

Proposal for Open Repositories 2014

University Institutional Repository and its Place in the University Information Infrastructure

Věra PILECKÁ¹, vera.pilecka@uk.cvut.cz

Helena KOVÁŘÍKOVÁ¹, helena.kovarikova@uk.cvut.cz

Lenka NĚMEČKOVÁ¹, lena.nemeckova@uk.cvut.cz

Marta MACHYTKOVÁ¹, marta.machytkova@uk.cvut.cz

Ivo PRAJER², prajer@cvut.cz

Petr KAREL², karel@cvut.cz

¹ Czech Technical University in Prague – Central Library

² Czech Technical University in Prague – Computing and Information Centre

Abstract

University repositories are no isolated archives. They provide significant added value to the existing university information infrastructure, and its integration within the university information system is essential. The Czech Technical University in Prague (CTU) systematically builds its repository and integrates it within various subsystems of the university information system. It provides mutual advantage. The repository uses existing data and in return it enriches the information infrastructure of the university with the archiving tool and its content.

This paper discusses technical and process solution to the interconnection of university repository (DSpace platform) with individual subsystems of the university information system, and the design of the optimized data flow and data management between the entities/components involved. The intention is to optimize the university data workflow – multiple uses of data (which have once been entered by various subsystems of the information system) to automatically manage repository structure and create its content without the need of frequent manual intervention.

The article introduces the interconnection of individual components, subsystems and processes related to the university institutional repository, and describes the role of individual item and data flow in the process.

Keywords

University repository infrastructure, University information system, University repository integration, University data workflow

Audience

This paper is aimed at repository administrators – technical system administrators, and repository data managers.

Background

The article aims to present how repositories might be created, positioned, and integrated within an existing university information infrastructure to provide an adequate value-added research information ecosystem in a university environment. The network includes university CRIS system, electronic theses and dissertations, journal publishers, library system, and identity management system. The organization and technical solution to data usage and automatic data flow between various subsystems of the university information system and the repository is essential. Our solution is also aimed at optimizing information workflow in terms of full-text archiving.

Presentation content

The university information system is composed of various subsystems aimed at different purposes and at different users. Those include educational, research, financial, organizational and other subsystems. Each subsystem/component uses some unique data and processes, and some others that are shared by multiple systems. In 2009, the CTU started developing an institutional repository and based on its scope it became obvious that it would have to interact with various data from multiple subsystems. The necessity to integrate these components and create an optimal data workflow between the involved entities arose. Because of the university information system structure, this solution is likely to be a unique one.

Subsystems of the university information system involved in this process include:

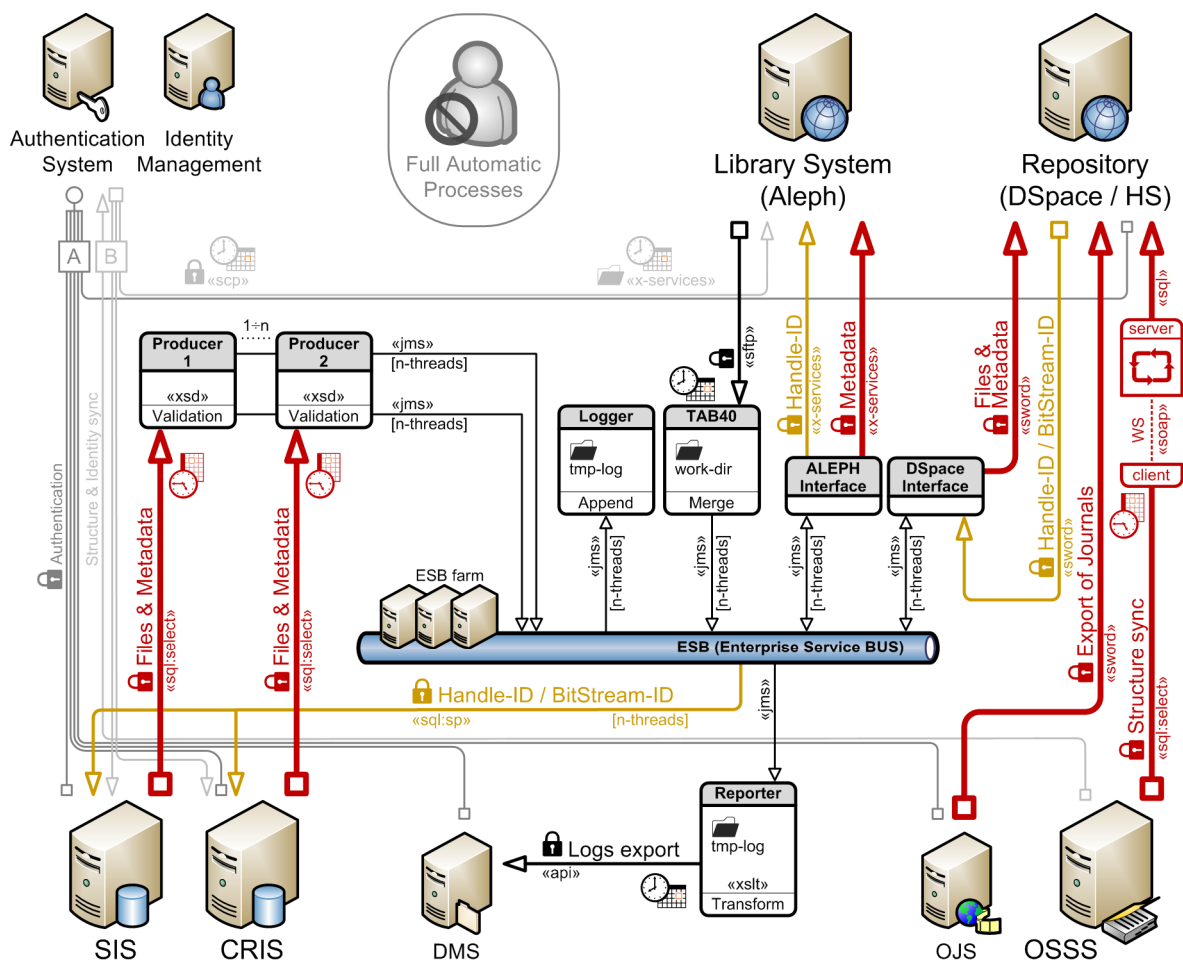
1. Metadata, full text sources, and documents identifiers:
 - Study Information System – **SIS** (contains: Electronic Theses and Dissertations - ETDs),
 - **CRIS** (contains: research outputs),
 - Open Journal Systems – **OJS** (a journal management and publishing system – contains: university journals and possibly also conference proceedings),
 - **Repository** (Handle server / DSpace – contains: unique handle identifier for items and identifier for bit streams)
2. Identity information / University organizational structure:
 - User identity management, User directory and authentication system (contains: community user IDs including user authentication information),
 - Organizational Structure Source System – **OSSS** (contains: information about academic bodies, faculties, higher education institutes and other constituent parts of university – e.g. its full identification, relation between them, its history)
3. Main target systems:
 - **DSpace** (contains: records and full texts from the above mentioned subsystems),
 - **Library System** (Aleph by ExLibris - contains: library entities' records, ETDs' records)

The solution involves creating various import/export software transactional “bridges” implemented at interfaces of each entity to communicate the correct data with each other. This is based on the SOA (Service Oriented Architecture) technology, and ESB (Enterprise Service Bus) open source platform. Processes are designed in BPMN (Business Process Model and Notation) and hardcoded into ESB processes.

What novel technical solutions for DSpace have we implemented?

- Automatic repository structure management** (DSpace communities and collections) are based on the university organizational structure as entered in the university information system (University – Faculties and other bodies – Departments and other parts). Data exchanger, a heart of the information system, and its part, the Organizational Structure Source System, monitors the lifecycle of the university structure and immediately reflects its updates to DSpace through software bridges to keep the up-to-date structure so all records are placed at the correct location. It is a simplex synchronization solution. For example, if a department is abolished, its collection record and associated items in DSpace are kept in the old structure, they are marked and not removed; new records are associated with the upgraded structure.
- Automatic data (metadata and full text) import from various subsystems/components** (as listed above). Each subsystem is an individual unit, containing customized communication interface with DSpace (so called Producer). The Producer reflects each subsystem's data presentation, validates and transforms data to unified message format to be used by the data transport bridges. Metadata and files (bit streams) are imported straight into the corresponding structure (collection) in DSpace, together with user rights specification/attributes all the way down to the level of individual items and files (bit streams). User rights are set in the original subsystem and retain with the file in DSpace as well.

Technical solution / Data flows / Integration scheme:



3. Further interconnections

- **Handle identifier** is the tool that enables the interconnection of individual records and full texts in various subsystems. It is generated in Handle server (part of Repository) and imported back to the particular components (SIS, CRIS and Library System)
- **Connection with the world** is enabled through OAI-PMH protocol that has been further implemented for the repository to be harvested by external engines (e.g. OpenAIRE, BASE, Czech National Repository – NUŠL, etc.)

Conclusion

Each institution is unique in its organizational structure and policy and needs its own data management solution. The Czech Technical University in Prague has brought up its own solution to integrate components, that would first sight have nothing in common, however have become connected because of the repository – they all deal with some kind of full text outputs that are most likely to be archived and made available (at least to the internal research community, or even better to the international community and to the public).

References

NĚMEČKOVÁ, Lenka, Věra PILECKÁ and Helena KOVÁŘÍKOVÁ. Library as a coordinator of university Open Access initiatives: institutional repository in the university information infrastructure. (poster) In: *34th IATUL Conference 2013, Cape Town, April 14th – 18th 2013*. Available at: <http://docs.lib.purdue.edu/iatul/2013/papers/9/>

NĚMEČKOVÁ, Lenka, Marta MACHYTKOVÁ, Helena KOVÁŘÍKOVÁ and Věra PILECKÁ. Open University: University Library and the Shifting in Scholarly Communication Paradigm. In: *LIBER 41st Annual Conference, 27 - 30 June 2012, Tartu, Estonia* [online]. Available at: http://www.utlib.ee/liber2012/index.php?id=prog_main

KOVÁŘÍKOVÁ, Helena. DSpace na ČVUT [DSpace at the CTU]. In: *2. setkání českých uživatelů systému DSpace [2nd Czech DSpace User Group Meeting]*, 15th April 2009, Ostrava (Czech Republic). Available at: <http://hdl.handle.net/10084/71291>

MOOS, Marek, Ivo PRAJER and Petr KAREL. Integrace DSpace do IS ČVUT: technologický pohled [The Integration of DSpace and the CTU Information System: a Technological Viewpoint]. In: *2. setkání českých uživatelů systému DSpace [2nd Czech DSpace User Group Meeting]*, 15th April 2009, Ostrava (Czech Republic). Available at: <http://hdl.handle.net/10084/71266>