Meet RODA, a Full-Fledged Digital Repository for Long-Term Preservation

Rui Castro
KEEP SOLUTIONS
Rua Rosalvo de Almeida, nº 5
4710-029 Braga, Portugal
+351 253066735
rcastro@keep.pt

Luís Faria
KEEP SOLUTIONS
Rua Rosalvo de Almeida, nº5
4710-029 Braga, Portugal
+351 253066735
lfaria@keep.pt

Miguel Ferreira
KEEP SOLUTIONS
Rua Rosalvo de Almeida, nº5
4710-029 Braga, Portugal
+351 253066735
mferreira@keep.pt

ABSTRACT
RODA is an open-source full-fledged digital preservation repository capable of ingesting, managing and providing continuous access to various types of digital objects, namely text-documents, raster images, relational databases, video and audio.

It is supported by open-source technologies and makes use of existing standards such as the OAIS [1], METS [2], EAD [3] and PREMIS [4].

Categories and Subject Descriptors
H.3.7 [**INFORMATION STORAGE AND RETRIEVAL**]: Digital Libraries – *Collection, Dissemination, Standards, System issues, User issues.*

General Terms
Management, Standardization, Design, Security.

Keywords
Digital preservation, authenticity, digital archive, digital objects, open-source, digital repository.

1. Introduction
RODA is an open source digital repository specially designed for archives, with long-term preservation and authenticity as its primary objectives. Created by the Portuguese National Archives in partnership with the University of Minho, it was designed to support the most recent archival standards and become a trustworthy digital repository.

RODA was developed on top of Fedora Commons and embodies high-level standards of security, scalability and usability. Its centralized architecture enables an easy management while the auto-deposit tools and ingest workflow account for the scalability of the human-resources.

RODA is a complete digital repository system that provides functionality for all of the main units that compose the OAIS reference model. RODA fully implements an Ingest workflow that validates SIPs and migrates digital objects to preservation formats, and provides Access by delivering different ways to search and navigate over available data as well as visualizing and downloading stored digital material.

Data Management functionalities allow archivists to create and modify descriptive metadata and define rules for preservation interventions, e.g. scheduling integrity checks on stored digital objects or initiate a migration process.

Administration procedures allow the definition of access rights to data and operational permissions for each user or group.

In this demonstration we will explore all the functionality provided by RODA: ingest, access, description, management and preservation by means of its well designed graphical user interfaces (Figure 1).

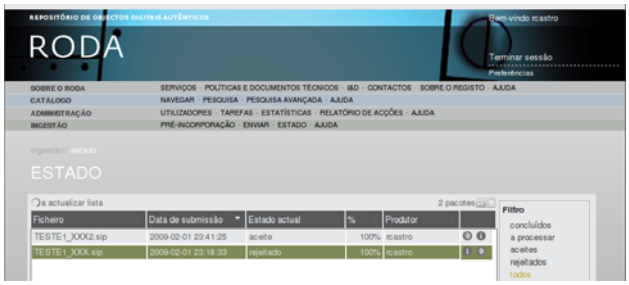

Figure 1 - Screenshot of RODA Web user interface.

2. Features to be demonstrated
In this section we describe the features of the system that will be demonstrated.

2.1 Ingest workflow
New entries to the repository come in the shape of Submission Information Packages (SIP). When the ingest process terminates, SIPs are transformed into Archival Information Packages (AIP), i.e. the actual packages that will be kept in the repository. Associated with the AIP is the structural, technical and preservation metadata, as they are essential for carrying out preservation activities.

The SIP is composed of one or more digital representations and all of the associated metadata, packaged inside a METS envelope. Producers take advantage of a small application called *RODA-in* that allows them to create these packages. The SIP is a

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
iPRES2011, Nov. 1–4, 2011, Singapore.
Copyright 2011 National Library Board Singapore & Nanyang Technological University

compressed ZIP file containing a METS envelope, the set of files that compose the representations and a series of metadata records. Within the SIP there should be at least one descriptive metadata record in EAD-Component format.

Before SIPs can be fully incorporated into the repository they are submitted to a series of tests to assess its integrity, completeness and conformity to the ingest policy. After decompressing the SIP, the validation process performs the following set of tasks: virus check, envelope syntax check, SIP completeness check, file integrity check, descriptive metadata check, preservation metadata check, representation check, specific representation check and format normalization. If all of these steps are completed successfully, the SIP waits for acceptance, which can be done manually or automatically (in batch mode).

After a successful validation, the SIP is then taken apart and each of its constituents is integrated into the repository. After this procedure, the SIP becomes an AIP and is ready to be disseminated to potential consumers that have clearance to access that information.

2.2 Access

Consumers are able to browse over available collections to view or download digital representations stored in the repository.

Depending on the type of the digital object, different viewers/disseminators are used. For example, text documents are delivered to consumers using a flash-based page-flip Web application and as a download so the consumer can use its favourite PDF viewing application. Representations composed of several images (e.g. digitised works) are displayed in a special Web viewing application that allows consumers to navigate through the pages of the representation.

2.3 Data management and archival storage

Because RODA was developed to be used by the Archival community, the underlying descriptive metadata supported by the repository is EAD/XML (Encoded Archival Description).

RODA's content model is atomistic and very much PREMIS-oriented. Each intellectual entity is described by an EAD-component. These metadata records are organized hierarchically in order to constitute a full archival description, but are kept separately within the Fedora Commons content model.

PREMIS entities are used intensively as they shape the overall content model of the repository. Archival storage is supported entirely by Fedora Commons.

2.4 Administration and Preservation planning

Administration allows for fine-grained control of user access to information and functionalities. Permissions can be granted to users or groups to perform certain actions within the repository and to access certain data or metadata objects.

All actions performed in the repository by any user are logged *ad aeternum* for security and authenticity reasons. Every user must be authenticated in order to use the repository.

Preservation management within RODA is handled by scheduled events. A preservation expert defines the set of rules that trigger specific preservation actions and when these should take place. Preservation actions comply to a common API, so creating and installing new preservation actions in the repository is as easy as copying the programme file to the correct directory on the server. Preservation actions include format migrators, checksum verification tools, reporting tools (e.g. to automatically send SIP acceptance/rejection emails), etc.

Each preservation event that takes place inside the repository is recorded as preservation metadata. Special events like format migrations establish relationships between representation nodes and record agents used and event outcomes automatically.

3. Architecture

RODA's architecture is service oriented and it's composed of 3 main components: RODA WUI, RODA Core Services, and RODA Migration Services.

RODA's Web User Interface (RODA-WUI) handles all the aspects of the graphic user interface for producers, consumers, archivists, system administrators and preservation experts. The RODA-WUI components are supported by AJAX and Web services. RODA WUI is a client of RODA Core Services.

RODA Core Services are responsible for carrying out more complex tasks than the ones offered by Fedora Commons such as the complete set of procedures that compose the ingest workflow, querying the repository in more advanced and abstract ways and carrying out administrative functions on the repository.

For performing format migrations the Core Services rely on RODA Migration services. RODA Migration services are several web services that perform format conversions on dedicated remote servers.

4. Future and current work

The RODA team is always engaged in addressing new challenges, and to go beyond the state-of-the-art. RODA is now part of the SCAPE project, an international project with several European national libraries and universities with the objective of enhancing the state of the art of digital preservation in three ways: by developing infrastructure and tools for scalable preservation actions; by providing a framework for automated, quality-assured preservation workflows; and by integrating these components with a policy-based preservation planning and watch system. Currently, the team is focusing in scalability, automatic monitoring of the world and preservation planning integration.The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

5. REFERENCES

[1] Consultative Committee for Space Data Systems (CCSDS), *Reference Model for an Open Archival Information System (OAIS)*. Consultative Committee for Space Data Systems (CCSDS), 2002.

[2] "Metadata Encoding and Transmission Standard (METS) Official Web Site." [Online]. Available: http://www.loc.gov/standards/mets/. [Accessed: 09-Aug-2011].

[3] "EAD: Encoded Archival Description Version 2002 Official Site (EAD Official Site, Library of Congress)." [Online]. Available: http://www.loc.gov/ead/. [Accessed: 09-Aug-2011].

[4] PREMIS Editorial Committee, *Data Dictionary for Preservation Metadata: PREMIS version 2.0*. 2008.