

ABSTRACT

The institutional repository as a general university administrative system

A practical example



Paper for the DORS DL workshop at ECDL 2006, version 1.0

Authors

Aalborg University Library
Thøger Kristensen
Head of Acquisitions
tk@aub.aau.dk

Atira A/S
Bo Alrø
Product Manager
ba@atira.dk

Contents

| | |
|---|----------|
| 1. Abstract | 3 |
| 1.1. Intro | 3 |
| 1.2. The process of joint development | 3 |
| 1.3. Meta-data model | 4 |
| 1.3.1. Changes to the meta-data model | 4 |
| 1.4. Key words | 4 |
| 1.5. Data gathering | 4 |
| 1.6. Dynamic integration | 4 |
| 1.7. Data import | 4 |
| 1.7.1. A step-by-step example | 5 |
| 1.8. Data storage | 5 |
| 1.9. Data exhibition | 5 |
| 1.9.1. Web services | 5 |
| 1.9.2. OAI | 5 |
| 1.9.3. Z39.50 | 5 |
| 1.9.4. Generic XML I/O | 5 |
| 1.10. Bibliometry | 5 |
| 1.11. Reporting | 5 |
| 1.12. Modules and the ASP-model | 6 |
| 1.12.1. Modules | 6 |
| 1.12.1.1. Reports, Bibliometry, Student projects, Press, External publications and PUREportal | 6 |
| 1.12.1.2. ASP model | 6 |

1. Abstract

1.1. Intro

PURE constitutes an institutional repository in the traditional sense, but it also adds a number of other features to the definition.

Development of PURE was initiated at Aalborg University Library and is now used at 7 other universities plus a number of other research institutions. PURE has become a central piece in the research administration and dissemination efforts of the universities and the core of emerging Current Research Information Systems.

Also, PURE is a tool for batch ingestion and conversion and for general management of large data collections - as well as a tool for creating technical access to data over web services, OAI and other industry specific protocols.

In addition, PURE is a tool for automated, rights-managed electronic publishing of bibliographical meta-data and full-text files. This is primarily realised by means of two web services (each with a rich library of exhibition methods). Further, it is facilitated by the PUREportal development framework, which allows fast development of whole research portals on top of PURE.

Leaving the traditional definition of an institutional repository, PURE is also a tool for planning and managing the work processes of data gathering - thus involving large groups of users and introducing GUI issues. The workflow orientation and the role-based distribution of user tasks constitutes PURE as more than the publication database, which was the initial focus point.

Furthermore - and of importance to the library- and university- management and administration - PURE is a central solution for advanced statistics and reporting about research publications, research funding, and other types of research activities.

And finally - and of particular importance the the individual researcher and research teams - PURE is an easy-to-use tool for routinely submitting research as well as a fast-track onto personal and project-specific websites and into the Google indexes.

1.2. The process of joint development

The so-called PURE board have representatives from each university using the system and from the private development company. The board decides on all important development steps and coordinates shared financing of development work.

1.3. Meta-data model

Any meta-data model can be implemented in PURE; however, PURE also comes with a ready-made meta-data model that handles content types such as Persons, Organisations, Projects and Publications, plus several secondary types. Many-to-many Relations exist between all primary content types. Particularly the Publication type and sub-types are rich in their definition, surpassing among others the CERIF meta-data model.

1.3.1. Changes to the meta-data model

The PURE meta-data model was developed jointly by the 7 universities and a number of other research institution. The process of joint development lead to a general-purpose meta-data model, which is used collectively by all universities.

1.4. Key words

A central keywords hierarchy can be defined by system administrators in PURE (unlimited in levels and with the option for researchers to add their own keywords at the lowest level). Keywords from this central hierarchy can be added to all content types: Persons, Projects, Organisations, Publications, etc.

1.5. Data gathering

Data can be entered centrally at a library, but most universities gathers data de-centrally at the source (research staffs at individual institutes). A roles model and applied work-flow ensures that only valid and approved data is submitted to the repository.

1.6. Dynamic integration

Most of the PURE-using universities demand the ability to dynamically integrate PURE to sources for all or some of these content types: Persons (and their roles), Projects, Organisations and News feeds. Also, in many cases, user authentication must be build on dynamic integration. PURE offers a technical framework with a plug-in architecture for customizing integration to such sources.

1.7. Data import

Methodical, controlled and validated import of historical data is made possible by means of the PURE XML Archive (PXA) format and a technical framework for developing converters from different data sources. A number of standard converters are currently being build; for PubMed, among others. However, much manual work is still needed when importing.

1.7.1. A step-by-step example

Will be part of the final paper.

1.8. Data storage

PURE stores to a Customer-definable SQL environment through the a Hibernate-facilitated O/R mapping layer. Binary objects are stored in the file system of the server. Further, ready-made connectors to The DSpace and FEDORA LTP systems also exists, facilitating easy integration to long term preservation systems.

1.9. Data exhibition

1.9.1. Web services

To web services exits in PURE: RPC/encoded and Document/Literal, each with a rich library of methods. Among other options, data can be exhibited directly from these web services in library formats such as APA, VANCOUVER, HARVARD and others.

Code examples will presented in the final paper.

1.9.2. OAI

PURE supports meta-data harvesting over the OAI-PMH protocol as a Data provider. Currently, both the DC and the far richer DDF-MXD format are supported. A technical framework allows for easy implementation of further formats.

1.9.3. Z39.50

Z39.50 is currently supported, and SRW/SRU will be at a later time.

1.9.4. Generic XML I/O

A generic XML I/O unit allows for data represented in XML to imported and exported.

1.10. Bibliometry

Various citation types and journal impact factors can be added. Import from ISI is planned. Import from Scopus is contemplated.

1.11. Reporting

A reporting module contains a number of ready-made reports for listing and calculating data in PURE. Each report can be customized, and new custom reports can be defined, too. A similar set exits for bibliometrical reporting and statistics.

1.12. Modules and the ASP-model

1.12.1. Modules

Below is a list of modules and a description of the ASP model:

| Module | PURECHASE | | ASP |
|-----------------------|--------------------|---------------------|-----------------------------|
| | Commercial license | License maintenance | License-free hosted service |
| PURE basic module | X | X | X |
| Reports | X | X | X |
| Bibliometry | X | X | X |
| Student projects | X | X | X |
| Press | X | X | X |
| External publications | X | X | X |
| PUREportal | X | X | X |

Table 1: Availability

1.12.1.1. Reports, Bibliometry, Student projects, Press, External publications and PUREportal

More information in the final paper.

1.12.1.2. ASP model

For smaller research institution, an ASP service is made available. More information in the final paper.