RODA
Repository of Authentic Digital Objects

PresDB’07
23 of March 2007

Francisco Barbedo
frbarbedo@iantt.pt

José Ramalho
jcr@di.uminho.pt

Luís Faria
lfaria@iantt.pt

Miguel Ferreira
mferreira@dsci.uminho.pt

Rui Castro
rcastro@iantt.pt

Luís Corujo
lcorujo@iantt.pt

RODA is a project from the National Archives of Portugal
About RODA

- Open-source
- For archivists
- Storage
- Continued Access
- Metadata Management
- Preservation
- Authenticity

RODA is a project for the implementation of a repository that guarantees the storage of digital objects, the continued access to them, the management of their metadata, and the preservation and authenticity of the digital objects in the context of a digital archive. The distinction between archives and libraries is very important because there is much done for digital libraries, but a lot less for digital archives.
RODA follows the OAIS model.
A repository in an archival context must follow a different data model than in a librarian context. EAD vs. DC, hierarchical vs. plain descriptive metadata
Use of Fedora
RODA 1.0 is just a prototype. It will only give support to still images, structured text and relational databases.
Relational Databases
in RODA 1.0

Long term archival – provided by the RODA repository
Authenticity and Provenance – provided by the RODA repository, specifically the preservation metadata (and it being a trusted repository)
Relational Databases in RODA

- Long term archival
- Authenticity and Provenance
- Separate data from a specific DBM
- Preserve data and structure
- Preserve semantics
- Scalability
- Preserve evolving data
- Distributed model

Long term archival – provided by the RODA repository
Authenticity and Provenance – provided by the RODA repository, specifically the preservation metadata (and it being a trusted repository)
Create an abstraction for the database, which is independent from the logic used: DBML
DBML

- Non-proprietary
- Platform and RDBMS independent
- XML language
- Stores the DB structure and information
- BLOBs are exported and preserved as stand-alone files in the representation
- Transformations to SQL and back are defined

*More information about DBML at http://hdl.handle.net/1822/601*

Separate data from a specific DBMS: Create an abstract representation of it.  
Preserve data and structure: using a declarative markup language  
Preserve semantics: Its not possible to keep the semantics without keeping the processing engine
DBML

- Non-proprietary
- Platform and RDBMS independent
- XML language
- Stores the DB structure
- BLOBs are exported and preserved as standalone files in the representation
- Transformations to SQL

*More information about DBML at [http://hdl.handle.net/1822/601](http://hdl.handle.net/1822/601)

Separate data from a specific DBMS: Create an abstract representation of it.
Preserve data and structure: using a declarative markup language
Preserve semantics: It's not possible to keep the semantics without keeping the processing engine
- Long term archival
- Authenticity and Provenance
- Separate data from a specific DBMS
- Preserve data and structure
- Preserve semantics
- Scalability
- Preserve evolving data
- Distributed model
An application – the SIP Creator – was implemented to help producers create Submission Information Packages (SIP) from their databases.
The SIP then can enter the repository, as described in the OAIS. The repository will guarantee the preservation of the database abstraction.
To provide the access from the consumer, the transformation to SQL is used, and the same SQL is injected into a state of the art RDBMS.

Scalability: using cache the dissemination we can get some scalability, depending on the cache size and the RDBMS efficiency.
To provide the access from the consumer, the transformation to SQL is used, and the same SQL is injected into a state of the art RDBMS.

Scalability: using cache the dissemination we can get some scalability, depending on the cache size and the RDBMS efficiency.
Future work

- RODA 2.0 from April 2007 until April 2008
- Full implemented solution
- Support other object classes
- Support for complex workflow (e.g. ingest)
- Full support of preservation events
- Data centre (vendor independent, scalable)

Preserve evolving data: the data in such repository is frozen. As data evolves, new intellectual entities can be created, like snapshots. Further research will be done on this subject next year. Distributed model: it is possible, but distributions raises other issues like synchronisation. Further research will be done next year.
Future work

 nạn 2.0 from April 2007 until April 2008

Full implemented solution

Support other object classes

Support for complex workflow (e.g., ingest)

Full support of preservation events

Data centre (vendor independent, scalable)

Long term archival

Authenticity and provenance

Separate data from a specific DBM

Preserve data and structure

Preserve semantics

Scalability

Preserve evolving data

Distributed model

Preserve evolving data: the data in such repository is frozen. As data evolves, new intellectual entities can be created, like snapshots. Further research will be done on this subject next year. Distributed model: it is possible, but distributions raises other issues like synchronisation. Further research will be done next year.
RODA
Repository of Authentic Digital Objects

http://roda.iantt.pt

For more information visit the web site.