

eResearch Administration: The game will change. *Designing a Research Data Registry for a world in which there are no information silos, no single sources of truth, and no one system to complete a transaction.*

Simon Porter¹,

¹University of Melbourne, Melbourne, Australia, sjporter@unimelb.edu.au

INTRODUCTION

In the relatively young history of research administration, administrative systems built to support research processes have undergone a reasonably literal translation from the paper based to the electronic. Under this model, the paper, or the electronic form operates as a blank slate upon which all the information required to complete a process must be entered, regardless of the fact that many of the details supplied bare a remarkable resemblance to information that a researcher has previously entered in another form for yet another process.

This presentation demonstrates a new model for research administration in which each new process is a continuance of a conversation that builds upon everything that has been previously entered about a researcher, both locally and via externally available data sources. This model has been developed in the context of designing a Research Data Registry that supports both the requirements of the University of Melbourne's to know about the location of research data that it is responsible for, and those of the Australian National Data Service to make research data more discoverable.

PRACTICAL EXAMPLES OF RESEARCH DATA ENTRY PROCESSES THAT REDUCE THE AMOUNT OF DATA NEEDED TO COMPLETE A TRANSACTION

In order to describe a research data set, it turns out that you need to describe almost everything else about the research that it relates to as well. What were the publications that the research data set supported? Which project created it? Was this project funded by a grant? To what department does it belong? Who were the researchers? What research classifications does it have?

In a traditional research administration system, this 'information context burden', leads to administrative process that are strongly resisted because of the work that it takes to complete the associated data entry. In systems where we already know about a researcher's grants, publications, and research classifications however, these contextual links become an asset that can significantly reduce data entry. For instance, if a research data set is related to a publication, simply by making the link between publication and the data set, we can infer the data sets probable creators, custodian department, and research classifications.

TECHNOLOGIES THAT SUPPORT A NEW APPROACH TO ADMINISTRATION

In building a research system capable of reusing a broad range of contextual information each research system almost needs to be its own data warehouse on research. This requirement emerges from the perspective of supporting a data model capable of representing the entire research context, and the challenge of storing or at least having all of the available information available via feeds. Without standard data models for research, as well as standard interfaces, this enterprise would seem doomed to fail.

As a starting point for the Research Data Registry, the project chose to build upon the work of the vivo web project, which has built a semantic web application with ontology capable of describing the entire research context. A semantic platform was chosen because of the fact that most data about research is not private, and there is significant promise in being able to pull in linked data sources for research data into the future.

NO SINGLE SOURCE OF TRUTH

Despite a focus on data reuse, it will sometimes be necessary to describe related entities such as projects, or publications. What are the processes required to make sure that the effort required to describe these entities is not wasted, and can also be reused. How should distributed sources of truth be managed?

NO ONE SYSTEM TO COMPLETE A TRANSACTION

In the context of a research data registry it may be that fully formed research data records are collected in the process of research, rather than as an adjunct to it. In this environment research data registry process must support models that both harvest and create records in equal measure.

REFERENCES

<http://creativecommons.org/licenses/by/4.0/>



Attribution 4.0 International