

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?

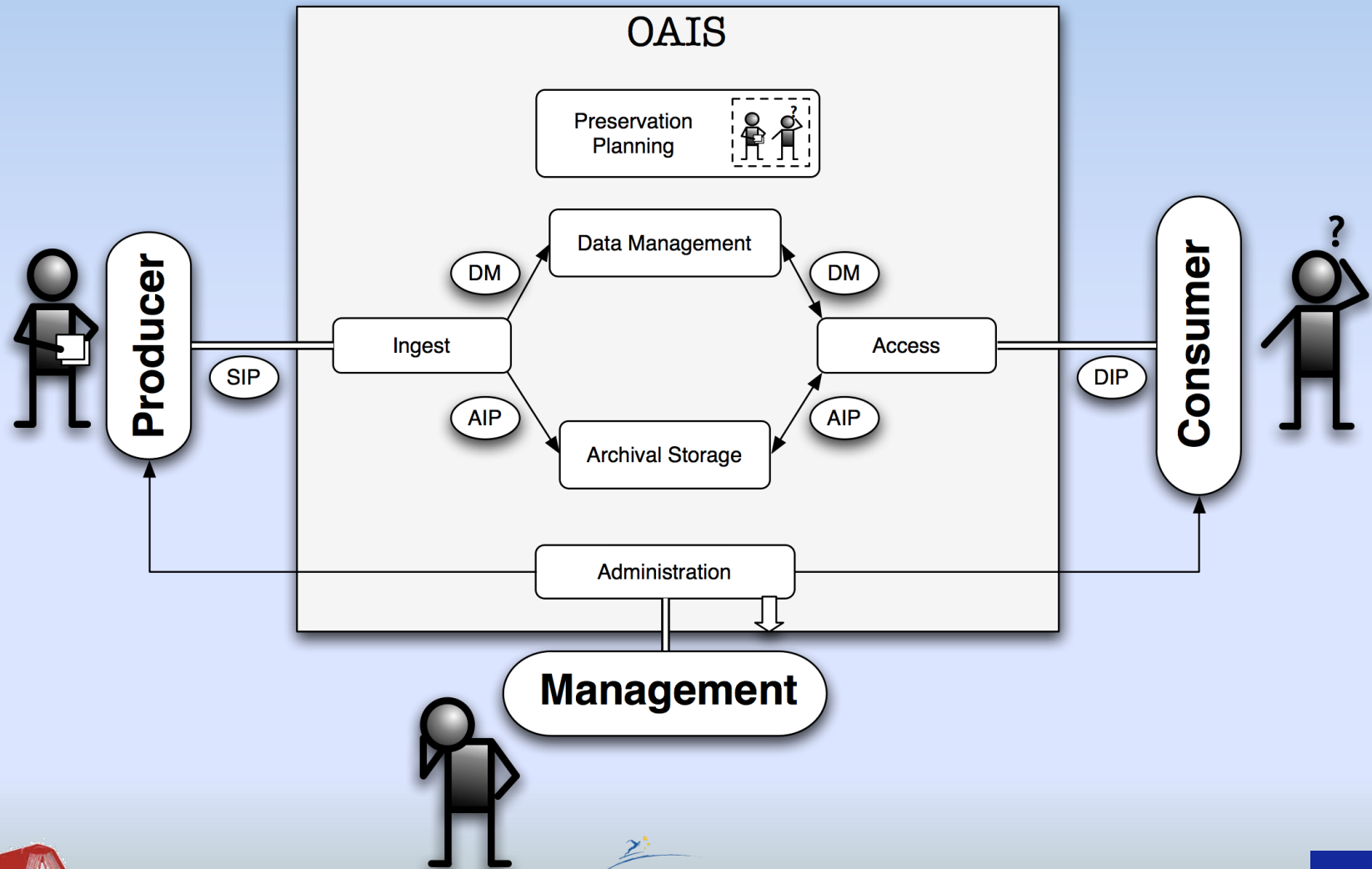


A digital repository specially designed for archives, with the following main features:

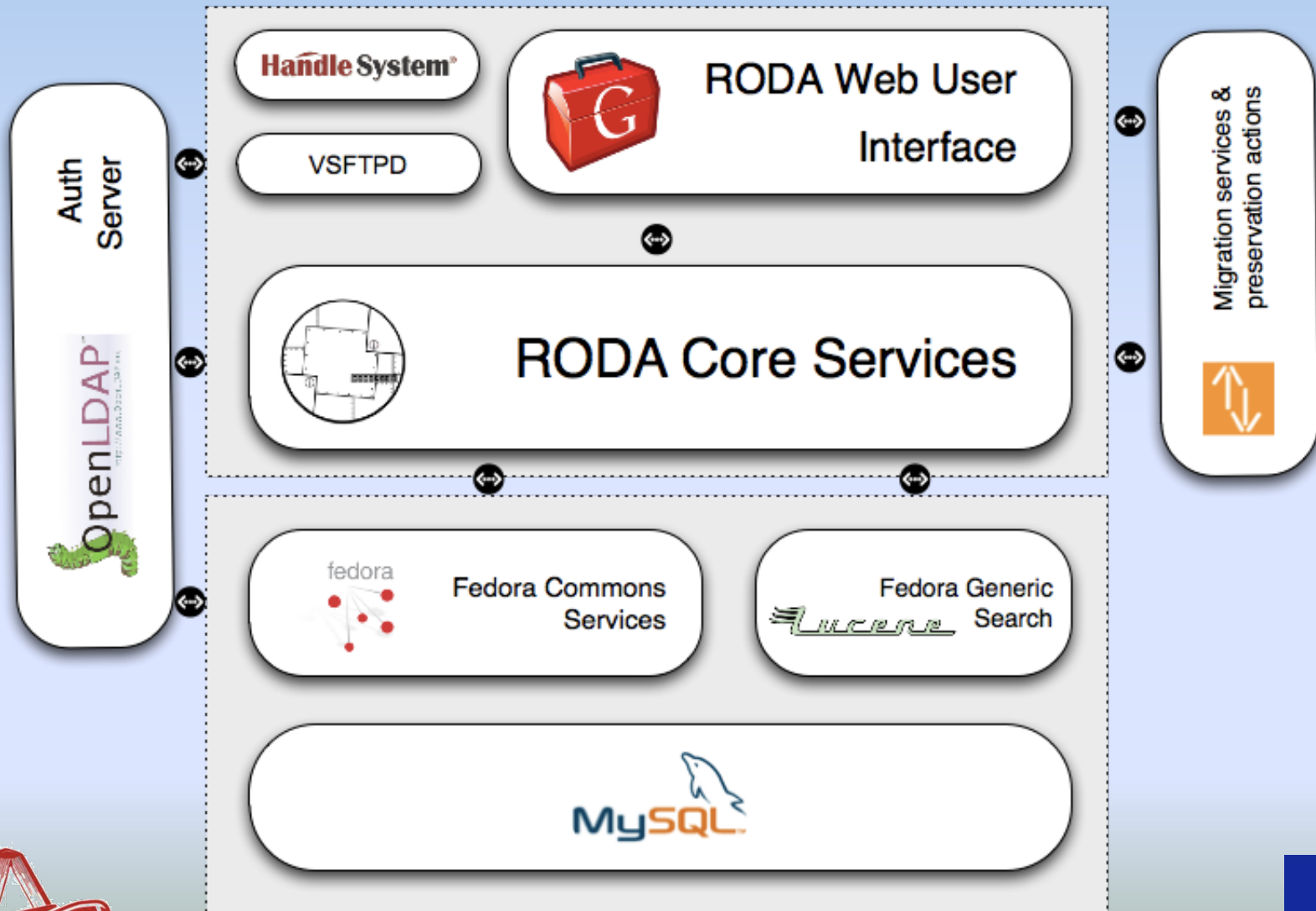
- Long-term Preservation and Authenticity
- Based on standards (OAIS, METS, EAD, PREMIS, etc.)
- **Secure** (fine-grain permissions, anonymous not allowed, LDAP)
- Scalable architecture (SOA)
- Clean web user interface and ingest desktop tool
- For archivists, for producers, for consumers
- Open source



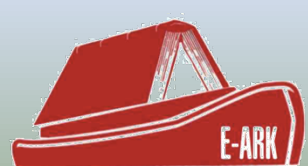
OAIS



Architecture



↔ Communication interface



RODA Web User Interface

The screenshot displays the RODA Web User Interface in a browser window. The address bar shows the URL 192.168.2.56. The page header includes the text "REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS" and the RODA logo. A navigation menu is visible with categories: ABOUT, DISSEMINATION, INGEST, and ADMINISTRATION, each with sub-links. The main content area is titled "BROWSE" and shows "3 fonds" with a filter for "IMAGES". A table displays the details for the selected fund, including identification and content structure information.

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

Welcome admin

Logout

Preferences

ABOUT SERVICES · POLICIES · I&D · CONTACTS · REGISTER · HELP

DISSEMINATION BROWSE · BASIC SEARCH · ADVANCED SEARCH · HELP

INGEST PRE-INGEST · SUBMIT · STATUS · HELP

ADMINISTRATION USER MANAGEMENT · SCHEDULER · STATISTICS · LOG · HELP

dissemination / browse / 35

BROWSE

3 fonds F +

Reference

- F FONDS
- F JISC
- F NATURE
- C IMAGES

Description	Permissions
Identification	
Reference	IMAGES
Complete reference	PT/KEEPS/NATURE/IMAGES
Handle	http://hdl.handle.net/10384/35
Title	IMAGES
Level	class
Initial date	2014-01-01
Final date	2016-01-01
Country code	PT
Repository code	KEEPS
Origination	<undefined>
Content and structure	
Scope and content	<undefined>



RODA In

The screenshot displays the RODA In 1.1.0 web interface. At the top, the header reads "REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS" and "RODA". Below the header, there are "Update plan" and "Send" buttons. The main content area is titled "FONDS/BNER_1881_07_09" and has two tabs: "Description" (selected) and "Representation".

On the left, a tree view shows the repository structure under "FONDS":

- BNER_1881_07_09 (selected)
- MISC_DATA_PART_1
- MISC_DATA_PART_2
- MISC_DATA_PART_3
- MISC_DATA_PART_4
- MISC_DATA_PART_5
- MISC_DATA_PART_6
- MISC_DATA_PART_7
- MISC_DATA_PART_8
- MISC_DATA_PART_9
- MISC_DATA_PART_10
- BNER_1882_02_22
- BNER_1883_12_08
- BNER_1884_04_12
- BNER_1885_03_25

The main record details are as follows:

Identification	
Reference:	BNER_1881_07_09
Title:	Banner and Times of Wales
Description level:	Item
Initial date:	1881-07-09
Final date:	1881-07-09
Country code:	GB
Repository code:	BL
Producer:	JISC

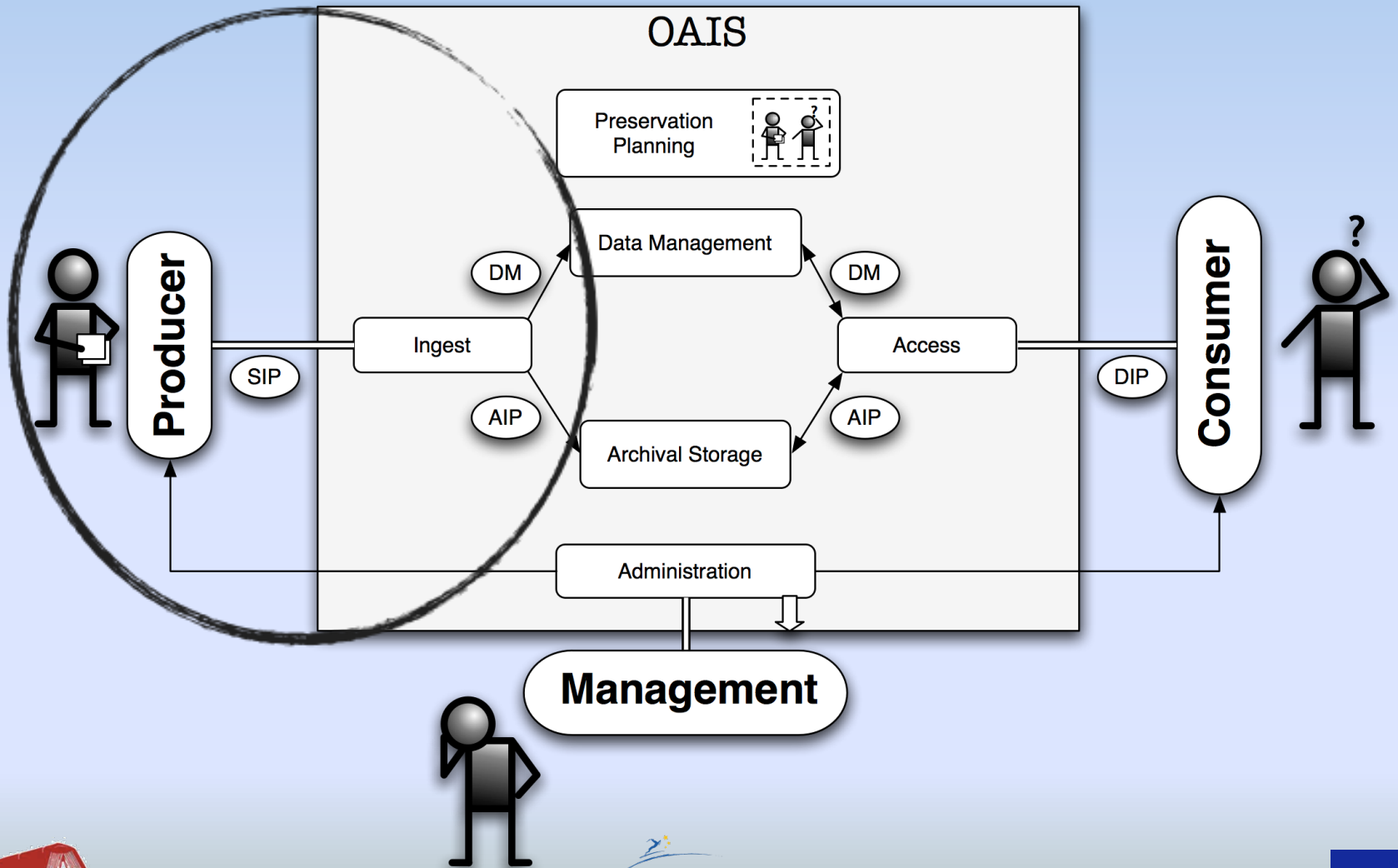
Context	
Scope and content:	Demo

At the bottom, there is a toolbar with buttons: "Create", "Remove", "Save", "Validate", "Create simple document", "Remove simple document", "Create representation", and "Remove representation".

RODA current problems

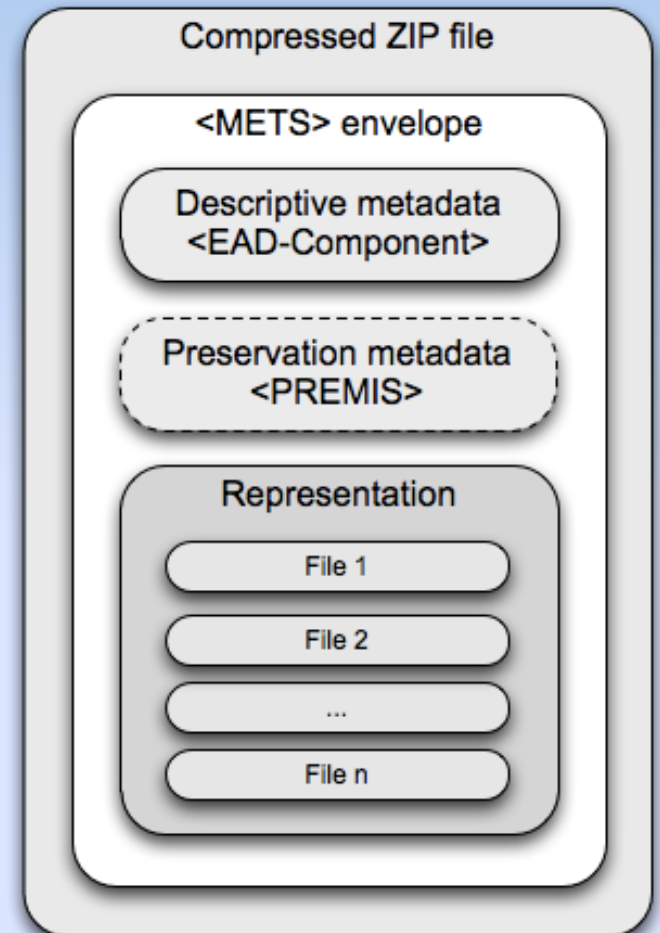


OAIS



RODA SIP Format

- Object class and format
- Place in classification plan
- Descriptive metadata
- Preservation metadata
- Technical metadata
- Representation files



Massive SIP creation (& upload)

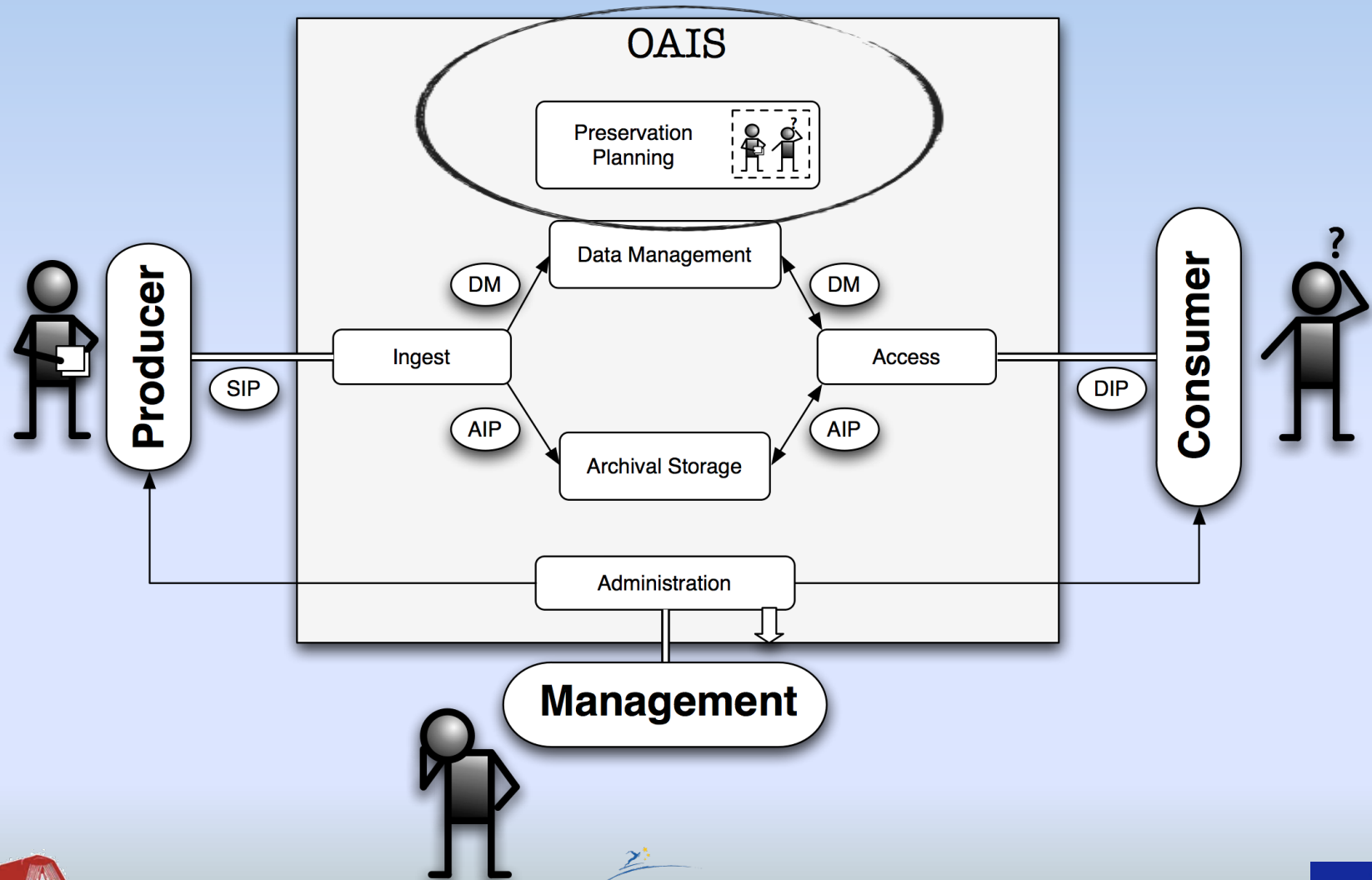
The image shows a screenshot of the RODA SIP Creator web application. The interface is divided into several sections:

- Top Bar:** "REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS" and "RODA" logo.
- Left Panel (CREATE):** Contains "Send packet" and "Create packet" buttons. Below is a "Fill in the descriptive metadata" form with sections for "Identification" and "Content and structure".
 - Identification:** Reference, Title, Level (item), Initial date (january), Final date (january), Country code (PT), Repository code (KEEPS), Origination.
 - Content and structure:** Scope and content (<undefined>).
- Center Panel:** A tree view showing a hierarchy of objects under "FONDS".
 - FONDS
 - BNER_1881_07_09 (selected)
 - MISC_DATA_PART_1
 - MISC_DATA_PART_2
 - MISC_DATA_PART_3
 - MISC_DATA_PART_4
 - MISC_DATA_PART_5
 - MISC_DATA_PART_6
 - MISC_DATA_PART_7
 - MISC_DATA_PART_8
 - MISC_DATA_PART_9
 - MISC_DATA_PART_10
 - BNER_1882_02_22
 - BNER_1883_12_08
 - BNER_1884_04_12
 - BNER_1885_03_25

- Right Panel:** "FONDS/BNER_1881_07_09" details view with tabs for "Description" and "Representation".
- Identification:** Reference: BNER_1881_07_09, Title: Banner and Times of Wales, Description level: Item, Initial date: 1881-07-09, Final date: 1881-07-09, Country code: GB, Repository code: BL, Producer: JISC.
- Context:** Scope and content: Demo.
- Bottom Bar:** "Create", "Remove", "Save", "Validate", "Create simple document", "Remove simple document" buttons.

RODA SIP Creator

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

c3po Clever, Crafty Content Profiling of Objects

Home Overview Objects Samples Export

Fork me on GitHub

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.

feedback

Scout - Find preservation risks

Scout Home Query Browse Login



Scout

A preservation watch system


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.

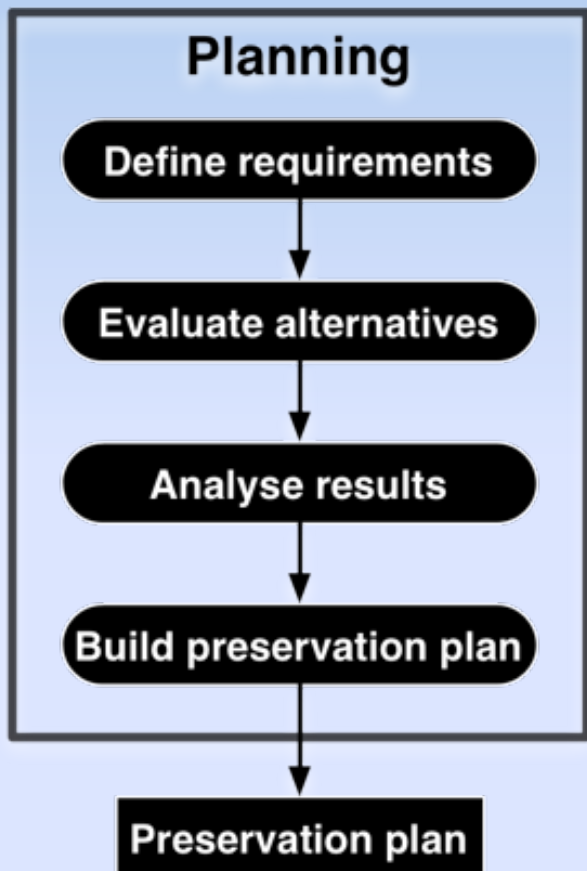
How does it work?

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



Plato - Create preservation plans



Plato

[Introduction](#) [Documentation](#) [Cite](#) [Case Studies](#) [Events](#) [Media](#) [History](#) [Team](#)

[Enter Plato 4](#)

What is Plato?

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

The mission of digital preservation is to ensure continued, authentic long-term access to digital objects for communities. This requires preservation actions to be carried out when the original environment of digital objects no longer exists, but each shows specific peculiarities, and a variety of factors influence the

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

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

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 D



Taverna - Execute the plan



The screenshot shows the Taverna website homepage. At the top left is the Taverna logo, a circular gear icon with the word "Taverna" next to it. To the right is the "myGrid" logo, a 3D cube structure. Further right is a search bar with the text "Google™ Custom Search". Below the logo and search bar is a navigation menu with the following items: Introduction, Documentation, Download, Developers, Cite, Collaborations, News, and About. The main content area features a large dark box with the text "Taverna Workflow Management System" and a sub-headline "Powerful, scalable, open source & domain independent tools for designing and executing workflows. Access to 3500+ resources." To the right of this box is a "RECENT NEWS" section with three bullet points: "Starting now – the Taverna Open Development Workshop", "Taverna has been accepted as an Apache Incubator project", and "Taverna Open Development Workshop 2014-10-30 / 2014-10-31". Below the main content area are five buttons: "Workbench", "Server", "Player", "Command Line", and "Taverna Online".

Taverna

myGrid

Google™ Custom Search

Introduction Documentation Download Developers Cite Collaborations News About

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

Workbench Server Player Command Line Taverna Online

COMMUNITY

- Taverna for astronomy, bioinformatics, biodiversity, digital preservation
- Workflow components
- Taverna 3 OSGi
- Taverna Online
- Next generation sequencing on Amazon cloud
- Taverna-Galaxy

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.

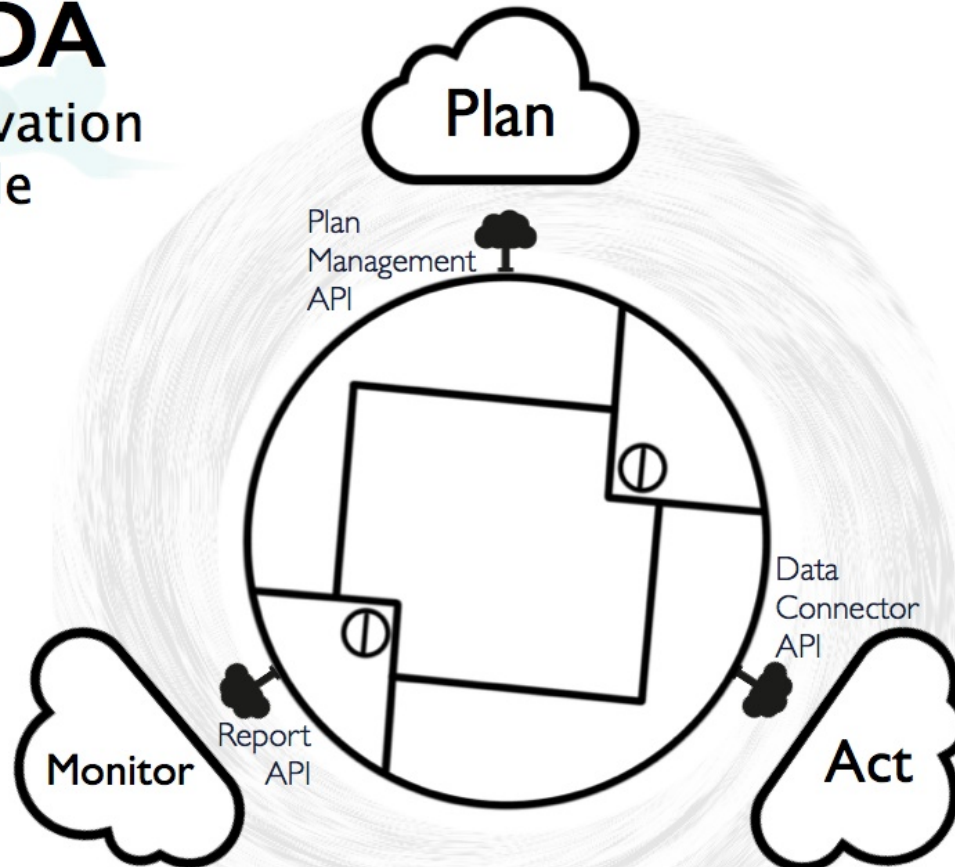
Taverna has been created by the [myGrid team](#) and is currently funded through FP7 projects [BioVeL](#), [SCAPE](#) and [Wf4Ever](#).

See Taverna in action



Advanced preservation planning

RODA preservation lifecycle



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



https://github.com/keeps/roda

keeps / roda

Unwatch 8 Unstar 9 Fork 6

RODA - Repository of Authentic Digital Objects <http://www.roda-community.org>

64 commits 2 branches 2 releases 4 contributors

branch: master roda / +

converting i18n files encoded in iso latin1 to utf8

luis100 authored 11 days ago latest commit 92d9d5d181

code-style	Renamed AIPDownload to RepresentationDownload (and all the repercussi...	a year ago
roda-client	Renamed AIPDownload to RepresentationDownload (and all the repercussi...	a year ago
roda-common	Renamed AIPDownload to RepresentationDownload (and all the repercussi...	a year ago
roda-core	Renamed AIPDownload to RepresentationDownload (and all the repercussi...	a year ago
roda-handle	Renamed AIPDownload to RepresentationDownload (and all the repercussi...	a year ago
roda-installer	Update roda-wui.properties	a year ago
roda-migrator	Renamed AIPDownload to RepresentationDownload (and all the repercussi...	a year ago
roda-ui	converting i18n files encoded in iso latin1 to utf8	11 days ago
.gitignore	added *~ to gitignore	a year ago
.travis.yml	adding travis-ci config	a year ago
CHANGELOG.txt	Update CHANGELOG.txt	a year ago
LICENSE.txt	Initial import of RODA into GitHub. Fixes #2, fixes #3 and fixes #4.	2 years ago
README.md	Update README.md	a year ago

Code

Issues 39

Pull Requests 0

Wiki

Pulse

Graphs

SSH clone URL

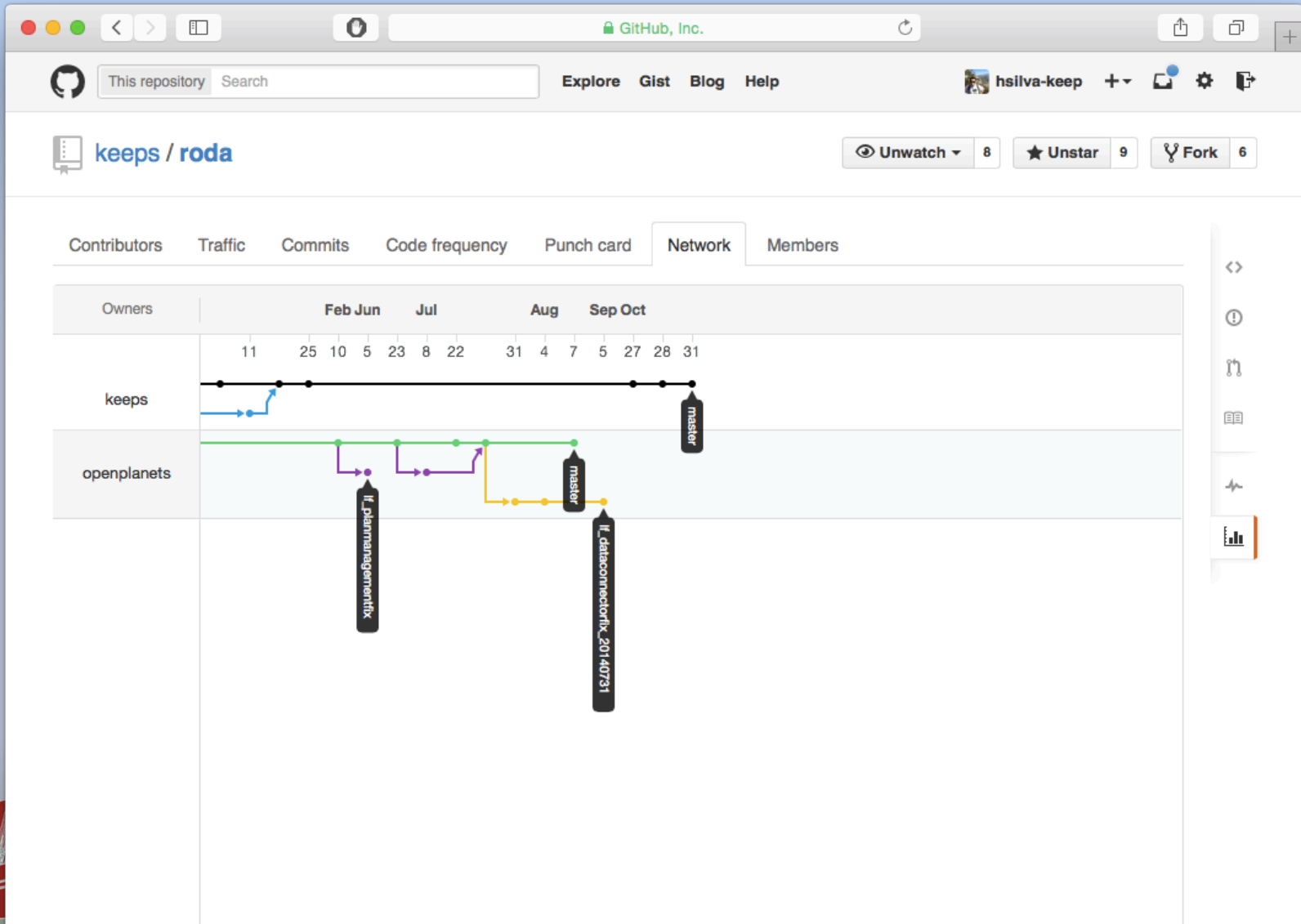
git@github.com:keep:

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

Clone in Desktop

Download ZIP

https://github.com/keeps/roda



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

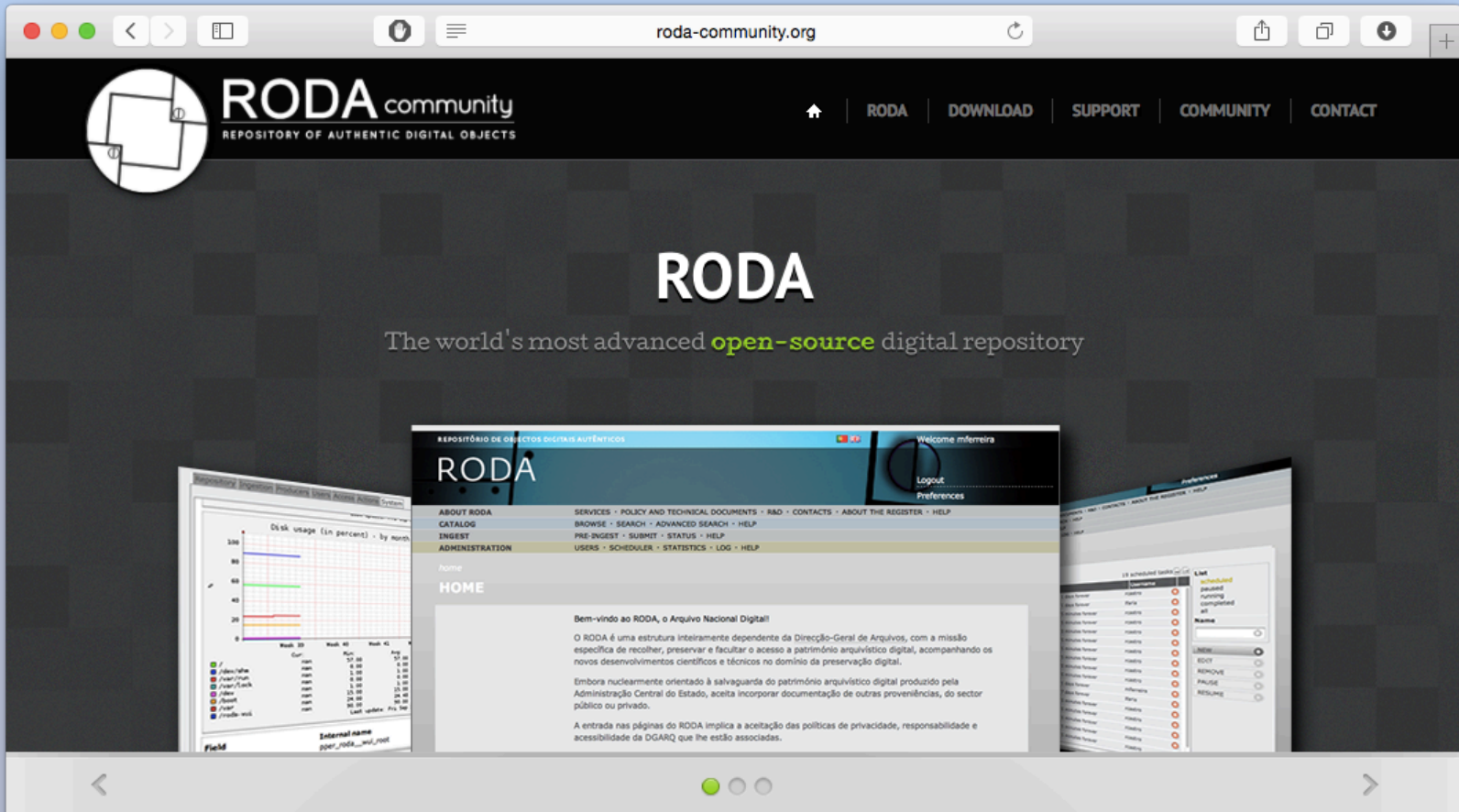
Embrace changes, but plan well and communicate!



Want to know more about RODA?



http://www.roda-community.org



An open-source **digital repository** designed for preservation

RODA is a complete digital repository that delivers functionality for all the main units of the OAIS reference model. RODA is capable of ingesting, managing and providing access to the various types of digital content produced by large corporations or public bodies. RODA is based on open-source technologies and is



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

