Research projects as a driving force for open source development and a fast route to market

RODA, SCAPE & E-ARK - a case study

Hélder Silva (Miguel Ferreira & Luís Faria)
KEEP SOLUTIONS
Outline

• What is RODA?
• RODA current problems
• Research initiatives & results
• Future research
• Route to market
• Final thoughts
What is RODA?
A digital repository specially designed for archives, with the following main features:

- Long-term Preservation and Authenticity
- Based on standards (OAIS, METS, EAD, PREMIS, etc.)
- Secure (fine-grain permissions, anonymous not allowed, LDAP)
- Scalable architecture (SOA)
- Clean web user interface and ingest desktop tool
- For archivists, for producers, for consumers
- Open source
Architecture
RODA Web User Interface
RODA In
RODA current problems
OAIS
RODA SIP Format

- Object class and format
- Place in classification plan
- Descriptive metadata
- Preservation metadata
- Technical metadata
- Representation files
Massive SIP creation (& upload)
OAIS
RODA preservation actions (set of plug-ins, more can be added easily):

• Fixity check
• File format migration
• Etc.

But...

• Well founded decision for selecting optimal preservation task must be made
• Environment monitoring must be carried out
Research initiatives & results
Advanced preservation planning

• C3PO - Inspect content features
• Scout - Find preservation risks
• Plato - Create preservation plans
• Taverna - Executable workflows
• RODA - Execute Plato action plans
OAIS - Preservation planning

...task responsible for “monitoring the environment of the OAIS and which provides recommendations and preservation plans to ensure that the information stored in the OAIS remains accessible to, and understandable by, and sufficiently usable by, the Designated Community”...
C3PO - Inspect content characteristics

What is c3po?

C3PO is a content profiling tool used during preservation planning. It uses the meta data extracted from digital objects and aggregates it in order to support the analysis during preservation planning activities. Furthermore, C3PO provides representative samples from collections and it offers different data export facilities, so that planners can analyze the content even more and obtain deeper knowledge of the objects they are trying to preserve.

If you want to find out more, please navigate here. And if you want to provide us feedback, please don't hesitate contacting us at c3po@ifs.tuwien.ac.at

Usage

Navigate to the overview tab and choose a collection by clicking on the first image button in the menu bar. This will present you some diagrams and basic overview of the selected collection. By clicking on the desired values of the diagrams, you will start filtering the collection. The filter is conveniently stored for you as you drill down the content set.

Filtering

C3PO allows you to filter the data based on some property value conditions. You can easily partition the selected collection based on some simple criteria. For example if you want to see all objects created in year 2007, you can just click on the filter button in the menu. The filter view will slide in and you can click on the big green '+' button. Choose the 'created' property (or any other) and C3PO will present you all years in your collection. When you choose a year (e.g. 2007), then C3PO will find all objects that had the created meta data field with the value 2007 and will recreate the overview only for this partition. If you want to add more conditions, just hit the '+' button again.
Scout - Find preservation risks

What is Scout?
Scout is a digital preservation watch system that allows you to be notified whenever some significant event in the world can bring you harm or profit.

How does it work?
It gathers information from all around the world and curates it using link data technology. It can fetch information from your content, organization policies, format and tool registries, the Web, and even human knowledge, putting all on a centralized and well defined knowledge base.

Then you can define questions to identify preservation risks. These questions can cross-relate information from different sources.
Plato - Create preservation plans

Planning

- Define requirements
- Evaluate alternatives
- Analyse results
- Build preservation plan

Preservation plan

What is Plato?

Digital content is short-lived, yet may prove to have value in the future. How can we keep it alive? Finding the right action to enable future access to digital content in a transparent way is the mission of digital preservation.

The mission of digital preservation is to ensure continued, authentic long-term access to digital objects for a wide range of communities. This requires preservation actions to be carried out when the original environment of digital content characterisation, preservation action and automatic object comparison in a service-oriented architecture is a key aspect.

The mission of preservation planning is to ensure authentic future access for a specific set of objects, defining the actions needed to preserve it.

The planning tool Plato is a decision support tool that implements a solid preservation planning process, including content characterisation, preservation action and automatic object comparison in a service-oriented architecture.

What's new?

September 2014: Best Demonstration at Digital Libraries 2014

We are happy to announce that The SCAPE preservation lifecycle was awarded Best Demonstration at Digital Libraries 2014.
Taverna - Execute the plan

Taverna Workflow Management System
Powerful, scalable, open source & domain independent tools for designing and executing workflows. Access to 3500+ resources.

RECENT NEWS
- Starting now – the Taverna Open Development Workshop
- Taverna has been accepted as an Apache Incubator project
- Taverna Open Development Workshop
2014-10-30 / 2014-10-31

Taverna Workbench
Server
Player
Command Line
Taverna Online

Taverna is an open source and domain-independent Workflow Management System – a suite of tools used to design and execute scientific workflows and aid in silico experimentation.

Taverna has been created by the myGrid team and is currently funded though FP7 projects BioVeL, SCAPE and Wf4Ever.
Advanced preservation planning

RODA
preservation lifecycle

Plan
- Plan Management API
- Data Connector API

Monitor
- Report API

Act
Future research
RODA - full scale pilot

- Implement pan-European SIP & DIP
- Improve ingest process
- Improve access to content (specially databases)
- Integrate RODA with live Record Management System
Route to market
SCAPE outputs

• C3PO, Scout, Plato, Taverna are external services

• Integrated with RODA via 3 new APIs:
  – Repository independent
  – Use standards (e.g. OAI-PMH, PREMIS, Dublin Core, METS)
  – Made none or the least changes possible in the underlying data model

This means: they will be merge into the main source code
E-ARK outputs

• Changes can have much deeper impact
• But we are willing to take that risk
• This way the outcomes of the project get to the end user

This means: they will be merge into the main source code
Final thoughts
• To be successful, be open to changes
• Open source and standards based approaches facilitate those changes
• Using available infrastructures (e.g. GitHub) and founding accelerates the access to new developments
• But planning and design is needed to facilitate the integration of new developments into the main source

Embrace changes, but plan well and communicate!
Want to know more about RODA?
http://www.roda-community.org
Research projects as a driving force for open source development and a fast route to market

RODA, SCAPE & E-ARK - a case study

Hélder Silva / hsilva@keep.pt
KEEP SOLUTIONS / www.keep.pt / info@keep.pt