Contribution ID: **51**   Type: **Presentation**

Frictionless Data Exchange Across Research Data, Software and Scientific Paper Repositories

Tuesday, 9 October 2018 12:00 (15 minutes)

A single scientific repository is, if considered by itself, of limited value. Real benefits come from the ability to exchange information effectively and in an interoperable way, enabling the development of a wide range of global cross-repository services. However, exchanging metadata and content across scientific repositories is mostly based on a 15-year-old technology, symbolized by the OAI-PMH protocol.

This protocol is:

1. is unsuitable when there is a need to exchange large quantities of metadata,

2. suffers from inconsistent implementations across providers and

3. was only designed for metadata transfer, omitting the much needed support for content exchange.

In light of these issues, the COAR Next Generation Repositories Working Group recommends the adoption of ResourceSync across repository platforms. As a result, it is important that we fully understand how ResourceSync performs against OAI-PMH. This work is being conducted under the umbrella of the European Open Science Cloud Pilot project from which we received funding to run experimental pilot to provide a fast and highly scalable exchange of data across repositories.

The work will assess how scholarly communication resources, i.e. research datasets, scientific manuscripts (research papers, theses, monographs, etc.) and scientific software, can be effectively, regularly and reliably exchanged across systems using the ResourceSync protocol.

The underlying aim of this work is set to provide an argument and evidence for modernising existing legacy communication mechanisms routinely used by thousands of research repositories. This will be achieved by running a set of experiments/benchmarks comparing OAI-PMH with ResourceSync along a set of dimensions, scenarios and implementation setups, including:

Architectural
- 1-to-1 synchronization
- 1-to-many synchronization (master copy or mirror) experiment
- many-to-1 synchronization (aggregator)

Conceptual
- Baseline synchronization
- Metadata
- Metadata and content
- Incremental synchronization
- Selective synchronization (PMH Sets, RS capability lists)

We will also compare/evaluate the efficacy of ResourceSync against OAI-PMH in terms of:
- speed (time)
- complexity (steps required to complete)
- reliability (recall)
- freshness (e.g. average time gap between syncs)

The evaluation will also consider different implementation set ups, such as sequential vs parallelized implementation of a ResourceSync client. The proposed talk will concentrate on presenting the first set of results form the evaluation.

Type of abstract

Presentation

Summary

Real benefits of repositories come from the ability to exchange information effectively and in an interoperable way, enabling the development of a wide range of global cross-repository services. However, exchanging metadata and content across scientific repositories is mostly based on a 15-year-old technology, symbolized by the OAI-PMH protocol. We perform a benchmarking evaluation of OAI-PMH against ResourceSync (a not yet widely adopted protocol) in a wide range of scenarios to provide an evidence-based argument for modernisation of technology for interoperable data exchange between repositories containing research data, software and scientific papers.

Primary author: Dr KNOTH, Petr (KMi, The Open University)

Presenter: Dr KNOTH, Petr (KMi, The Open University)

Session Classification: Thematic Services

Track Classification: Area 1. Cross-Domain challenges / Data exchange across domains: researchers, technologist and policy makers perspectives