

Conference proposal for Open Repositories, Atlanta, Georgia, May 18-21, 2009.

Fedora Content Modelling for Improved Services for Research Databases

By Mikael Karstensen Elbæk, Alfred Heller, and Gert Schmeltz Pedersen, DTU Library¹, Technical University of Denmark.

Abstract: A re-implementation of the research database of the Technical University of Denmark, DTU, is based on Fedora. The backbone consists of content models for primary and secondary entities and their relationships, giving flexible and powerful extraction capabilities for interoperability and reporting. By adopting such an abstract data model, the platform enables new and improved services for researchers, librarians and administrators.

ORBIT² - **Online Research dataBase In Technology** - is the official research database of the Technical University of Denmark, DTU. It is developed by DTU Library.

ORBIT provides all information on scientific, popular and educational publications used in research activities at DTU. ORBIT serves also as an Institutional repository with pre-, in- and postprints and a Current Research Information System (CRIS) for statistical and administrative purposes. ORBIT is the central tool for departments, sections and the individual researchers to present their publications and externally funded research projects.

The ORBIT search interface gives worldwide public access to information about research and development activities at the University and provides an overview of publications to which DTU researchers have contributed as authors or co-authors. ORBIT also provides the possibility of cross search on publications, projects, resources, events, department profiles and personal staff profiles. ORBIT is interoperable by exposing the data through OAI-PMH and SRU/SRW, delivering research data to the national research database, the European project DRIVER and Google Scholar, to mention some.

ORBIT aims to target researchers, administration and management, the business community, the press, teachers and information brokers.

Fedora has been chosen as backend for a re-implementation of ORBIT, in order to facilitate development of services and adapt the system with more ease and transparency. DTU Library has a broad and deep knowledge of Fedora and has contributed to its development with GSearch, the generic search service module.

¹ DTU Library, <http://www.dtic.dtu.dk>

² ORBIT - **Online Research dataBase In Technology**, <http://orbit.dtu.dk>

In the current work, the goal is to improve user services that are unrealistic to be implemented in the current setup. Hence Fedora is utilized to overcome these obstacles. By simply using Fedora the following added services are to be harvested:

1. standardized abstraction and persistence layer for resource urls
2. full text indexing
3. ability for relating independent records without influencing the records themselves
 - a. Through the triple store, this gives the opportunity to query reports of related materials that are not possible today.
 - b. It is also possible to enable an authoring tool that requests and interprets records within the database.

In a longer perspective, Fedora facilitates:

- Complex, Compound, Dynamic Objects
- Enhanced publications
- Primary research data
- Repository-based research environments
- Connecting repository islands
- Open interfaces and repository interoperability

The technical aspects

The architecture of the Fedora-based ORBIT implementation is seen in Figure 1. Components are:

- The Fedora web application with its repository of *foxml* objects and its *triplestore*, supplemented by the *fedoragsearch* web application with its *lucene* index.
- OAI-PMH harvest and ingest, which wraps all sorts of metadata into *foxml* objects, including the generation of triples in the *RELS-EXT* datastream, representing the entities and relationships in a model of research publications and primary and derived research data.
- The *d2fcris* web application, which services query clients and cataloguing clients.

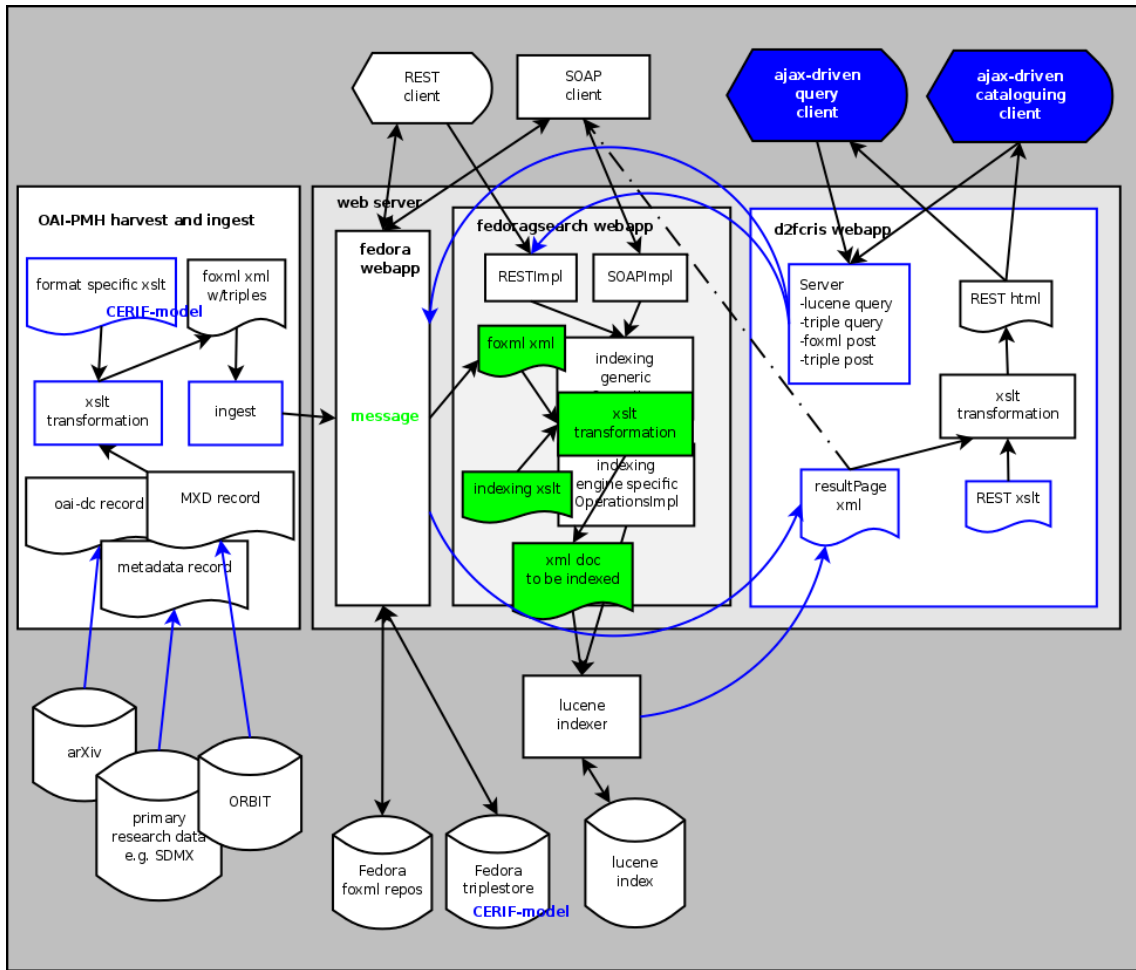


Figure 1: Architecture of the Fedora-based ORBIT implementation

Abstract Content Modelling

The backbone of the architecture is the model of entities and relationships. It has several important aspects:

- It serves as a high level abstraction
- It determines the set of relationships used in the triples
- It is used in the format specific xslts that transform metadata records into triples
- It provides flexibility

The initial model of entities and relationships is a subset of the CERIF-model³, which has been developed within the euroCRIS association, see figure 2. It is an important point that the architecture allows the model to be evolving over time to accommodate new requirements. In such cases, the format specific xslts are edited to generate triples that

³ CERIF 2008, <http://www.eurocris.org/cerif/cerif-releases/cerif-2008/>

represent the new model, and then all the metadata in the foxml objects are retransformed.

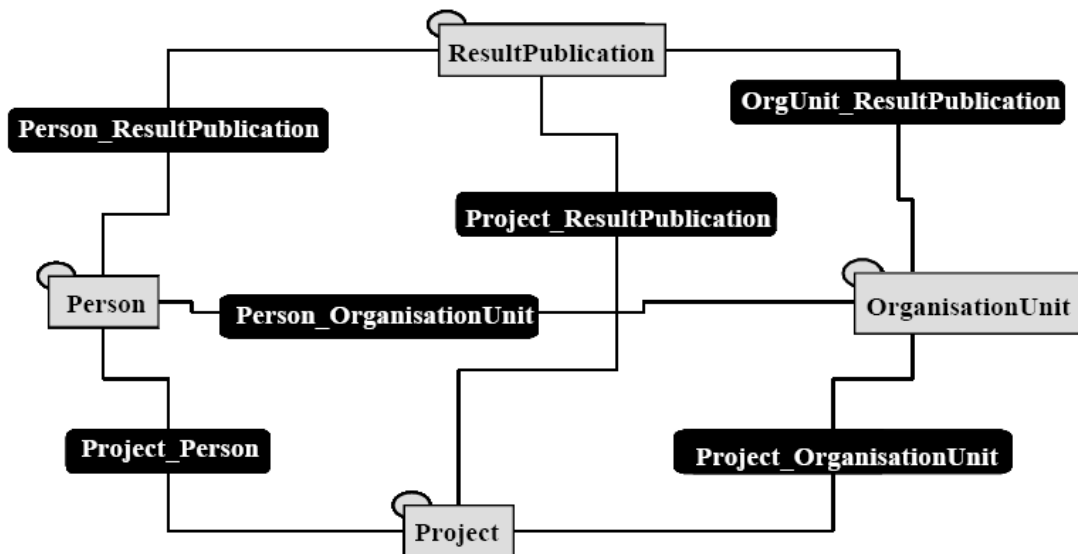


Figure 2: Initial entities and relationships for research information, CERIF2008

Examples of new services that are facilitated by the new implementation are:

- flexible and extended tools for reporting, making customized reports possible
- interlinking between research entities like projects and publications, thus:
 - enhancing information retrieval providing several entries to research, i.e. from project to publications to person to related events etc
 - enabling author tools that makes creating relations between research entities easy

The presentation will demonstrate and evaluate the benefits achieved by using Fedora for the re-implementation of ORBIT.