



Ingest with RODA

The present and the future of repository ingest

Luís Faria

KEEP SOLUTIONS

Outline

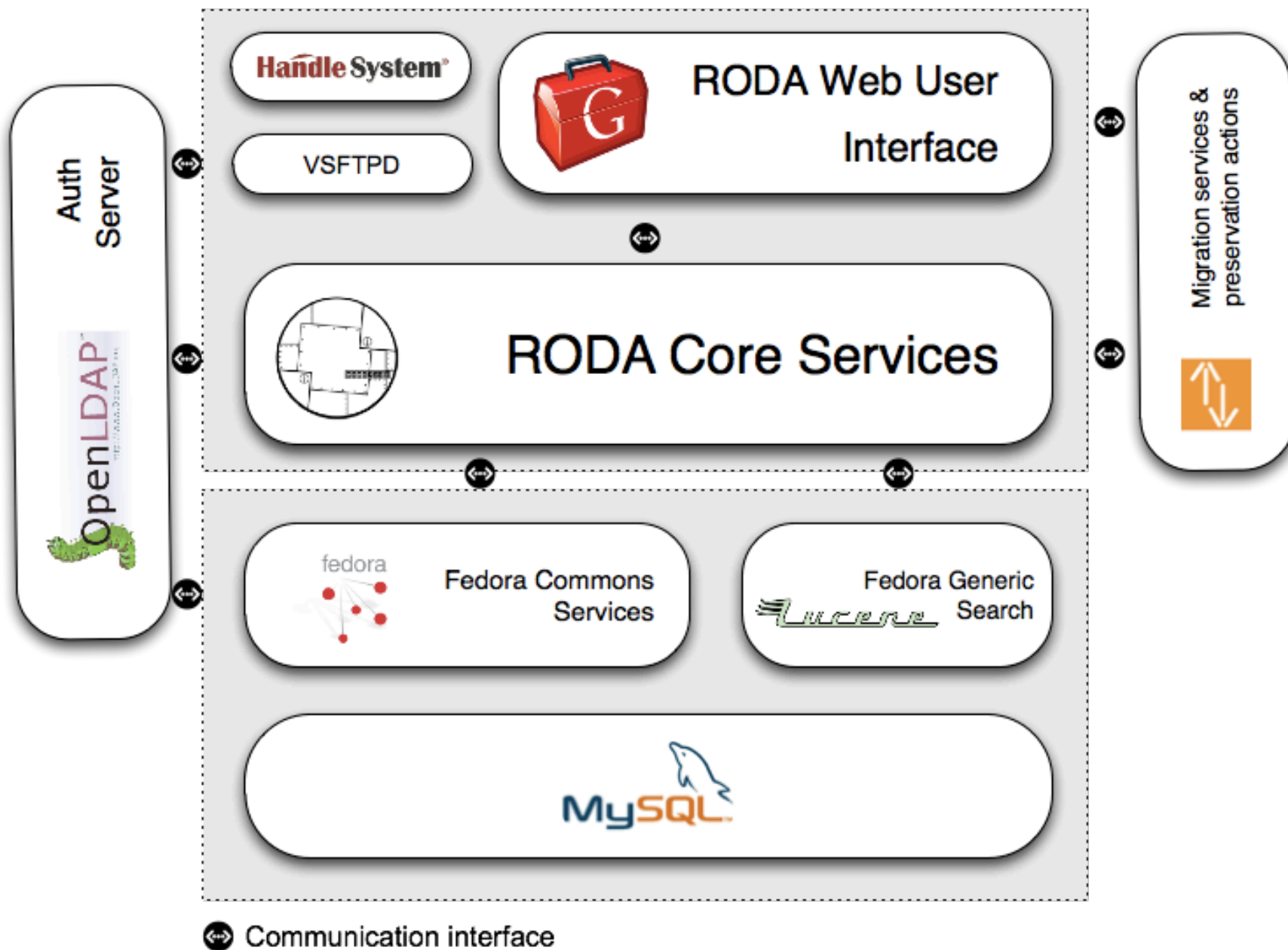
- What is RODA?
- The pre-ingest procedure
- RODA-in: producer ingest desktop tool
- SIP format
- Ingest workflow
- Current work in SCAPE
 - Integration of preservation planning and watch tools
 - Scalability: platform and preservation components

What is RODA?

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

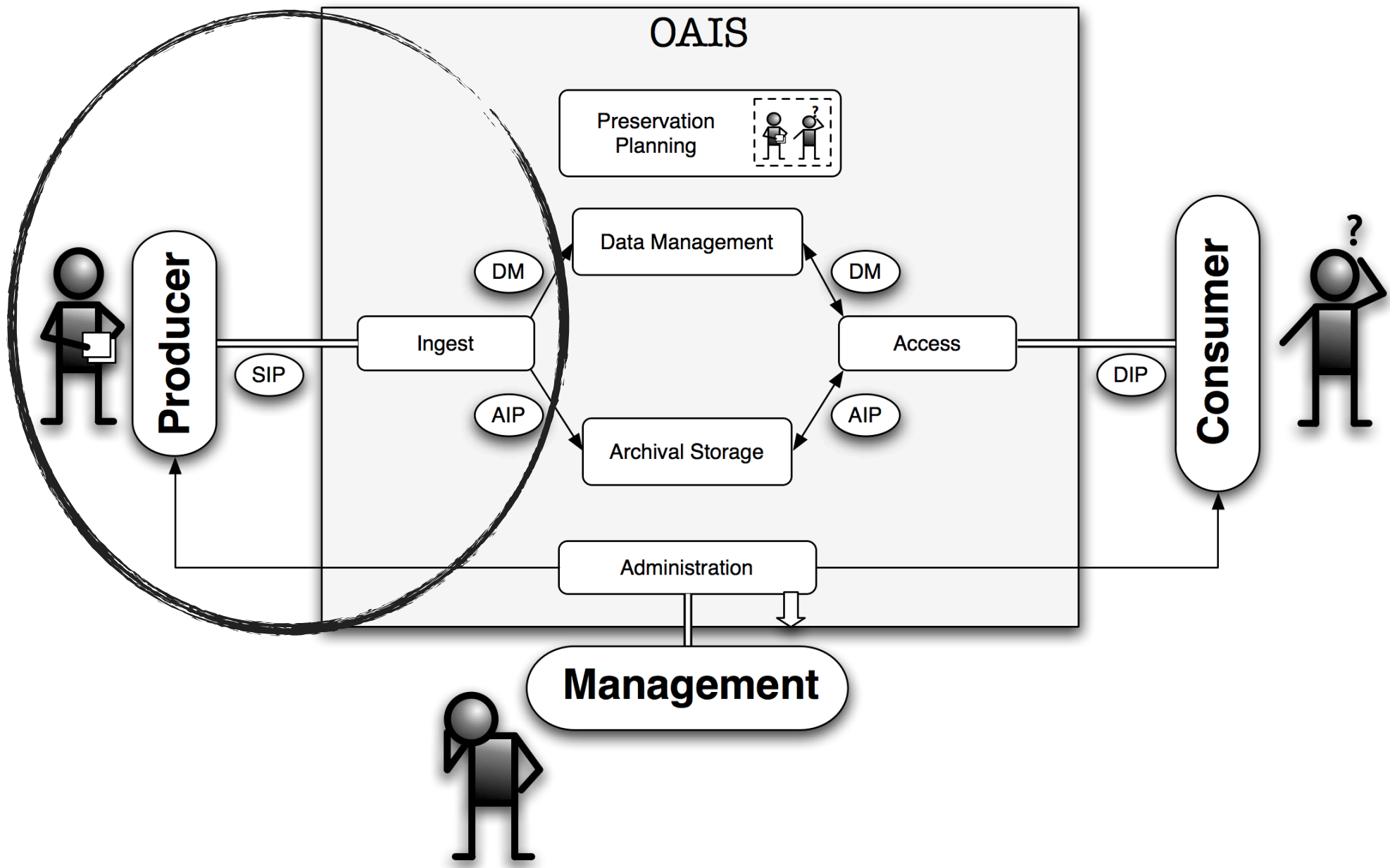
- Long-term Preservation and Authenticity
- Based on standards (OAIS, EAD, PREMIS, METS, etc.)
- TRAC focused (Trustworthy Repositories Audit & Certification)
- Security focused
- Service Oriented Architecture (SOA)
- Clean design
- Open source
- Portuguese National Archives and University of Minho

Architecture



Distinguishing features

- Modern web user interface design
- Preservation metadata management and generation
- Pluggable preservation actions
- Schedule actions to run repeatedly
- Pluggable content viewers
- Configurable normalization rules
- Configurable ingest workflow
- Thorough statistics, reporting and logs
- Published API and services



Pre-ingest procedure

- Institutional agreement between producer and archive
 - Transfer of custody
 - Patrimonial value of content
 - Definition of requisites for content
- Definition of a classification plan
- User authorization

ABOUT RODA

SERVICES · POLICY AND TECHNICAL DOCUMENTS · R&D · CONTACTS · ABOUT THE REGISTER · HELP

CATALOG

BROWSE · SEARCH · ADVANCED SEARCH · HELP

INGEST

PRE-INGEST · SUBMIT · STATUS · HELP

ADMINISTRATION

USERS · SCHEDULER · STATISTICS · LOG · HELP

catalog / browse / 50218

BROWSE

4 fonds

Reference

- AACC
- DGARQ
- TEST
 - 692
 - LFARIA
 - sf1
 - c1
 - sr1
 - 101
 - 815
 - 1030
 - 50192
 - 50218
- fbarbedo

Description Visualization Preservation Permissions

Identification

Reference	50218
Complete reference	PT/DGARQ/TEST/sf1/c1/sr1/50218
Handle	http://hdl.handle.net/10384/50218
Title	National Geographic Photos
Level	item
Initial date	2008-01-01
Final date	2009-01-01
Country code	PT
Repository code	DGARQ
Origination	National Geographic
Acquisition number	NG
Acquisition date	2009-05-11

Content and structure

Scope and content tests for or09

Notes

Note Pacote de submissão validado semanticamente por Ifaria

EDIT

SUBLEVEL

MOVE

EAD























CLOSE

catalog / browse / 2

BROWSE

4 fonds  

Reference

-   AACC
-   DGARQ
-   TEST
 -  692
 -   LFARIA
 -   sf1
 -   c1
 -   sr1
 -  101
 -  815
 -  1030
 -  50192
 -  50218
-   fbarbedo

Description **Producers** Permissions

Users or groups:

-  José Miguel Araújo Ferreira (mferreira)
-  administrators
-  producers

Supported formats



Supported formats



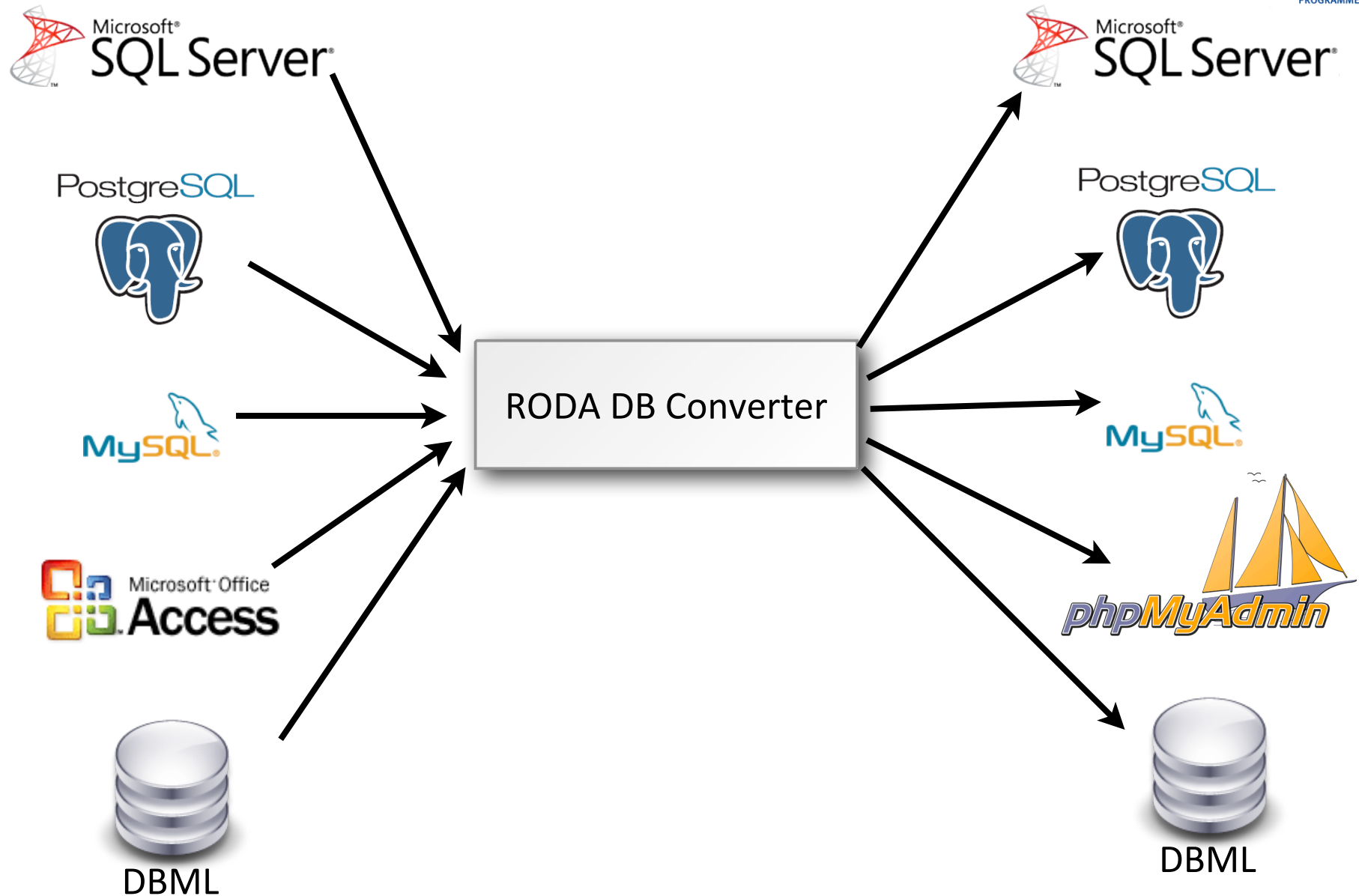
PostgreSQL



Migration tools

- Text: Open Office and Microsoft Office
- Images: Imagemagick
- Video: Mencoder and FFMpeg
- Audio: GStreamer
- Databases: RODA DB Converter

RODA DB Converter



SIP creation tools

- RODA-in
 - Occasional submissions
- RODA web upload form
 - Small and occasional submissions
- RODA SIP manipulation libraries and upload API
 - Custom build SIP creation
 - Software integration



Update plan Send



TEST/sf1/c1/sr1/NG

Description

 Show all fields

Identification

Reference: NG

Title: National Geographic Photos

Description level: Item

Initial date: 2008-01-01

Final date: 2009-01-01

Country code: PT

Repository code: DGARQ

Producer: National Geographic

Context

Scope and content: tests for or09

Access and use restrictions

Access restrictions: Free access

Create Remove

Save Validate Create simple document Remove simple document Create representation Remove representation

Text document



Text document like PDF, Microsoft Word or Rich Text Format

Raster image



One or more bidimensional still images with associated structure. Images format can be JPEG, GIF, PNG, BMP, etc.

Audio



Audio file like WAV, MP3, OGG, FLAC, etc.

Video



Video file like MPEG-2, MPEG-4, AVI, WMV or Quicktime

Relational database



Relational database like MySQL, PostgreSQL, Microsoft SQL Server, Microsoft Access, or even DBML.

Unknown



If document doesn't fit in above classes, choose this generic class. But beware that no preservation will be guaranteed!

Update plan Send



TEST/sf1/c1/sr1/NG

Description

Representation

 Only TIFFs Only JPEGs Only PNGs Only BMPs Only GIFs Only ICOs Only XPMs Only TGAs

Select and organize images:

- R2009-05-11T13.38.58.46Z
 - ocelot-california-1198172-xl.jp
 - right-whales-1189052-xl.jp
 - haw-creek-falls-1190091-xl.jp
 - mexican-gray-wolf-1198153-xl.jp
 - gray-fox-dusk-1201638-xl.jp
 - coconut-island-fish-xl.jp
 - bee-eater-hungary-118899-xl.jp
 - moose-mayfield-lake-1191505-xl.jp
 - atlantic-walrus-MM7468-xl.jp
 - toad-river-canada-1191505-xl.jp
 - elephant-seal-pups-119216-xl.jp
 - crystal-cave-1191533-xl.jp
 - chicago-lights-1191505-xl.jp

Preview

Metadata



Create Remove


Save Validate

Create simple document Remove simple document

Create representation Remove representation

Send packages

Select packages to send:

 TEST/sf1/c1/sr1/NG

Select sending means:

- Online, directly (package < 1GB)
- Online, via FTP
- Offline, via postal mail

Cancel

Send

Authentication



Username:

Password:

Cancel

Login

Sending packages

Uploading TEST/sf1/c1/sr1/NG to RODA



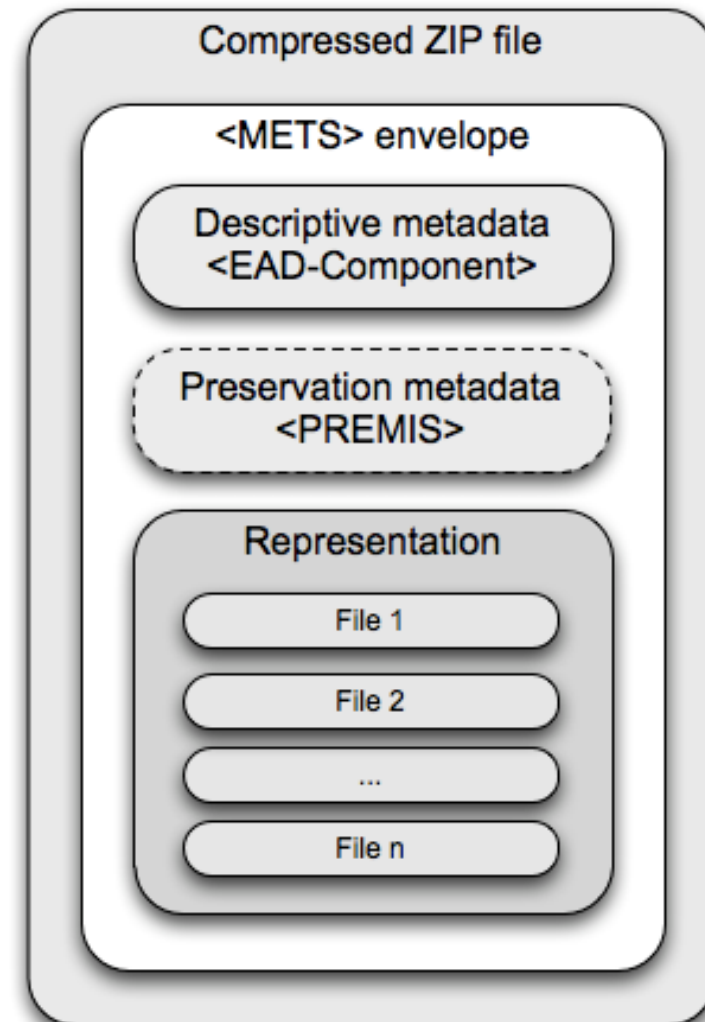
50%

Processing 1 of 1

Cancel

SIP structure

- Object class and format
- Place in classification plan
- Descriptive metadata
- Preservation metadata
- Technical metadata
- Representation files
- Identification of root file



Ingest workflow

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

STATUS

38 packets

File	Submission date	Current state	%	Producer	
TEST_sf1_c1_sr1_NG.sip	2009-05-11 14:41:47	accepted	100%	Ifaria	
TEST_sf1_c1_sr1_fox.sip	2009-05-07 12:19:36	accepted	100%	Ifaria	
TEST_LFARIA_MP48298.sip	2009-03-20 17:50:39	accepted	100%	Ifaria	
TEST_LFARIA_PNG2.sip	2009-03-20 15:06:14	accepted	100%	Ifaria	
TEST_LFARIA_PNG492.sip	2009-03-20 15:05:28	accepted	100%	Ifaria	
TEST_LFARIA_XPM.sip	2009-03-20 11:40:25	accepted	100%	Ifaria	
TEST_LFARIA_WAV.sip	2009-03-20 11:40:24	accepted	100%	Ifaria	
TEST_LFARIA_TXT.sip	2009-03-20 11:40:16	accepted	100%	Ifaria	
TEST_LFARIA_TIFF.sip	2009-03-20 11:40:15	accepted	100%	Ifaria	
TEST_LFARIA_TGA.sip	2009-03-20 11:40:15	accepted	100%	Ifaria	
TEST_LFARIA_RTF.sip	2009-03-20 11:40:14	accepted	100%	Ifaria	
TEST_LFARIA_PDF.sip	2009-03-20 11:40:13	accepted	100%	Ifaria	
TEST_LFARIA_OGG.sip	2009-03-20 11:40:06	accepted	100%	Ifaria	
TEST_LFARIA_ODT.sip	2009-03-20 11:40:05	accepted	100%	Ifaria	
TEST_LFARIA_MP3.sip	2009-03-20 11:40:01	accepted	100%	Ifaria	
TEST_LFARIA_JPEG.sip	2009-03-20 11:39:28	accepted	100%	Ifaria	
TEST_LFARIA_ICO.sip	2009-03-20 11:39:25	accepted	100%	Ifaria	
TEST_LFARIA_GIF.sip	2009-03-20 11:39:24	accepted	100%	Ifaria	

Filter

processed
processing
accepted
rejected
All

Producer

Ifaria


REPORT





































VIEW

ACCEPT

REJECT

STATUS

38 packets 

File	Submission date	Current state	%	Producer	
TEST_sf1_c1_sr1_NG.sip	2009-05-11 14:41:47	accepted	100%	lfaria	 
TEST_sf1_c1_sr1_fox.sip	2009-05-07 12:19:36	accepted	100%	lfaria	 
TEST_LFARIA_MP48298.sip	2009-03-20 17:50:39	accepted	100%	lfaria	 
TEST_LFARIA_PNG2.sip	2009-03-20 15:06:14	accepted	100%	lfaria	 
TEST_LFARIA_PNG492.sip	2009-03-20 15:05:28	accepted	100%	lfaria	 
TEST_LFARIA_XPM.sip	2009-03-20 11:40:25	accepted	100%	lfaria	 
TEST_LFARIA_WAV.sip	2009-03-20 11:40:24	accepted	100%	lfaria	 
TEST_LFARIA_TXT.sip	2009-03-20 11:40:16	accepted	100%	lfaria	 
TEST_LFARIA_TIFF.sip	2009-03-20 11:40:15	accepted	100%	lfaria	 
TEST_LFARIA_TGA.sip	2009-03-20 11:40:15	accepted	100%	lfaria	 
TEST_LFARIA_RTF.sip	2009-03-20 11:40:14	accepted	100%	lfaria	 
TEST_LFARIA_PDF.sip	2009-03-20 11:40:13	accepted	100%	lfaria	 
TEST_LFARIA_OGG.sip	2009-03-20 11:40:06	accepted	100%	lfaria	 
TEST_LFARIA_ODT.sip	2009-03-20 11:40:05	accepted	100%	lfaria	 
TEST_LFARIA_MP3.sip	2009-03-20 11:40:01	accepted	100%	lfaria	 
TEST_LFARIA_JPEG.sip	2009-03-20 11:39:28	accepted	100%	lfaria	 
TEST_LFARIA_ICO.sip	2009-03-20 11:39:25	accepted	100%	lfaria	 
TEST_LFARIA_GIF.sip	2009-03-20 11:39:24	accepted	100%	lfaria	 

RODA

Logout

ABOUT RODA
CATALOG
INGEST
ADMINISTRATION

[ingest / status](#)

STATUS

File

TEST_sf1_c1_sr1_NG
TEST_sf1_c1_sr1_fo
TEST_LFARIA_MP482
TEST_LFARIA_PNG2
TEST_LFARIA_PNG49
TEST_LFARIA_XPM.s
TEST_LFARIA_WAV.s
TEST_LFARIA_TXT.s
TEST_LFARIA_TIFF.s
TEST_LFARIA_TGA.s
TEST_LFARIA_RTF.s
TEST_LFARIA_PDF.s
TEST_LFARIA_OGG.s
TEST_LFARIA_ODT.s

TEST_LFARIA_MP3.sip	2009-03-20 11:40:01	accepted	100%	Ifaria
TEST_LFARIA_JPEG.sip	2009-03-20 11:39:28	accepted	100%	Ifaria
TEST_LFARIA_ICO.sip	2009-03-20 11:39:25	accepted	100%	Ifaria
TEST_LFARIA_GIF.sip	2009-03-20 11:39:24	accepted	100%	Ifaria

Ingest report

Original file name: TEST_sf1_c1_sr1_NG.sip

- ▶ received (HTTP)
- ▶ unpacked
- ▶ virus-free
- ▶ well formed
- ▶ authorized
- ▶ ingestd
- ▶ normalized
- ▼ accepted

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 [ro

CLOSE ✕

Ingest report

Original file name: TEST_sf1_c1_sr1_NG.sip

▶ received (HTTP)

▶ unpacked

▶ virus-free

▶ well formed

▶ authorized

▶ ingestd

▶ normalized

▼ accepted

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 [ro

CLOSE X

[ABOUT RODA](#)[SERVICES](#) · [POLICY AND TECHNICAL DOCUMENTS](#) · [R&D](#) · [CONTACTS](#) · [ABOUT THE REGISTER](#) · [HELP](#)[CATALOG](#)[BROWSE](#) · [SEARCH](#) · [ADVANCED SEARCH](#) · [HELP](#)[INGEST](#)[PRE-INGEST](#) · [SUBMIT](#) · [STATUS](#) · [HELP](#)[ADMINISTRATION](#)[USERS](#) · [SCHEDULER](#) · [STATISTICS](#) · [LOG](#) · [HELP](#)[catalog](#) / [browse](#) / 50218

BROWSE

4 fonds

Reference

- AACC
- DGARQ
- TEST
 - 692
 - LFARIA
 - sf1
 - c1
 - sr1
 - 101
 - 815
 - 1030
 - 50192
 - 50218
- fbarbedo

[Description](#) [Visualization](#) [Preservation](#) [Permissions](#)

Identification

Reference	50218
Complete reference	PT/DGARQ/TEST/sf1/c1/sr1/50218
Handle	http://hdl.handle.net/10384/50218
Title	National Geographic Photos
Level	item
Initial date	2008-01-01
Final date	2009-01-01
Country code	PT
Repository code	DGARQ
Origination	National Geographic
Acquisition number	NG
Acquisition date	2009-05-11

Content and structure

Scope and content tests for or09

Notes

Note Pacote de submissão validado semanticamente por Ifaria

[EDIT](#) [SUBLEVEL](#) [MOVE](#) [EAD](#) [CLOSE](#)

▼ Identification





Reference	50218
Complete reference	PT/DGARQ/TEST/sf1/c1/sr1/50218
Handle	http://hdl.handle.net/10384/50218
Title	National Geographic Photos
Level	item
Initial date	2008-01-01
Final date	2009-01-01
Country code	PT
Repository code	DGARQ
Origination	National Geographic
Acquisition number	NG
Acquisition date	2009-05-11

▼ Content and structure

Scope and content	tests for or09
-------------------	----------------

▼ Notes

Note	Pacote de submissão validado semanticamente por Ifaria
------	--

EDIT SUBLEVEL MOVE EAD CLOSE 

Description

Visualization

Preservation

Permissions

Disseminassions of 50218 - "National Geographic Photos"



JPEG
(original)



TIFF
(normalized)



Photo pre-viewer



Book pre-viewer

CLOSE



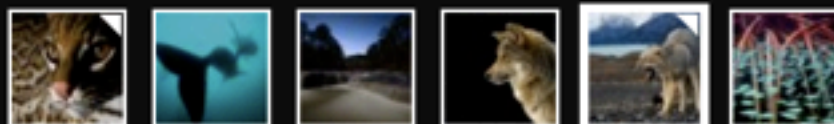


NATIONAL GEOGRAPHIC

Photograph by Mattias Klum

© COPYRIGHT NATIONAL GEOGRAPHIC SOCIETY. ALL RIGHTS RESERVED.

National Geographic Photos



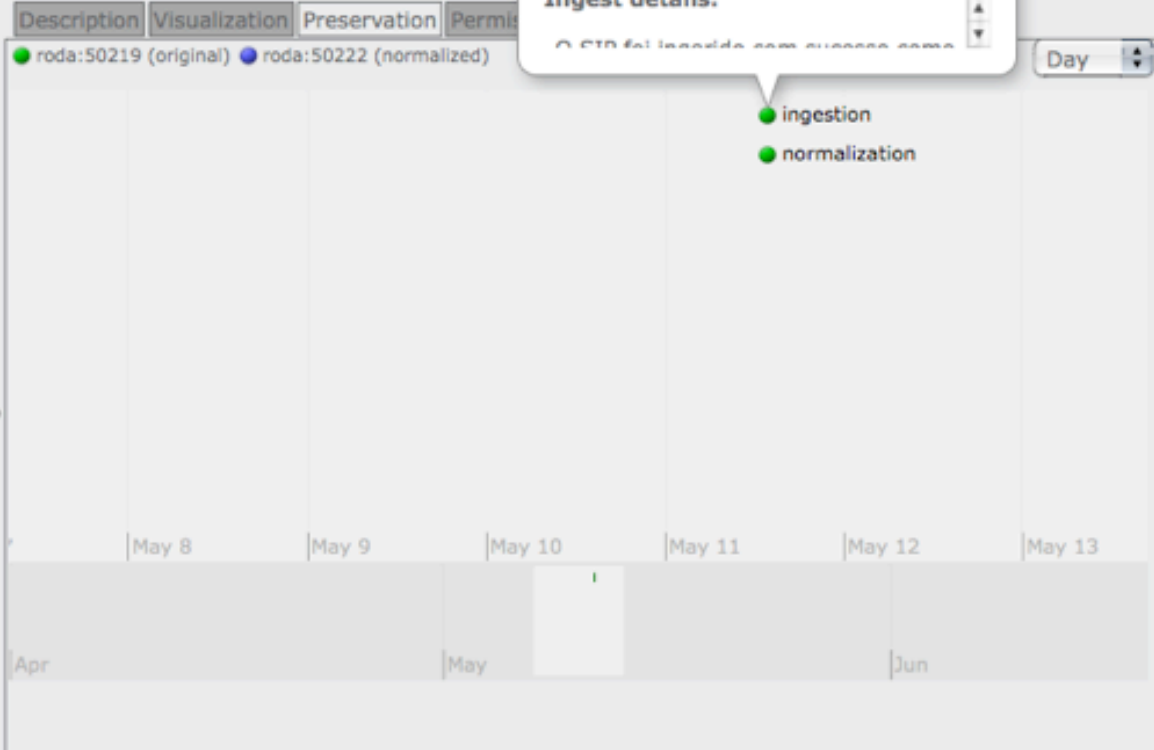
(5/13) gray-fox-dusk-1201638-xl.jpg.tiff

catalog / browse / 50218

BROWSE

4 fonds

- Reference
 - AACC
 - DGARQ
 - TEST
 - 692
 - LFARIA
 - sf1
 - c1
 - sr1
 - 101
 - 815
 - 1030
 - 50192
 - 50218
 - fbarbedo



ingestion

Description: The objects inside the SIP were added to the preservation repository

Result: success

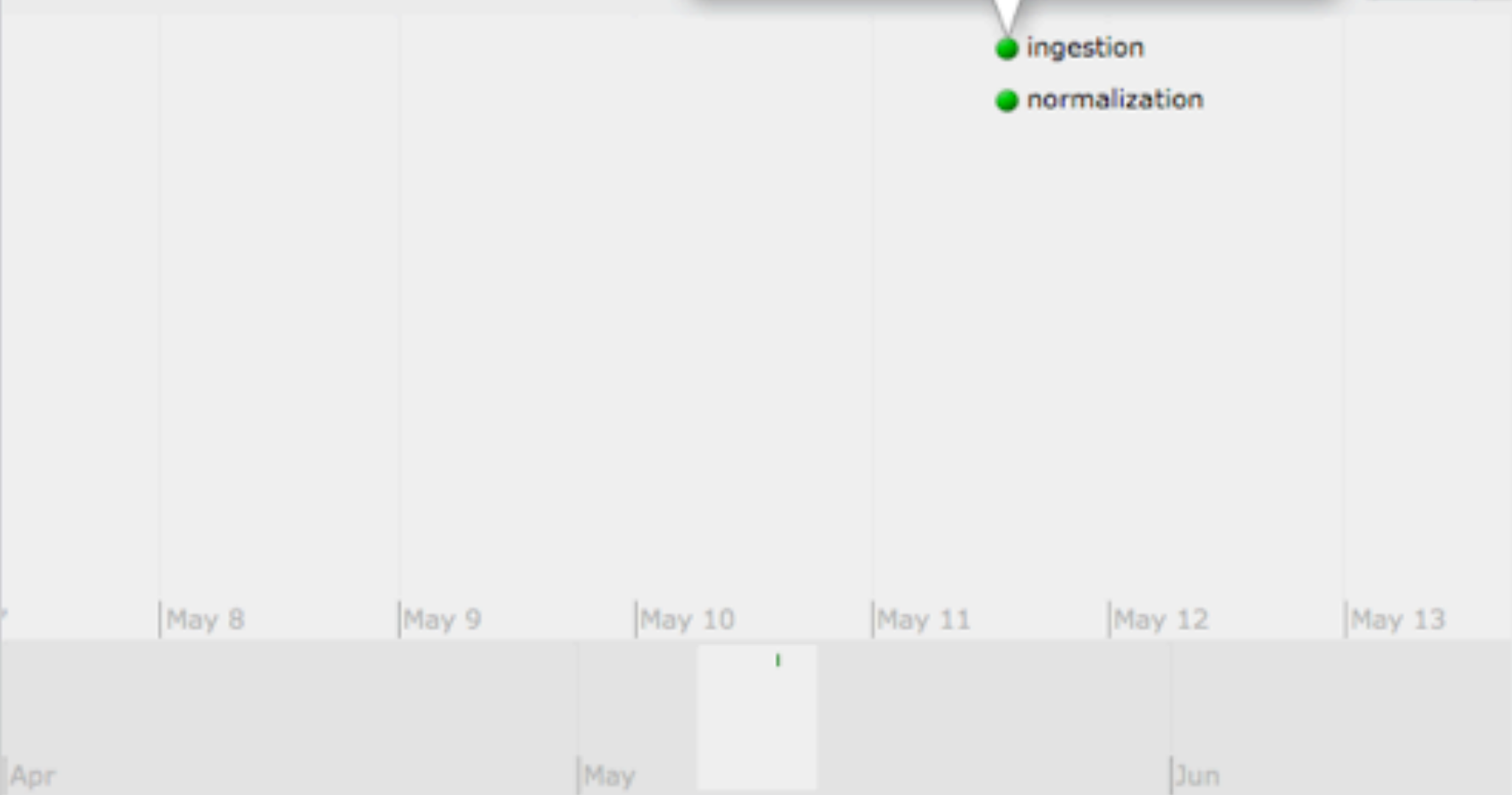
Ingest details:

● SIP file inside.com success.com

50218

Description Visualization Preservation Permissions

● roda:50219 (original) ● roda:50222 (normalized)



ingestion

Description: The objects inside the SIP were added to the preservation repository

Result: success

Ingest details:

○ SIP for ingested.com success.com

A

sr1

- 101
- 815
- 1030
- 50192
- 50218

PREMIS

CLOSE

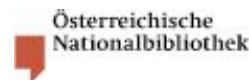
Current problems



- Scalability
 - Information quantity, size, type and complexity
- Monitor the world
 - Detect changes in environment
- Automatic planning
 - Systematize decision making
- Changing processes
 - Dynamic and flexible process execution

Ongoing work

- FP7 project with international consortium
- Develop platform and scalable tools
- Framework for preservation workflows
- Integration with planning and watch tools



Future of RODA ingest

- Ingest workflow per collection
- Optimize tasks to run on a scalable platform
- Integration with preservation planning tool
 - Ingest will trigger planning
 - Automatic execution of preservation plan
- Continuous monitoring
 - Ingest as a Watch information source
 - Monitor events that can trigger changes in planning

Ingest with RODA

The present and the future of repository ingest

Luís Faria

<http://bit.ly/scape-tools>

<http://www.scape-project.eu>

<http://redmine.keep.pt>

<http://roda.keep.pt>

