RODA
An open-source digital repository designed for preservation
What is RODA?
A digital repository designed for archives with the following characteristics:

- **Open-source**
- Long-term **Preservation and Authenticity**
- Based on standards: EAD, PREMIS, OAIS, METS, ...
- **Secure**
  - fine-grain permissions, no anonymous usage, LDAP
- **Scalable** architecture (SOA)
- Clean web user **interface**
- Vendor independent
Built on top of Fedora Commons
OAIS compatible
but... who isn't anyway?
Atomistic data model
Supported media types

- Text documents (PDF)
- Video (MPEG-2)
- Audio (Wav)
- Digitised works (TIFF unc.)
- Relational databases (DBML)
Supported media types

- Text documents (PDF)
- Video (MPEG-2)
- Audio (Wav)
- Digitised works (TIFF unc.)
- Relational databases (DBML)
- Spreadsheets (CSV, PDF)
- Presentations (PDF)
- E-mail
All other media types can also be ingested
All other media types can also be ingested
But will only be preserved at the bit level
# Default ingest workflow

<table>
<thead>
<tr>
<th>Step</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Upload</td>
<td>Upload SIP to the ingest stage area by HTTP, FTP or local copy</td>
</tr>
<tr>
<td>2</td>
<td>Unzip</td>
<td>Open ZIP package</td>
</tr>
<tr>
<td>3</td>
<td>Virus Check</td>
<td>Check all files for viruses with Clam Anti Virus</td>
</tr>
<tr>
<td>4</td>
<td>Syntax and fixity check</td>
<td>Check if metadata is well formed and integrity of all files against checksums in METS envelope</td>
</tr>
<tr>
<td>5</td>
<td>Authorization check</td>
<td>Check if user has authorization to ingest into the defined classification plan</td>
</tr>
<tr>
<td>6</td>
<td>Ingest</td>
<td>Ingest original into Fedora. This ingestion is not final, can be undone in next steps.</td>
</tr>
<tr>
<td>7</td>
<td>Normalization</td>
<td>Check if representation needs normalization, and enforce it as needed</td>
</tr>
<tr>
<td>8</td>
<td>Semantic check</td>
<td>Manual semantic check by an archivist, then object is marked active, indexed and published.</td>
</tr>
</tbody>
</table>
RODA ingest tool
### STATUS

<table>
<thead>
<tr>
<th>File</th>
<th>Submission date</th>
<th>Current state</th>
<th>%</th>
<th>Producer</th>
<th>Filter</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEST_sf1_c1_sr1_NG.sip</td>
<td>2009-05-11 14:41:47</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_sf1_c1_sr1_fox.sip</td>
<td>2009-05-07 14:19:36</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_MP48298.sip</td>
<td>2009-03-20 17:01:30</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_PNG2.sip</td>
<td>2009-03-20 15:06:14</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_PNG492.sip</td>
<td>2009-03-20 15:05:28</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_XPM.sip</td>
<td>2009-03-20 11:40:25</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_WAV.sip</td>
<td>2009-03-20 11:40:24</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_TXT.sip</td>
<td>2009-03-20 11:40:16</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_TIFF.sip</td>
<td>2009-03-20 11:40:15</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_TGA.sip</td>
<td>2009-03-20 11:40:15</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_RTF.sip</td>
<td>2009-03-20 11:40:14</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_PDF.sip</td>
<td>2009-03-20 11:40:13</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_OGG.sip</td>
<td>2009-03-20 11:40:06</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_ODT.sip</td>
<td>2009-03-20 11:40:05</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_MPC3.sip</td>
<td>2009-03-20 11:40:01</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_JPEG.sip</td>
<td>2009-03-20 11:39:28</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_ICO.sip</td>
<td>2009-03-20 11:39:45</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_GIF.sip</td>
<td>2009-03-20 11:39:24</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_GFARIA_RAG.sip</td>
<td>2009-03-20 11:39:28</td>
<td>accepted</td>
<td>100%</td>
<td>Ifaria</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Ingest report**

<table>
<thead>
<tr>
<th>Original file name: TEST_sf1_c1_sr1_NG.sip</th>
</tr>
</thead>
<tbody>
<tr>
<td>received (HTTP)</td>
</tr>
<tr>
<td>unpacked</td>
</tr>
<tr>
<td>virus-free</td>
</tr>
<tr>
<td>well formed</td>
</tr>
<tr>
<td>authorized</td>
</tr>
<tr>
<td>ingestd</td>
</tr>
<tr>
<td>normalized</td>
</tr>
<tr>
<td>accepted</td>
</tr>
</tbody>
</table>

**Date and time:** 2009-05-11 15:10:44.000

**Task:** Semantic Check  
**Result:** Success  
**Details:** SIP aceite pelo utilizador Ifaria: SIP valid for demoSIP objects marked active [roda.dgarq.gov.pt]
Access
All the words "tests" must appear in one of the fields Scope and content of the results.
BROWSE

Disseminations of 50218 - "National Geographic Photos"

JPEG (original)

TIFF (normalized)

Photo pre-viewer

Book pre-viewer
Description: The objects inside the SIP were added to the preservation repository.

Result: success

Ingest details:
- ingestion
- normalization
Lets talk about preservation, shall we?
Normalisation on ingest
Configurable normalisation rules

Automatic generation of preservation metadata
PREMIS events, representation information, agents

Preservation plans can be executed on the demand
New preservation components are easily installed as add-ons

Original representations are kept unchanged
Enables future approaches, e.g. emulation on-demand

Fixity checks
Notifications are sent when bit rot is detected
Schedule task

Name
Description

Start date
- Now
- Schedule: January

Repeat
- no repeat
- repeat each: seconds until: forever

Plugin
- Converter/AIFF to WAV (version 1)
- Converter/RTF to PDF/A (version 1)
- Converter/TGA to TIFF (version 1)
- Converter/TIFF to TIFF (version 1)
- Converter/WAV to WAV (version 1)
- Converter/WMV to MPEG2 (version 1)
- Converter/XPM to TIFF (version 1)
- Description tools/Auto fill unit dates (version 1)
- Description tools/Notify access restrict expiration (version 1)
- Ingest Maintenance/Reject 'AUTHORIZED' SIPs (version 1)
- Ingest Maintenance/Reject 'DROPPED_FTP' SIPs (version 1)
- Ingest Maintenance/Reject 'DROPPED_UPLOAD_SERVICE' SIPs (version 1)
- Ingest Maintenance/Reject 'SIP_INGESTED' SIPs (version 1)
- Ingest Maintenance/Reject 'SIP_NORMALIZED' SIPs (version 1)
- Ingest Maintenance/Reject 'SIP_VALID' SIPs (version 1)
- Ingest Maintenance/Reject 'UNPACKED' SIPs (version 1)
- Ingest Maintenance/Reject 'VIRUS_FREE' SIPs (version 1)
- Ingest tools/Auto accept SIP (version 1)
- Ingest tools/Notify producers (version 1)
- Ingest/Check SIP syntax (version 1)
- Ingest/Check producer authorization (version 1)
Management features
Advanced user management

<table>
<thead>
<tr>
<th>User</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>admin</td>
<td>Administrador RODA</td>
</tr>
<tr>
<td>arodrigues</td>
<td>Ana Maria do Rosário Silva Rodrigues</td>
</tr>
<tr>
<td>chenriques</td>
<td>Maria Cecília de Jesus Henriques</td>
</tr>
<tr>
<td>fbarbedo</td>
<td>francisco barbedo</td>
</tr>
<tr>
<td>guest</td>
<td>Utilizador anónimo</td>
</tr>
<tr>
<td>jmaferreira</td>
<td>José Ferreira</td>
</tr>
<tr>
<td>lcoruja</td>
<td>Luis Miguel Nunes Coruja</td>
</tr>
<tr>
<td>Faria</td>
<td>Luís Francisco da Cunha Cardoso de Faria</td>
</tr>
</tbody>
</table>

- **List of Users**
  - Users: admin, arodrigues, chenriques, fbarbedo, guest, jmaferreira, lcoruja, Faria, guests, administrators
  - Roles: Administrador RODA, Ana Maria do Rosário Silva Rodrigues, Maria Cecília de Jesus Henriques, francisco barbedo, Utilizador anónimo, José Ferreira, Luis Miguel Nunes Coruja, Luís Francisco da Cunha Cardoso de Faria

- **Search**: Field for searching users.

- **Users**
  - Options: REPORT, NEW, EDIT, INACTIVATE, REMOVE

- **Groups**
  - Options: NEW, EDIT, REMOVE
Comprehensive logging system
auditing purposes

<table>
<thead>
<tr>
<th>Date and Time</th>
<th>Action</th>
<th>Parameters</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-05-12 12:05:10.0</td>
<td>Browser getProducers</td>
<td>doPID: roda:2</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:10.0</td>
<td>UserBrowser getUser</td>
<td>name: mferreira</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:10.0</td>
<td>UserBrowser getGroup</td>
<td>groupName: administrators</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:10.0</td>
<td>UserBrowser getUser</td>
<td>groupName: producers</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:07.0</td>
<td>UserBrowser getUser</td>
<td>name: admin</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:07.0</td>
<td>UserBrowser getUser</td>
<td>name: fbarbedo</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:07.0</td>
<td>UserBrowser getUser</td>
<td>name: Ifaria</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:07.0</td>
<td>UserBrowser getUser</td>
<td>name: mferreira</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:07.0</td>
<td>UserBrowser getGroup</td>
<td>groupName: administrators</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:07.0</td>
<td>UserBrowser getUser</td>
<td>name: admin</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:07.0</td>
<td>UserBrowser getGroup</td>
<td>groupName: producers</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:06.0</td>
<td>Browser getDescriptionObject</td>
<td>pid: roda:2</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:06.0</td>
<td>Browser getRODAObjectUserPermissions</td>
<td>pid: roda:2, dissemination.browse.2</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:06.0</td>
<td>Browser getRODAObjectPermissions</td>
<td>pid: roda:2</td>
<td>Ifaria</td>
</tr>
<tr>
<td>2009-05-12 12:05:05.0</td>
<td>Logger addLogEntry</td>
<td>logEntry: LogEntry (datetime=null, address=192.168.111.154, username=Ifaria, action=RODA WUI pageHit, parameters=</td>
<td>roda-wui</td>
</tr>
</tbody>
</table>
Take up so far...
Running at

National Archives of Portugal
National Archives of Brazil (will go live soon)

Organisations that have requested the VM

University of Michigan, USA
National Library of Wales, UK
The Bavarian State Archives, Germany
King’s College London, UK
University of Dundee, UK
Casa de las Americas, Cuba
Marine Biological Laboratory, USA
ELDA: Evaluations and Language resources Distribution Agency
Unicamp, Brasil
National Library of Australia
University of St Andrews, UK
Community support
RODA

The world's most advanced open-source digital repository

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 supported by existing standards such as the OAIS, METS, EAD and PREMIS.

Conforms to open standards
RODA follows open standards using EAD for description metadata, PREMIS for preservation metadata, METS for structural metadata, several standards for technical metadata (e.g. NISO Z39.87 for digital still images).

Vendor independent
RODA is 100% built on top of open-source technologies. The entire infrastructure required to support RODA is vendor independent. This means that you may use the hardware and Linux distributions that best fit your institutional needs.

Scalable
The service-oriented nature of RODA's architecture makes it extremely scalable. RODA can be easily deployed and managed on any hardware platform.

Embedded preservation actions
Preservation actions and management within RODA are seamlessly embedded in the workflow. This ensures that digital content is adequately protected and preserved throughout its lifecycle.
RODA
The world's most advanced open-source digital repository

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 supported by existing standards such as the OAIS, METS, EAD and PREMIS.

Conforms to open standards
RODA follows open standards using EAD for description metadata, PREMIS for preservation metadata, METS for structural metadata, several standards for technical metadata (e.g. NISO Z39.87 for digital still images).

Vendor independent
RODA is 100% built on top of open-source technologies. The entire infrastructure required to support RODA is vendor independent. This means that you may use the hardware and Linux distributions that best fit your institutional needs.

Scalable
The service-oriented nature of RODA's architecture allows the system to scale.

Embedded preservation actions
Preservation actions and management within RODA is handled internally. This makes RODA a powerful tool for ensuring digital preservation.
Know more about the project and product

Download RODA
Source-code, binary, VM

Online demo
demo.roda-community.org

Find help
Documentation, developers guide, articles, mailing lists

Find vendors for paid support
Developing the future today

Miguel Ferreira
mferreira@keep.pt

KEEP SOLUTIONS, LDA
http://www.keep.pt