcmi.952138030