

Agenda Day 1 - afternoon

3 December

Activity

Presenter

14:00-15:00

Software presentation

- End-to-end eArchiving solutions for digital preservation
- eArchiving tools for database preservation

Chair: Kuldar Aas, National Archives of Estonia
Hélder Silva, KEEP Solutions Portugal
Luís Faria, KEEP Solutions Portugal

15:00-15:15

Coffee break

15:15-16:30

How to take up eArchiving ?

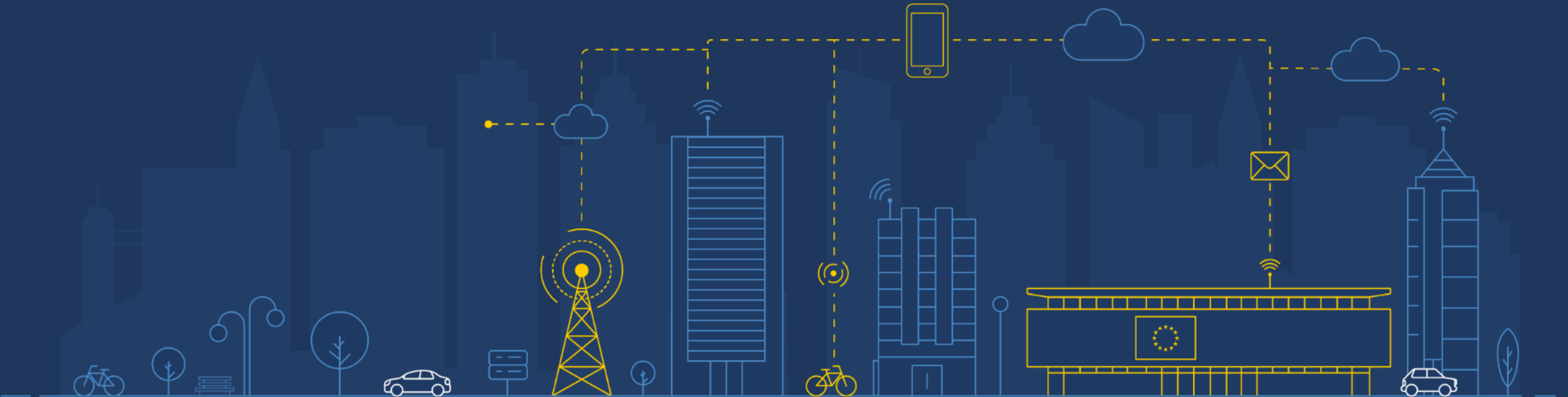
- Presentation on the Danish National Archives eArchiving journey
- Presentation on the Publications Office eArchiving journey

Chair: Jože Škofljanec, National Archives of Slovenia
Anders Bo Nielsen, National Archives of Denmark
Maria Kardami, Publications Office of the European Union,
Corinne Frappart, Publications Office of the European Union

16:30-17:00

Wrap up

Software presentation



Software presentation

End-to-end eArchiving solutions for digital preservation



Hélder Silva

Infrastructure Director,
KEEP Solutions, Portugal





Digitalbevaring.dk

If, by the end of the presentation, you don't know:

- 1. What software was presented*
- 2. What it is for*
- 3. How it works*
- 4. Where you can find it*
- 5. ...*

come and talk with us, we are happy to talk with you!

***End-to-end eArchiving solutions for digital
preservation***

*End-to-end eArchiving solutions for **digital
preservation***

“set of actions to ensure that digital information remains accessible and usable throughout time”

MIME
EAD
Migration
WARC
JPEG
TIFF
OPF
METS
OAIS
RODA
AIP
DC
SIP
JHove
ISO
PREMIS
Authentication
NDSA
Refreshing
PAIMAS
PDF/A
DCC
MPEG
Digitisation
DPC
DRAMBORA
Authenticity
Formats
DDI
PRONOM
DIP
XML
SGML
CCSDS
Checksum
DOI
DROID
Emulation
TRAC
HTML

*End-to-end **eArchiving** solutions for digital
preservation*

20180912_eArchiving_Flyer.pdf



1



2

eArchiving

Connecting Europe

Simplify long-term access to information with the CEF eArchiving building block: a technical solution for developers that need to migrate, preserve and reuse data.

The eArchiving Building Block can provide long-term information assurance. It provides the specifications, reference software, training and service desk support for digital archiving, including digital preservation. This benefits both the design and implementation of repositories and enables business systems to send data to those repositories.

Information Package specifications are the foundation of eArchiving. These describe platform-independent formats to structure information assets as bulk data and metadata that remains authentic and understandable over time. They are thus ideal for:

- migrating information assets between generations

How it works

eArchiving is based on Information Package specifications that provide interoperability across borders, types of institution and user communities. These are supported by the Common Specification for Information Packages (CS IP), expressed with the Metadata Encoding & Transmission Standard (METS), with specialisations for:

- Submission Information Package

20180912_eArchiving_Flyer.pdf



1



2

eArchiving

Connecting
Europe

Simplify long-term access to information with the CEF eArchiving building block: a technical solution for developers that need to migrate, preserve and reuse data.

It provides the specifications, reference software, training and service desk support for digital archiving, including digital preservation.

The eArchiving building block provides a reference solution for digital archiving, including digital preservation. This benefits both the design and implementation of repositories and enables business systems to send data to those repositories.

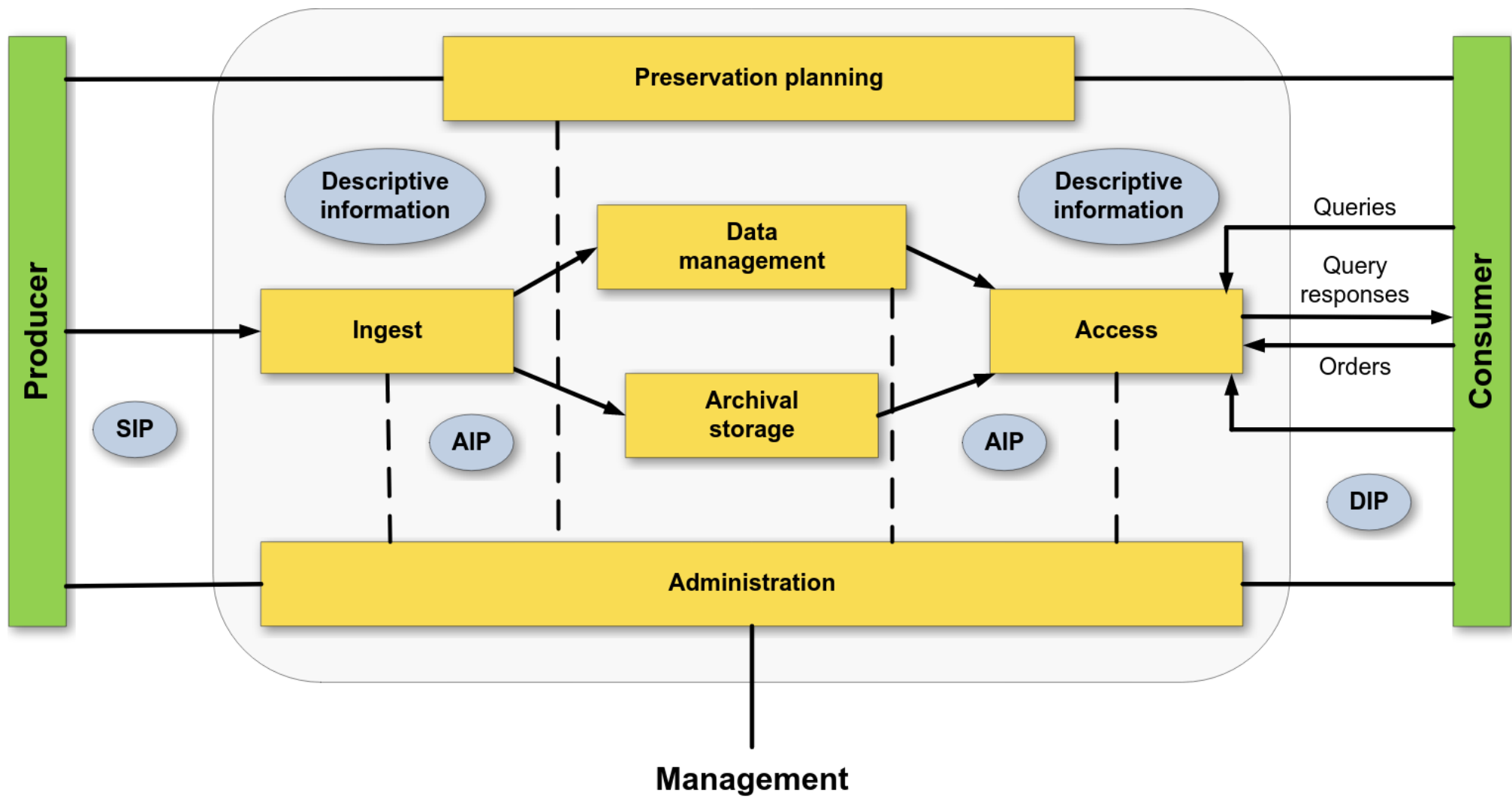
Information Package specifications are the foundation of eArchiving. These describe platform-independent formats to structure information assets as bulk data and metadata that remains authentic and understandable over time. They are thus ideal for:

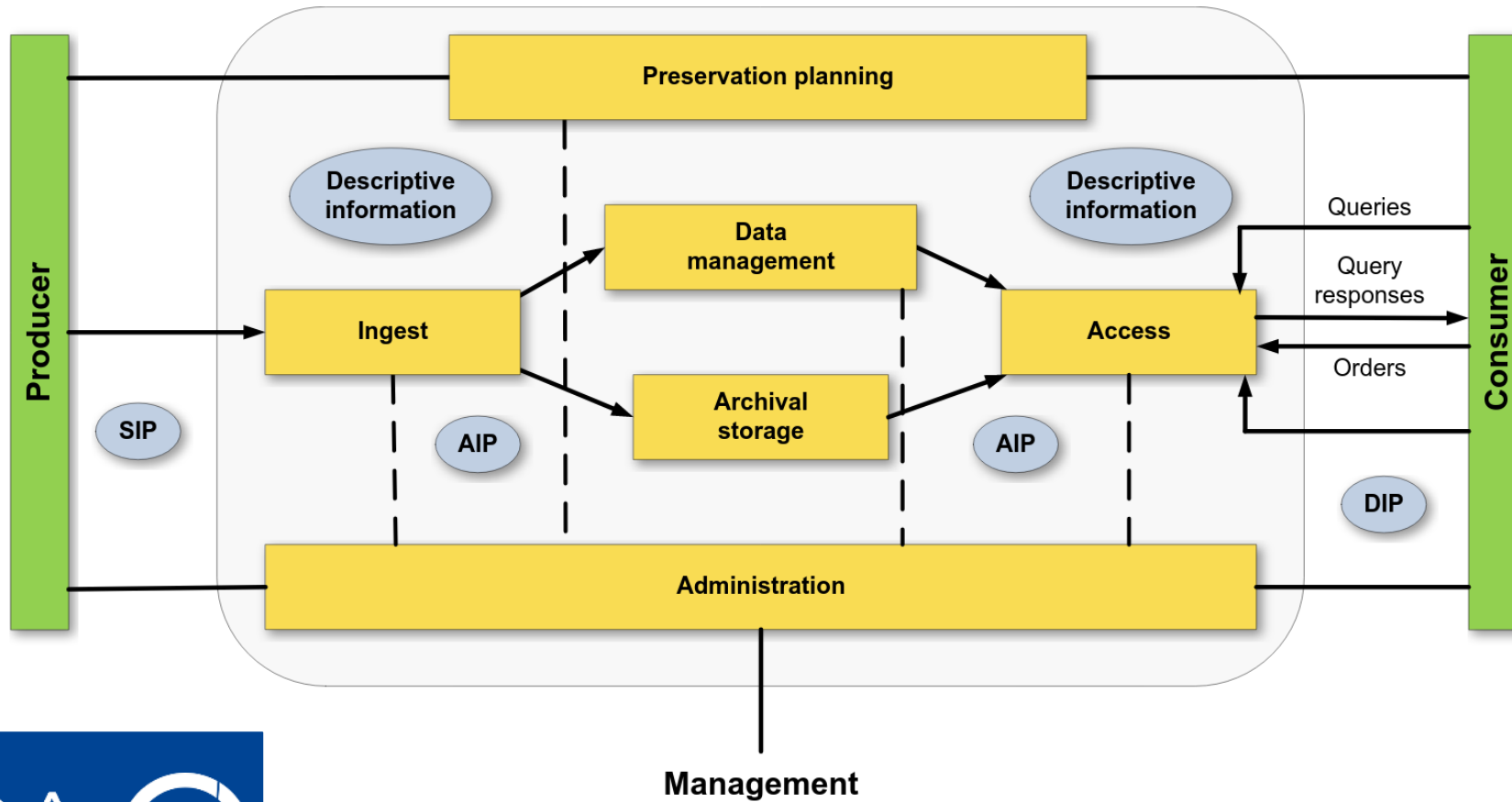
- migrating information assets between generations

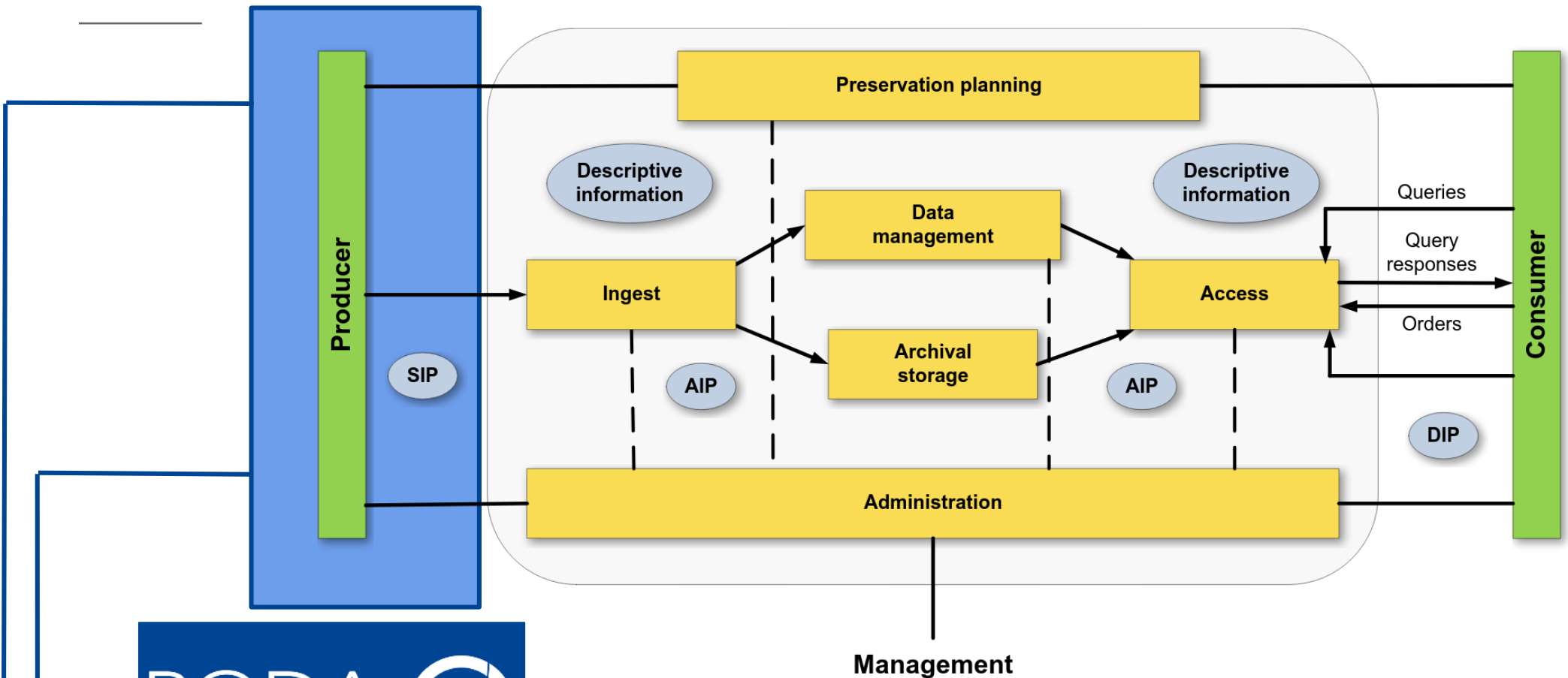
Information Package specifications that provide interoperability across borders, types of institution and user communities. These are supported by the Common Specification for Information Packages (CS IP), expressed with the Metadata Encoding & Transmission Standard (METS), with specialisations for:

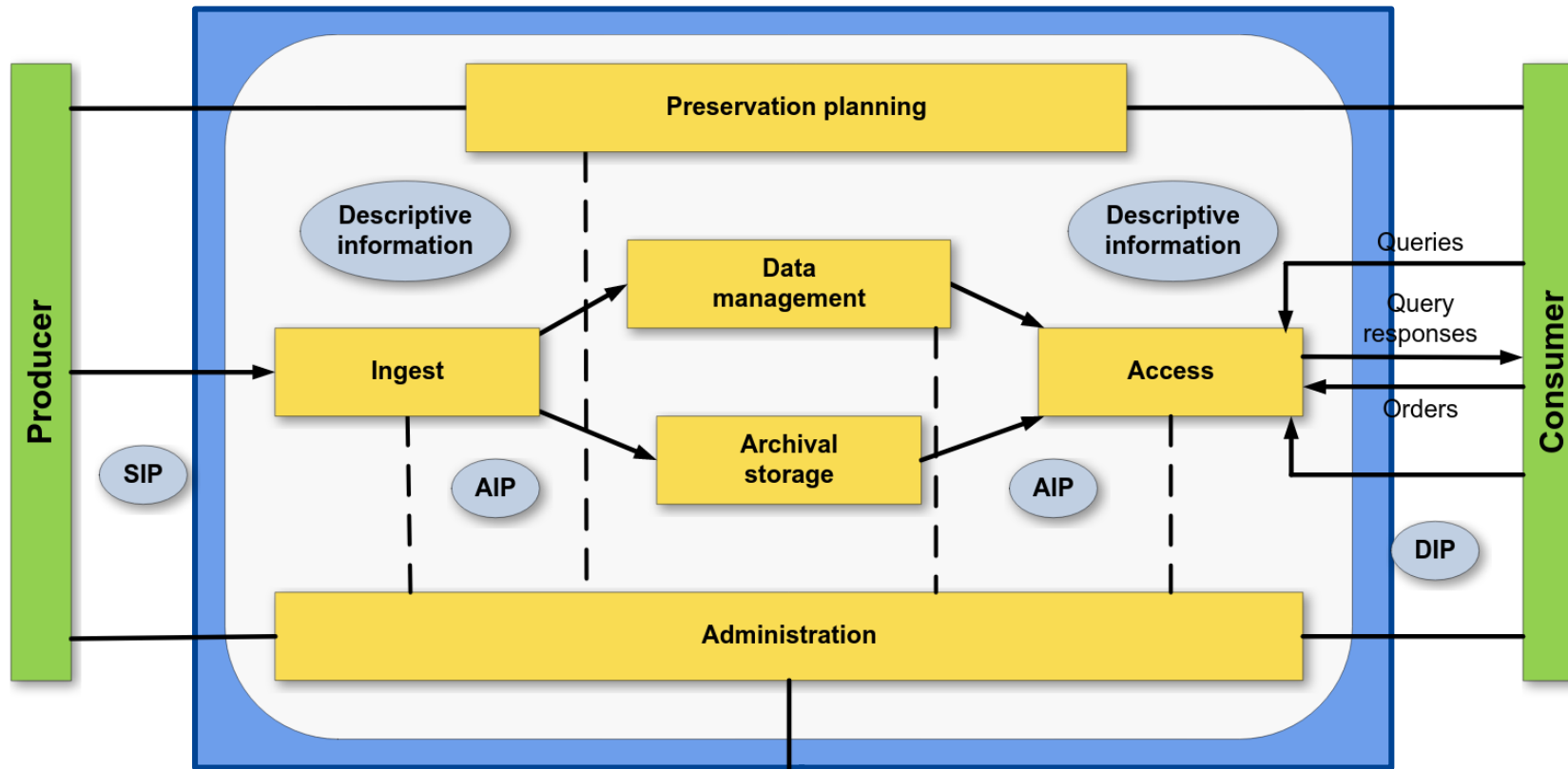
- Submission Information Package

***End-to-end eArchiving solutions for digital
preservation***



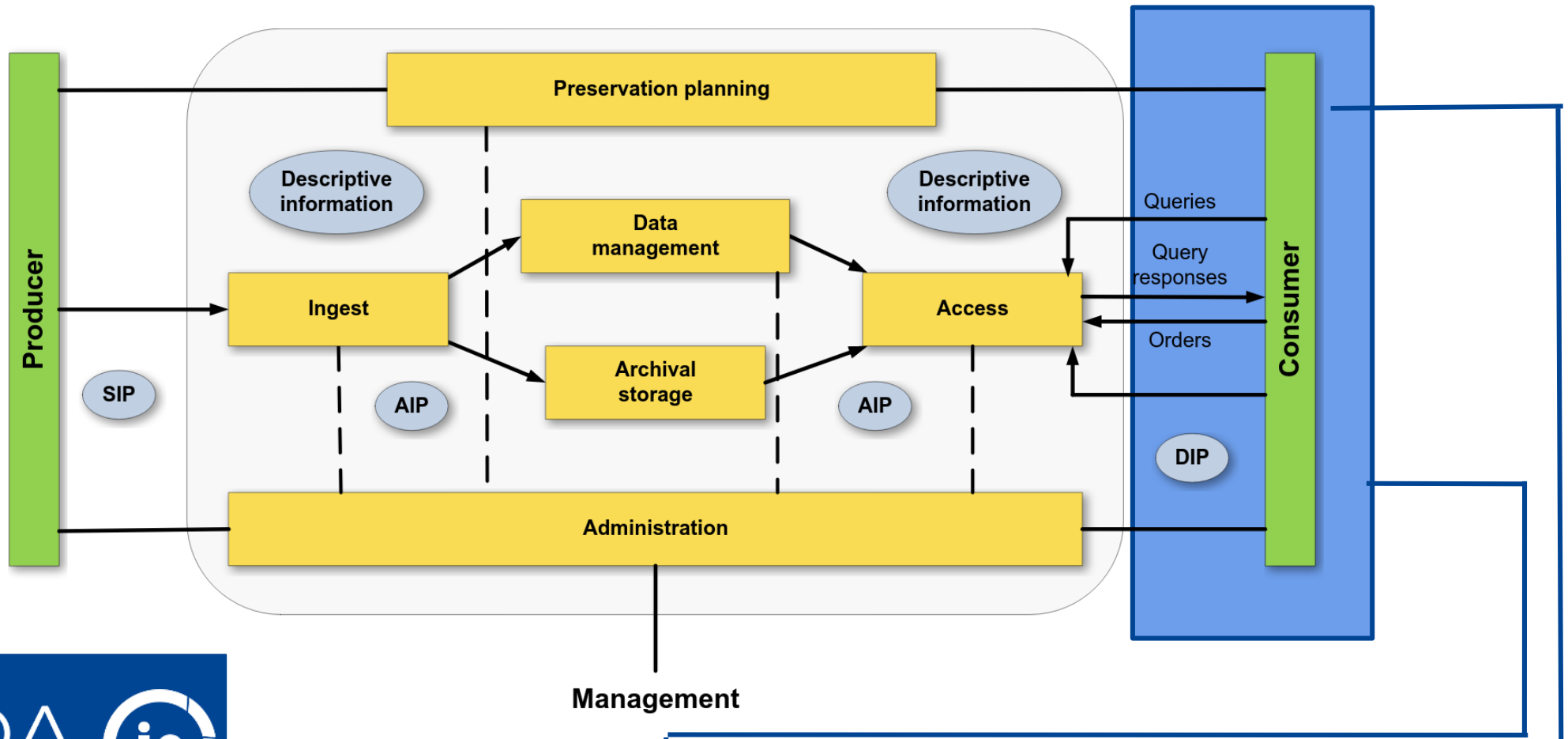






Management





eArchiving tools

RODA
SIP creation tool



What is RODA-in?



SIP **creation tool**

Desktop tool to create Submission Information Packages (SIP)



For **producers**

To prepare data to be sent to the Archive



Various output formats

E-ARK SIP, BagIt, Hungarian SIP (type 4)



Conforms to **open standards**

EAD, DC, METS, etc.



Multiple **descriptive metadata** schemas

EAD 2002, EAD 3 and Dublin Core



Multi-language

English, Portuguese, Spanish and Hungarian



Multi-platform

Windows, Mac OS X, Linux



Offline operation

No network required to operate the tool

What scenarios does it cover?



Files in a file system



Packaging



Transfer & ingest

EARK compatible
Digital Repository

How does it work?

RODA-In

File Edit Classification Scheme View Help

1. FILE EXPLORER

2. INFORMATION PACKAGES

3. INSPECTOR

4. SUBMISSION PACKAGES

Create SIP(s)

Choose the folders that contain your files

Choose folder

Select an item from the information packages to inspect it

Memory usage: 97.1 MB of 503.5 MB

The process is composed of 4 stages

RODA-In

File Edit Classification Scheme View Help

1. FILE EXPLORER

2. INFORMATION PACKAGES

3. INSPECTOR

1. Choose files or folders

Choose the folders that contain your files

Choose folder

Load your classification scheme

Load

or create a [new classification scheme](#)

Select an item from the information packages to inspect it

4. SUBMISSION PACKAGES

Create SIP(s)

Memory usage: 97.1 MB of 503.5 MB

RODA-In

File Edit Classification Scheme View Help

1. FILE EXPLORER

- 06-Dspace themes /Users/mferreira/My Documents/13-KEE...
 - 01-ESEPF
 - 02-ULusiada
 - 03-IPCB
 - 04-UTAD

2. INFORMATION PACKAGES

2. Choose an association strategy

Load your classification scheme

Load

or create a [new classification scheme](#)

Ignore Associate

3. INSPECTOR

Select an item from the information packages to inspect it

4. SUBMISSION PACKAGES

Create SIP(s)

Memory usage: 115.4 MB of 503.5 MB

RODA-In

File Edit Classification Scheme View Help

1. FILE EXPLORER

- 06-Dspace themes /Users/mferreira/My Documents/13-KEE
 - 01-ESEPF
 - 02-ULusiada
 - 03-IPCB
 - 04-UTAD

2. INFORMATION PACKAGES

3. INSPECTOR

4. SUBMISSION PACKAGES


File Explorer: Ignore Associate


Submission Packages: Create SIP(s)


Memory usage: 115.4 MB of 503.5 MB


CREATE ASSOCIATION TO "Root"

CHOOSE THE ASSOCIATION METHOD

 **One information package for each selected files or folders**
 Use this option to create a SIP for each of the selected files or folders. If you selected five files and/or folders, the association will create five SIPs, i.e. five descriptive items. Each one of these items can correspond to a distinct description level.

 **One information package with all selected files and/or folders**
 Use this option to create one and only one SIP containing all selected files and folders. If you selected five folders and/or files the association will result in one SIP corresponding to one item with a set description level.

 **One information package for each file under the selected folder(s)**
 Use this option to create one SIP for each file under the selected folder(s). This operation will create a set of SIPs equal to the number of files under the selected folder(s).

 **One classification scheme from folder structure**
 This option creates information packages in addition to submission packages. It should be used when the folder structure is well organized and resembles the desired output plan. If a folder only has sub-folders, it creates a Series. When a folder's children are just files, the whole folder will be used to create one submission package. In a situation where a folder has a mixed content, files and sub-folders, each file will be a submission package and each sub-folder will be a Series.

Cancel x Continue >

an item from
 nation packages
 inspect it

RODA-In

File Edit Classification Scheme View Help

1. FILE EXPLORER

06-DSpace themes /Users/mferreira/My Documents/13-KEEF

2. INFORMATION PACKAGES

- 01-ESEPF
- Universidade Lusfada
- 03-IPCB
- 04-UTAD

3. INSPECTOR

Universidade Lusfada MIXED

METADATA

EAD 2002

Identity

Identifier: uuid-33476bbe-0a00-48c1-b486-1dfcff20d664

Description level: File

Title: Universidade Lusfada

Initial date: 2017-09-11

Final date: 2018-09-24

Descriptive date:

Country code: PT

DATA

Documentation

- rep1
 - 02-ULusiada
 - 00-studio
 - 01-slices

Add representation Remove

3. Add descriptive metadata

4. SUBMISSION PACKAGES

Create SIP(s)

02-ULusiada: 1 reps, 3 folder(s), 22 files, 340.8 MB

Memory usage: 107.4 MB of 493.0 MB

RODA-In

File Edit Classification Scheme View Help

1. FILE EXPLORER

06-Dspace themes /Users/mferreira/My Documents/13-KEEF

2. INFORMATION PACKAGES

- 01-ESEPF
- 03-IPCB
- 04-UTAD
- Universidade Lusíada

3. INSPECTOR

New node MIXED

METADATA

Identity

Identifier: uuid-7b4dae0f-7c92-4308-8039-d4d775c9e5df

Description level: Fonds

Title: New node

Initial date

Final date

Acquisition date

Accession code

Classification

4. SUBMISSION PACKAGES

Create SIP(s)

Ignore Associate Add Remove

Memory usage: 354.3 MB of 493.0 MB

Naming scheme and output format

Creating SIPs

Selected 4/4 SIP

Export all items

Include hierarchy

Create inventory report

Output directory Choose...

SIP format E-ARK

SIP names ID E-ARK BagIt

Cancel

RODA-In

File Edit Classification Scheme View Help

1. FILE EXPLORER

06-Dspace themes /Users/mferreira/My Documents/13-KEE...

2. INFORMATION PACKAGES

- 01-ESEPF
- 03-IPCB
- 04-UTAD
- Universidade Lusfada

3. INSPECTOR

New node MIXED

METADATA </> ✓ + EAD 2002

Identity

Identifier uuid-7b4dae0f-7c92-4308-8039-d4d775c9e5df

Creating SIPs

Name	Date Modified	Size
inventory_report - 2018.09.24 16.46.57.180.csv	Today at 16:46	Zero bytes
inventory_report - 2018.09.24 16.47.21.242.csv	Today at 16:47	17 KB
uuid-3f53d299-cf9e-4382-9c3f-c57f57c2a39d.zip	Today at 16:47	6,6 MB
uuid-085db43c-12bd-4528-b7fe-94c8acfa5cb2.zip	Today at 16:47	27,8 MB
uuid-01801ae4-04c9-4f2b-8509-537c66d70536.zip	Today at 16:47	6,3 MB
uuid-33476bbe-0a00-48c1-b486-1dfcff20d664.zip	Today at 16:47	309,9 MB

Search

Associations

Associate files/directories to an information package to create a new association

4. SUBMISSION PACKAGES

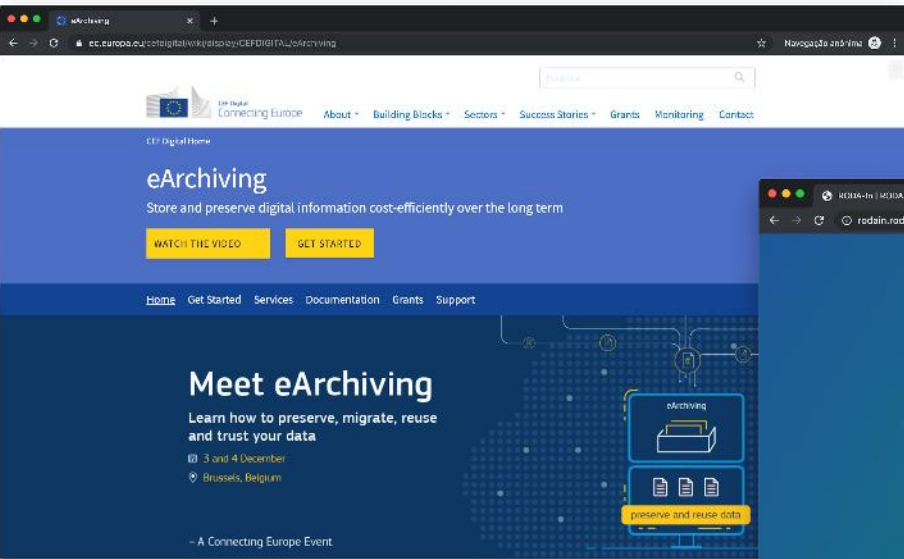
Create SIP(s)

Ignore Associate Add Remove

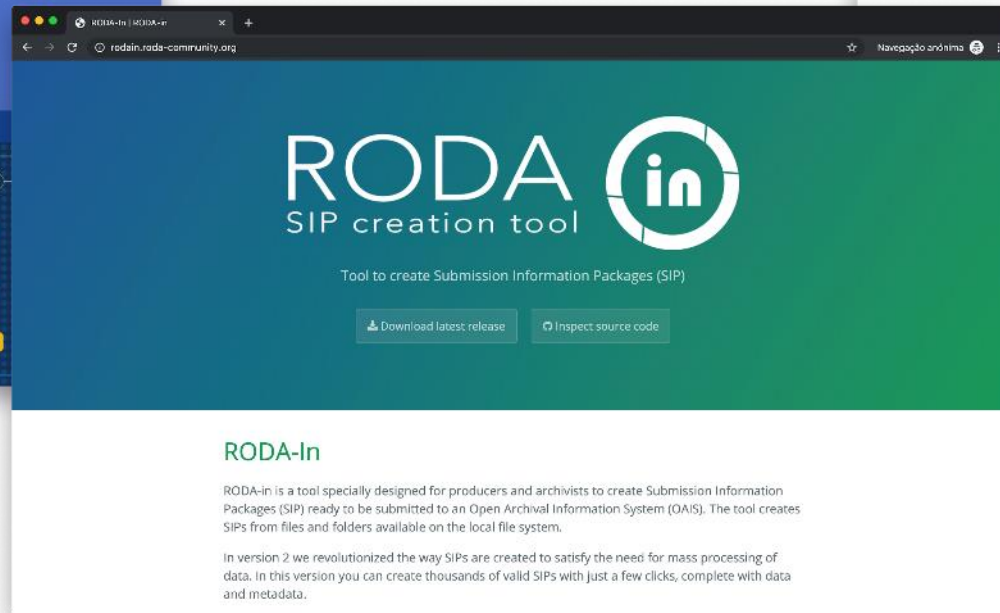
Memory usage: 343.5 MB of 478.5 MB

Inspect the outcome

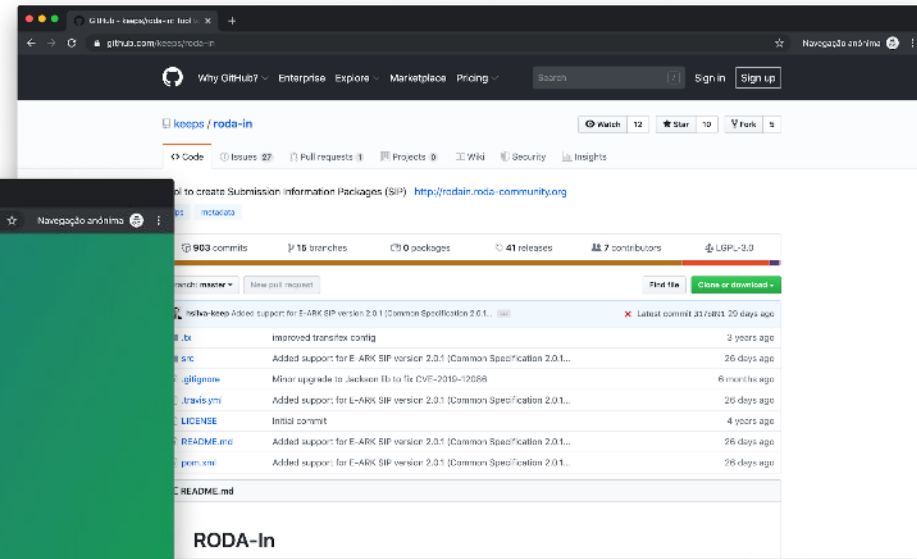
Where can I find it?



<https://ec.europa.eu/cefdigital/wiki/x/FgXvB>
(under Services)



rodain.roda-community.org



github.com/keeps/roda-in



RODA

What is RODA?



Long-term **digital repository**

Implements the main functional units of the OAIS reference model



Large corporations or **public bodies**

Scalable to millions of records



Fully featured repository

With the ability to integrate with existing systems via well defined protocols and formats



Conforms to **open standards**

OAIS, PREMIS, EAD, DC, METS, **E-ARK IPs**, BagIt, etc.



Vendor **independent**

RODA is 100% built on top of open-source technologies



Authenticity

The use of preservation metadata (**PREMIS**), together with **ISO 16363**, ensures reliability of the service and authenticity of the enclosed digital records



Pluggable **preservation actions**

Management and **Preservation Planning** tasks are handled by a job execution module



Horizontal scalability

An advanced indexing system enables discovery services to be spread through various servers for greater performance



Auditable

Users must be authenticated before accessing the repository. **All user actions and preservation events are logged** for future accountability



Integrations with 3rd party systems

Well documented **REST API** and convenient Java libraries available on Github

How does it work?

RODA 3

Welcome

Login English

Welcome to RODA!

An open-source digital repository designed for preservation

RODA is a digital repository solution 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 objects produced by large corporations or public bodies. RODA is based on open-source technologies and is supported by existing standards such as the Open Archival Information System (OAIS), Metadata Encoding and Transmission Standard (METS), Encoded Archival Description (EAD), Dublin Core (DC) and PREMIS (Preservation Metadata).

This is a only demonstration site. A general cleaning of the repository will be performed every night.

Credentials:

Username: **admin**
Password: **roda**



Conforms to open standards

RODA follows open standards using EAD for description metadata, PREMIS for preservation metadata, METS for structural metadata, and 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 be highly scalable, enabling the distribution of the processing load between several servers. Furthermore, the use of advanced indexing components enable RODA's discovery services to be spread through various servers on a cluster for even greater performance.



Embedded preservation actions

Preservation actions and management within RODA is handled by a job execution module. The job execution module allows the repository manager to run preservation tasks over a given set of data. Preservation actions include format conversions, checksum verifications, reporting (e.g. to automatically send SIP acceptance/rejection emails), virus checks, etc.



Authenticity



Support for multiple formats

Ingest...

RODA - WELCOME x +

demo.roda-community.org/?locale=en#welcome

Navegação anónima

RODA 3







Welcome Catalogue Search **Ingest** Administration Planning Help admin English

Pre-ingest
Transfer
Process
Assessment

Welcome to RODA!

An open-source digital repository designed for preservation

RODA is a digital repository solution 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 objects produced by large corporations or public bodies. RODA is based on open-source technologies and is supported by existing standards such as the Open Archival Information System (OAIS), Metadata Encoding and Transmission Standard (METS), Encoded Archival Description (EAD), Dublin Core (DC) and PREMIS (Preservation Metadata).

-  **Conforms to open standards**
RODA follows open standards using EAD for description metadata, PREMIS for preservation metadata, METS for structural metadata, and 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 be highly scalable, enabling the distribution of the processing load between several servers. Furthermore, the use of advanced indexing components enable RODA's discovery services to be spread through various servers on a cluster for even greater performance.
-  **Embedded preservation actions**
Preservation actions and management within RODA is handled by a job execution module. The job execution module allows the repository manager to run preservation tasks over a given set of data. Preservation actions include format conversions, checksum verifications, reporting (e.g. to automatically send SIP acceptance/rejection emails), virus checks, etc.
-  **Authenticity**
-  **Support for multiple formats**

RODA - INGEST x +

demo.roda-community.org/?locale=en#ingest/pre

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

Pre-ingest

The pre-ingest process depicts the ability of a Producer to create Submission Information Packages (SIP) containing both data and metadata (in a well-defined structure) in order to submit them to the repository for ingest. The SIPs created are expected to comply to the policies established by (or negotiated with) the repository.

The pre-ingest process usually comprises some or all of the following activities:

Submission agreement

This activity consists of the definition of the terms, pre-conditions and requirements for content, and accompanying information (e.g. metadata, documentation, contracts, etc.), to be sent to the repository by the Producer. It is materialised in a written agreement between the Producer and the Repository that specifies the type of content and all the legal and technical requirements that both parties are expected to comply.

Classification plan

During the signing of the submission agreement, the Producer must have agreed to a base Classification Scheme (or list of Collections) on which she will have explicit authorisation to deposit new information.

The base Classification Scheme is usually created by the Repository and can be downloaded in this section in machine readable format. The downloaded file can be loaded into RODA-in to better arrange and prepare Submission Information Packages before transferring them to the repository to be ingested.

[Download classification scheme](#) (note: downloading the classification scheme requires a RODA instance)

Submission Information Packages (SIP)

This activity consists of preparing of one or more Submission Information Packages (SIP) according to the technical and non-technical requirements defined on the Submission Agreement. To facilitate the creation of SIPs, Producers may take advantage of the RODA-in tool.

The tool and its documentation are available at <http://rodain.roda-community.org>.

Transfer of materials

RODA - TRANSFER x +

demo.roda-community.org/?locale=en#ingest/transfer

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

Ingest transfer

The Transfer area provides the appropriate temporary storage to receive Submission Information Packages (SIPs) from Producers. SIPs may be delivered via electronic transfer (e.g. FTP) or loaded from media attached to the repository. This page also enables the user to search files in the temporary transfer area, create/delete folders and upload multiple SIPs to the repository at the same time for further processing and ingest. The ingest process may be initiated by selecting the SIPs you wish to include in the processing batch. Click the "Process" button to initiate the ingest process.

Search... advanced

No parent

Could not find any transferred resources in this context.

REFRESH UPLOAD NEW FOLDER

About RODA Download Development Contact us powered by

RODA - TRANSFER x +

demo.roda-community.org/?locale=en#ingest/transfer

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

Ingest transfer

The Transfer area provides the appropriate temporary storage to receive Submission Information Packages (SIPs) from Producers. SIPs may be delivered via electronic transfer (e.g. FTP) or loaded from media attached to the repository. This page also enables the user to search files in the temporary transfer area, create/delete folders and upload multiple SIPs to the repository at the same time for further processing and ingest. The ingest process may be initiated by selecting the SIPs you wish to include in the processing batch. Click the "Process" button to initiate the ingest process.

1 transferred resource selected Search... advanced

No parent

<input checked="" type="checkbox"/>	<input type="checkbox"/>	Name	Size
<input checked="" type="checkbox"/>	<input type="checkbox"/>	uuid-8ecd9bc5-da63-4f69-87d7-5d515211e6b0.zip	2.4 MB

EXPORT

- Rename
- Move
- Remove
- Start new process

RODA - INGEST x +

demo.roda-community.org/?locale=en#process/create/ingest

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

New process

Name
Default ingest workflow (1.0)

Selected submission information packages (SIP)

Name	Size	Date created
uuid-8ecd9bc5-da63-4f69-87d7-5d515211e6b0.zip	2.4 MB	2019-11-27 10:48:20

EXPORT 1-1 of 1

Workflow

- Default ingest workflow (1.0)
- Minimal ingest workflow (1.0)

Default ingest workflow (1.0)

Performs all the tasks needed to ingest a SIP into the repository and therefore creating an AIP.

Categories: ingest

Format of the Submission Information Packages
Select the format of the Submission Information Packages to be ingested in this ingest process.

BagIt (1.0)
BagIt as a zip file

E-ARK SIP (1.0)
E-ARK SIP as a zip file.

Uploaded file/folder (1.0)
Treats a file/folder as a SIP.

Parent node

Process

CREATE

OBTAIN COMMAND

CANCEL

Parent node

Specifications  

Force parent node

Force the use of the selected parent node even if the SIPs provide information about the desired parent.

AIP Virus check

Scans Information Package(s) for malicious software using the Antivirus application ClamAV. Clam AntiVirus (ClamAV) is a free and open-source, cross-platform antivirus software toolkit able to detect many types of malicious software, including viruses. If malicious software is detected a report will be generated and a PREMIS event will record this occurrence.

Metadata validation

Checks if the descriptive metadata included in the Information Package is present, and if it is valid according to the XML Schemas installed in the repository. A validation report is generated indicating which Information Packages have valid and invalid metadata.

Fixity information computation

Computes file fixity information (also known as checksum) for all data files within an AIP, representation or file and stores this information in PREMIS objects within the corresponding entity. This task uses SHA-256 as the default checksum algorithm, however, other algorithms can be configured in "roda-core.properties". File fixity is the property of a digital file being fixed, or unchanged. "AIP corruption risk assessment" is the process of validating that a file has not changed or been altered from a previous state. In order to validate the fixity of an AIP or file, fixity information has to be generated beforehand.

File format identification (Siegfried)

Identifies the file format and version of data files included in Information Packages using the Siegfried tool (a signature-based file format identification tool that supports PRONOM identifiers and Mimetypes). The task updates PREMIS objects metadata in the Information Package to store the results of format identification. A PREMIS event is also recorded after the task is run.

PDF/A format validation (VeraPDF)

This action validates PDF files to make sure they comply to the PDF/A specification. PDF/A is an ISO-standardized version of the Portable Document Format (PDF) specialized for use in the archiving and long-term preservation of electronic documents. PDF/A differs from PDF by prohibiting features ill-suited to long-term archiving, such as font linking (as opposed to font embedding) and encryption. The specification for PDF/A is a set of restrictions and requirements applied to the "base" PDF standards (PDF 1.4 for PDF/A-1 and ISO 32000 for PDF/A-2 and PDF/A-3) plus a specific set of 3rd party standards. The outcome of this action is the creation of a new technical metadata file in the Archival Information Package (under the folder "metadata/other") that records the output of the VeraPDF tool. A PREMIS event is also recorded after running this task. For more information about VeraPDF, please visit <http://verapdf.org>

Feature extraction

Extraction of technical metadata using Apache Tika

Full-text extraction

Process

CREATE 

OBTAIN COMMAND 

CANCEL 

identifies the file format and version of data files included in Information Packages using the Sigstream tool (a signature-based file format identification tool that supports PRONOM identifiers and Mimetypes). The task updates PREMIS objects metadata in the Information Package to store the results of format identification. A PREMIS event is also recorded after the task is run.

PDF/A format validation (VeraPDF)

This action validates PDF files to make sure they comply to the PDF/A specification. PDF/A is an ISO-standardized version of the Portable Document Format (PDF) specialized for use in the archiving and long-term preservation of electronic documents. PDF/A differs from PDF by prohibiting features ill-suited to long-term archiving, such as font linking (as opposed to font embedding) and encryption. The specification for PDF/A is a set of restrictions and requirements applied to the "base" PDF standards (PDF 1.4 for PDF/A-1 and ISO 32000 for PDF/A-2 and PDF/A-3) plus a specific set of 3rd party standards. The outcome of this action is the creation of a new technical metadata file in the Archival Information Package (under the folder "metadata/other") that records the output of the VeraPDF tool. A PREMIS event is also recorded after running this task. For more information about VeraPDF, please visit <http://verapdf.org>

Feature extraction

Extraction of technical metadata using Apache Tika

Full-text extraction

Extraction of full-text using Apache Tika

Digital signature validation

Checks if digital signatures embedded in files are valid. The task supports the following formats: PDF, Microsoft Office Formats (.docx, .xlsx, .pptx) and OpenDocument Formats (.odt, .ods, .odp). The outcome of this action is three-fold: 1) the outcome of verification is stored in a PREMIS event; 2) the extracted digital signatures are stored within the AIP under the "metadata/other" folder; and 3) the files with the digital signature removed are stored under a new representation in the Archival Information Package (AIP).

Verify user authorization

Checks if the user has enough permissions to place the AIP under the desired node in the classification scheme

Auto accept

Adds information package to the inventory without any human appraisal. After this point, the responsibility for the digital content's preservation is passed on to the repository.


Ingest finished email notification


Send a notification after finishing the ingest process to one or more e-mail addresses (comma separated)


Ingest finished notification only when failed

If checked, the ingest finished notification will only be sent if a fail occurs during ingestion

Process

CREATE 

OBTAIN COMMAND 

CANCEL 

Ingest process

The Ingest process contains services and functions to accept Submission Information Packages (SIPs) from Producers, prepare Archival Information Packages (AIPs) for storage, and ensure that Archival Information Packages and their supporting Descriptive Information become established within the repository. This page lists all the ingest jobs that are currently being executed, and all the jobs that have been run in the past. In the right side panel, it is possible to filter jobs based on their state, user that initiated the job, and start date. By clicking on an item from the table, it is possible to see the progress of the job as well as additional details.

Jobs Search... advanced

<input type="checkbox"/>	Name	Creator	Start date	Duration	Status	Progress	Total	Successful	Failed
<input type="checkbox"/>	Default ingest workflow (1.0)	admin	2019-11-27 10:58:13	335ms	waiting	0%	1	0	0

EXPORT 1-1 of 1

Creators
 admin (1)

Status
 waiting to start (1)

Failures
 with failures (1)

RODA - 2A4BE97B-7A1C-42BE x +

demo.roda-community.org/?locale=en#process/job/2a4be97b-7a1c-42be-a3dd-f6c22f0ecf34

Welcome Catalogue Search Ingest Administration Planning Help

admin English

Process

Name
Default ingest workflow (1.0)

Creator
admin

Start date
2019-11-27 10:58:13 UTC

Duration
6 seconds

Status
running

Progress
14% done 1 total 1 processing

Source objects
A manually selected list with 1 transferred resources [DOWNLOAD](#)

Plugin
Default ingest workflow (1.0)

Format of the Submission Information Packages

Select the format of the Submission Information Packages to be ingested in this ingest process.

E-ARK SIP (1.0)
E-ARK SIP as a zip file.

Parent node
≡ Specifications

Force parent node
Force the use of the selected parent node even if the SIPs provide information about the desired parent.

AIP Virus check
Scans Information Package(s) for malicious software using the Antivirus application ClamAV. Clam AntiVirus (ClamAV) is a free and open-source, cross-platform antivirus software toolkit able to detect many types of malicious software,

Actions
[LIST CREATED PACKAGE ▶](#)

Package (under the folder "metadata/other") that records the output of the VeraPDF tool. A PREMIS event is also recorded after running this task. For more information about VeraPDF, please visit <http://verapdf.org>

- Feature extraction**
Extraction of technical metadata using Apache Tika
- Full-text extraction**
Extraction of full-text using Apache Tika
- Digital signature validation**
Checks if digital signatures embedded in files are valid. The task supports the following formats: PDF, Microsoft Office Formats (.docx, .xlsx, .pptx) and OpenDocument formats (.odt, .ods, .odp). The outcome of this action is three-fold: 1) the outcome of verification is stored in a PREMIS event; 2) the extracted digital signatures are stored within the AIP under the "metadata/other" folder; and 3) the files with the digital signature removed are stored under a new representation in the Archival Information Package (AIP).
- Verify user authorization**
Checks if the user has enough permissions to place the AIP under the desired node in the classification scheme
- Auto accept**
Adds information package to the inventory without any human appraisal. After this point, the responsibility for the digital content's preservation is passed on to the repository.
- Ingest finished notification only when failed**
If checked, the ingest finished notification will only be sent if a fail occurs during ingestion

Actions

[LIST CREATED PACKAGE ▶](#)

Reports

Search reports...



JobId is 2a4be97b-7a1c-42be-a3dd-f6c22f0ecf34

Submission Information	Archival Information Package	Last updated at	Status	Progress	Failed
uuid-8ecd9bc5-da63-4f69	CEF eArchiving - Release	2019-11-27 10:58:15	Running	2 of 7 (29%)	0

[EXPORT](#)

1-1 of 1

Report status

Running (1)

Successful inner processes

E-ARK SIP (1)

About RODA

What is RODA?
License
Acknowledgements

Download

Demo
Binary
Source code

Development

Developer guide
Translations
Roadmap
Bug reporting

Contact us

Community support
Commercial support
Send us a message

powered by
keep.
Preserving the future

RODA - 2A4BE97B-7A1C-42BE x +

demo.roda-community.org/?locale=en#process/job/report/2a4be97b-7a1c-42be-a3dd-f6c22f0ecf34-b6fed5f5-3c24-3601-80e6-f36842767194-cfeb677... ☆ Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

Information package processing report

Job
Default ingest workflow (1.0)

Source transferred resource
uuid-8ecd9bc5-da63-4f69-87d7-5d515211e6b0.zip (uuid-8ecd9bc5-da63-4f69-87d7-5d515211e6b0)

Outcome intellectual entity
CEF eArchiving - Release Management Strategy v.1.0.docx Active

Date created
2019-11-27 10:58:13 UTC

Date updated
2019-11-27 10:58:21 UTC

Duration
7 seconds

Status
Success

Progress
Executed 7 of 7 tasks (100%)

Run tasks

1. E-ARK SIP

Agent

Actions
BACK

Run tasks

1. E-ARK SIP

Agent

org.roda.core.plugins.plugins.ingest.EARKSIPToAIPPlugin (1.0)

Start datetime

2019-11-27 10:58:13 UTC

End datetime

2019-11-27 10:58:15 UTC

Outcome

Success

2. AIP Virus check

Agent

org.roda.core.plugins.plugins.antivirus.AntivirusPlugin (ClamAV 0.100.2/25206/Fri Dec 14 10:22:28 2018)

Start datetime

2019-11-27 10:58:15 UTC

End datetime

2019-11-27 10:58:20 UTC

Outcome

Success

Outcome details

```
/roda/data/storage/aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50: OK
```

```
----- SCAN SUMMARY -----
```

```
Infected files: 0
```

```
Time: 5.094 sec (0 m 5 s)
```

3. Metadata validation

Actions

BACK

Data management...

RODA - CATALOGUE x +

demo.roda-community.org/?locale=en#browse/d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help

admin English

Intellectual Entity > Digital Informatio...

Digital Information LifeCycle Interoperability Standards Board

0 risk incidences, 0 preservation events and 12 log entries

Created by admin on 2017-10-23 and last updated by admin on 2017-10-23

Encoded Archival Description 2002

Identity

Reference code
d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Description level

Fonds

Title
Digital Information LifeCycle Interoperability Standards Board

Initial date
2017-01-01

Final date
2017-01-30

Context

Biography or history
The Digital Information LifeCycle Interoperability Standards Board

interoperability specifications which allow for the transfer, long-te

ustain maintain a set of
e of the information

**EAD 2002 & Dublin Core included
by default**

RODA - CATALOGUE x +

demo.roda-community.org/?locale=en#browse/d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help

admin English

Intellectual Entity > Digital Informatio...

Digital Information LifeCycle Interoperability Standards Board

0 risk incidences, 0 preservation events and 12 log entries

Created by admin on 2017-10-23 and last updated by admin on 2017-10-23

Encoded Archival Description 2002

Identity

Reference code
d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Description level

Fonds

Title
Digital Information LifeCycle Interoperability Standards Board

Initial date
2017-01-01

Final date
2017-01-30

Context

Biography or history
The Digital Information LifeCycle Interoperability Standards Board
interoperability specifications which allow for the transfer, long-te

sustain maintain a set of
of the information

**Any schema is possible as long as
it is XML-based**

RODA - CATALOGUE x +

demo.roda-community.org/?locale=en#browse/d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help

admin English

Intellectual Entity > Digital Informatio...

Digital Information LifeCycle Interoperability Standards Board

0 risk incidences, 0 preservation events and 29 log entries

Created by admin on 2017-10-23 and last updated by admin on 2017-10-23

Encoded Archival Description 2002 **Dublin Core (2002-12-12)** +

Title
Digital Information LifeCycle Interoperability Standards Board

Creator
Hélder Silva

Identifier
d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Language
English

Sublevels Search... advanced

<input type="checkbox"/>	Level	Title
<input type="checkbox"/>	Series	Specifications

EXPORT

Description levels
 Series (1)

Representations
 without files (1)

Multiple schemas per record

Digital Information LifeCycle Interoperability Standards Board

0 risk incidences, 0 preservation events and 49 log entries

Created by admin on 2017-10-23 and last updated by admin on 2017-10-23

Encoded Archival Description 2002 Dublin Core (2002-12-12)

Identity

Reference code

d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Description level

Fonds

Title

Digital Information LifeCycle Interoperability Standards Board

Initial date

2017-01-01

Final date

2017-01-30

Context

Biography or history

The Digital Information LifeCycle Interoperability Standards Board (DILCIS Board) is an international group of experts committed to maintain interoperability specifications which allow for the transfer, long-term preservation, and reuse of digital information regardless of the origin

Intellectual Entity

AIP identifier

d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Created

by admin on 2017-10-23 17:54:55 UTC+1

Updated

by admin on 2017-10-23 17:54:55 UTC+1

Level

Fonds

Type

MIXED

Permissions

admin

administrators



RODA - CATALOGUE x +

demo.roda-community.org/?locale=en#browse/d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help

admin English

Intellectual Entity > Digital Informatio...

Digital Information LifeCycle Interoperability Standards Board

0 risk incidences, 0 preservation events and 36 log entries
Created by admin on 2017-10-23 and last updated by admin on 2017-10-23

Encoded Archival Description 2002 Dublin Core (2002-12-12)

Identity

Reference code
d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99

Description level

Fonds

Title
Digital Information LifeCycle Interoperability Standards Board

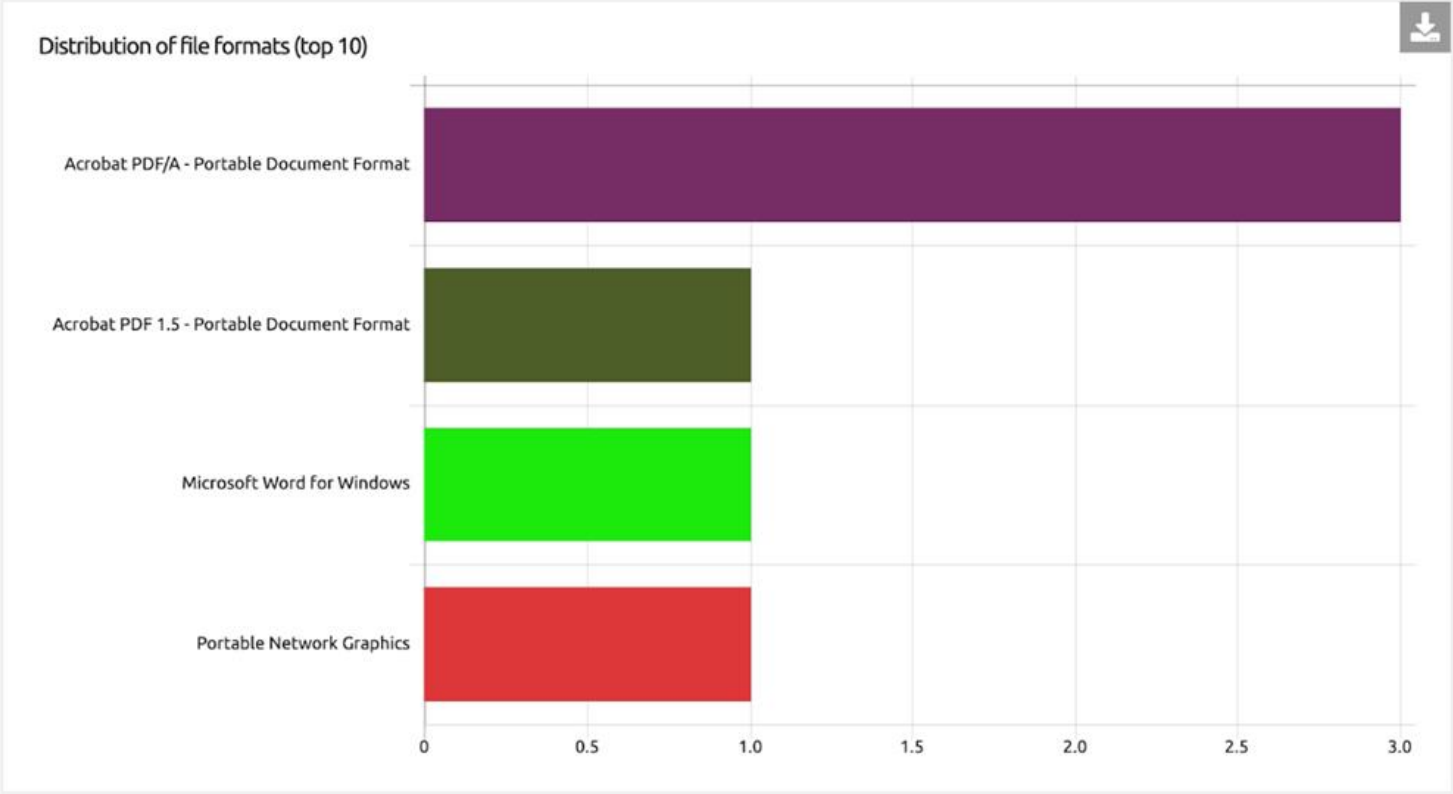
Initial date
2017-01-01

Final date
2017-01-30

Context

Biography or history
The Digital Information LifeCycle Interoperability Standards Board (DILCIS Board) is an international group of experts committed to maintain and sustain maintain a set of interoperability specifications which allow for the transfer, long-term preservation, and reuse of digital information regardless of the origin or type of the information.

- Change type
- Move
- Permissions
- Remove
- Start new process
- Download intellectual entity
- Download preservation metadata
- Download documentation



Archival storage...

AIPs are stored directly on the **file system**

Better **transparency**

Easier backups and **replication**

Well defined specifications for SIP, AIP, DIP

Managed by the **DILCIS Board** - a subgroup of DLM Forum

Part of the eArchiving CEF Building Block to promote **European adoption**

```
root@nucha: /home/hsilva
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage#
root@de88e4408c0e:/roda/data/storage# ls -lh
total 60K
drwxr-x--- 2 root roda 4.0K Nov 27 04:22 action-log
drwxr-x--- 9 root roda 4.0K Nov 27 10:58 aip
drwxr-x--- 2 root roda 4.0K Nov 27 04:22 dip
drwxr-x--- 2 root roda 4.0K Nov 27 10:58 job
drwxr-x--- 3 root roda 4.0K Nov 27 10:58 job-report
drwxr-x--- 3 root roda 4.0K Nov 27 04:22 preservation
drwxr-x--- 2 root roda 28K Nov 27 04:23 representation-information
drwxr-x--- 2 root roda 4.0K Nov 27 04:23 risk
drwxr-x--- 2 root roda 4.0K Nov 27 04:22 risk-incidence
root@de88e4408c0e:/roda/data/storage# ls -lh aip/
total 28K
drwxr-x--- 4 root roda 4.0K Nov 27 04:23 3b6aeb0d-4c01-4c9b-a4fe-7541660dc8d5
drwxr-x--- 4 root roda 4.0K Nov 27 04:23 92330592-0641-44e7-ab6d-86b593478beb
drwxr-x--- 4 root roda 4.0K Nov 27 04:23 99f33572-4b9f-4e58-b853-975db23ae831
drwxr-x--- 3 root roda 4.0K Nov 27 04:23 abe559bb-3cd9-485c-8181-3198a155dad2
drwxr-x--- 5 root roda 4.0K Nov 27 10:58 cfeb677d-3564-49db-9fb4-c98a7dde0f50
drwxr-x--- 3 root roda 4.0K Nov 27 04:23 d5ce5ffe-4a56-44b6-9939-1bd25d3f5a99
drwxr-x--- 4 root roda 4.0K Nov 27 04:23 ea416d77-39a4-4a13-8f68-49dc19a89acc
root@de88e4408c0e:/roda/data/storage#
```

AIPs are stored directly on the **file system**

Better **transparency**

Easier backups and **replication**

Well defined specifications for SIP, AIP, DIP

Managed by the **DILCIS Board** - a subgroup of DLM Forum

Part of the eArchiving CEF Building Block to promote **European adoption**

```
root@nucha: /home/hsilva
root@de88e4408c0e:/roda/data/storage# find aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/schemas
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/schemas/ead2002.xsd
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/schemas/mets.xsd
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/schemas/xlink.xsd
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1/data
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1/data/CEF eArchiving - Release Management Strategy v.1.0.docx
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1/metadata
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1/metadata/other
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1/metadata/other/Siegfried
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1/metadata/other/Siegfried/CEF eArchiving - Release Management
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1/metadata/preservation
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1/metadata/preservation/urn:roda:premis:representation:efa180cb
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep1/metadata/preservation/urn:roda:premis:file:CEF eArchiving - R
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/aip.json
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/descriptive
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/descriptive/ead2002.xml
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:f18eebb5-457c-4a1c-a759-5a5e35f2412d.x
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:ecb984aa-52f9-4642-a386-aff5b08306c8.x
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:18dea81b-2b86-4301-bd34-c6a6d8c3e7eb.x
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:afb7a695-ed00-400d-84b0-c00873a41fed.x
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:459458db-6e5f-4169-8815-923da9f46ea6.x
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:ed710ed3-30e2-4f18-bf6c-ddef42ee3026.x
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:89ee6193-f92a-4e63-8757-e75eb7f0720e.x
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:a5c1c3b8-384c-4002-87e6-4fd300cc180d.x
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:5697624b-dc27-4594-ae4e-b72a301eab23.x
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/metadata/preservation/urn:roda:premis:event:61fb74f4-0e57-4fef-bf2d-0c408aa3969e.x
root@de88e4408c0e:/roda/data/storage#
```

Preservation planning...

Representation network

Representation information is any information required to understand and render both the digital material and the associated metadata. Digital objects are stored as bitstreams, which are not understandable to a human being without further data to interpret them. Representation information is the extra structural or semantic information, which converts raw data into something more meaningful. This page allows the preservation expert to define Representation Information and to link it to Intellectual Entities in the repository.

Representation Information Search representation information... advanced

Name	Support level	Family
A-Law telephony companding algorithm, from ITU-T G.711 sound	Known	fileformat
AMR, Adaptive Multi-Rate Speech Codec sound	Known	fileformat
AMR-WB+, Extended Adaptive Multi-Rate - Wideband Speech Codec sound	Known	fileformat
AMR-WB, Adaptive Multi-Rate - Wideband Speech Codec (G.722.2) sound	Known	fileformat
ARC_IA, Internet Archive ARC file format. aggregate	Known	fileformat
AVI (Audio Video Interleaved) moving-image	Known	fileformat
AVI (Audio Video Interleaved) with Cinepak Video Codec moving-image	Known	fileformat
AVI (Audio Video Interleaved) with DivX Video Codec moving-image	Known	fileformat
AVI (Audio Video Interleaved) with Indeo Video Codec moving-image	Known	fileformat
AVI File Format with OpenDML Extensions, Version 1.02 moving-image	Known	fileformat
AVI OpenDML File Format with FFV1 video encoding moving-image	Known	fileformat
AVI OpenDML File Format with UYUV Video Encoding moving-image		
AVI OpenDML File Format with V210 Video Encoding moving-image		
AVI OpenDML File Format with YUY2 Video Encoding moving-image		
AVI OpenDML with DV Digital Video moving-image		

Tags

- aggregate (9)
- any (11)
- database (7)
- dataset (27)
- moving-image (103)
- other (2)
- presentation (3)
- sound (58)
- spreadsheet (2)
- still-image (62)
- text (37)

Support level

- Known (377)




Family

- File format (377)

Relations with RI, AIP, Representations or Files

Risk register

The risk register lists all identified risks that may affect the repository. It should be as comprehensive as possible to include all identifiable threats, and generally contain an estimated probability of each risk event occurring, the severity or possible impact of the risk, and its probable timing or anticipated frequency. Risk mitigation is the process of defining actions to enhance opportunities and reduce threats to repository objectives.

 Risks advanced  

<input type="checkbox"/>	Name	Risk owner	Identified on	▼ Severity	Incidences	Not mitigat
<input type="checkbox"/>	Loss of information reliability	Preservation	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Hardware failure or incompatibility	Technical	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Non-availability of core utilities	Management	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Identifier to information referential integrity is compromised	Preservation	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Repository loses mandate	Management	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Media degradation or obsolescence	Technical	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Preservation plans cannot be implemented	Preservation	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Accidental system disruption	Technical	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Liability for regulatory non-compliance	Legal	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Financial shortfalls or income restrictions	Budgeting	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Destruction of primary documentation	Management	2016-04-01	Moderate	0	0
<input type="checkbox"/>	Inability to validate effectiveness of ingest process	Preservation				
<input type="checkbox"/>	Externally motivated changes or maintenance to information during ingest	Ingest				

Categories

- Data storage (1)
- Hardware, software or communication (1)
- Operations and service delivery (49)
- Personnel, management and administration (1)
- Physical environment (11)

Owner

- Budgeting (5)
- Dissemination (5)
- Documentation (5)
- Ingest (4)
- Legal (4)
- Management (23)
- Personnel (3)
- Preservation (15)
- Technical (15)

Severity

- High (1)
- Moderate (78)

Technical & non technical risks

RODA - PLANNING x +

demo.roda-community.org/?locale=en#planning/events

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

🕒 Preservation events

A preservation event aggregates metadata about actions, specifically documenting which objects it affects and which human or software agents intervened. Documentation of actions that modify an object is critical to maintaining digital provenance, a key element of authenticity. Actions that create new relationships or alter existing relationships are important in explaining those relationships. Even actions that alter nothing, such as validity and integrity checks on objects, can be important to record for management purposes.

Search... advanced 🔍

<input type="checkbox"/>	▼ Date	Type	Detail	Outcome
<input type="checkbox"/>	2019-11-27 10:58:21	ingest end	The ingest process has ended.	Success
<input type="checkbox"/>	2019-11-27 10:58:21	accession	Added package to the inventory. After this point, the responsibility for the digital content's preservation is passed on to the repository.	Success
<input type="checkbox"/>	2019-11-27 10:58:21	authorization check	User permissions have been checked to ensure that he has sufficient authorization to store the AIP under the desired node of the classification scheme.	Success
<input type="checkbox"/>	2019-11-27 10:58:21	format identification	Identified the object's file formats and versions using Siegfried.	Success
<input type="checkbox"/>	2019-11-27 10:58:21	message digest calculation	Created base PREMIS objects with file original name and file fixity information (SHA-256).	Success
<input type="checkbox"/>	2019-11-27 10:58:20	wellformedness check	Checked whether the descriptive metadata is included in the SIP and if this metadata is valid according to the established policy.	Success
<input type="checkbox"/>	2019-11-27 10:58:20	virus check	Scanned package for malicious programs using ClamAV.	Success
<input type="checkbox"/>	2019-11-27 10:58:15	wellformedness check	Checked that the received SIP is well formed, complete and that no unexpected files were included	Success
<input type="checkbox"/>	2019-11-27 10:58:15	unpacking	Extracted object	
<input type="checkbox"/>	2019-11-27 10:58:13	ingest start	The ingest process	

PREMIS events

Entity

- Intellectual entity (22)
- Representation (13)

Type

- accession (1)
- authorization check (1)
- creation (11)
- format identification (9)
- ingest end (1)
- ingest start (1)
- message digest calculation (1)
- unpacking (1)
- update (6)
- virus check (1)
- wellformedness check (2)

Outcome

- Success (35)

RODA - CREATE_JOB x +

demo.roda-community.org/?locale=en#process/create_job

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help

admin English

New process

Name
AIP feature extraction (ExifTool) (1.0)

Categories
 validation characterization management risk management feature extraction experimental format identification conversion dissemination misc reindex

Workflow

- ✓ AIP Virus check (ClamAV 0.100.2/25206/Fri Dec 14 10:22:28 ...)
- 👤 AIP ancestor hierarchy fix (1.0)
- 👤 AIP batch export (1.0)
- ⚠️ AIP corruption risk assessment (1.0)
- i AIP feature extraction (ExifTool) (1.0)**
- i AIP feature extraction (FITS) (1.0)
- i AIP feature extraction (JHOVE) (1.0)
- i AIP feature extraction (MediaInfo) (1.0)
- i AIP feature extraction (avprobe) (1.0)
- 🔄 AIP file format identification (DROID) (1.0)
- 👤 Activity log truncation (1.0)
- ⚙️ Audio conversion (SoX) (14.4.1)

AIP feature extraction (ExifTool) (1.0)

ExifTool is a platform-independent application capable of reading technical metadata from a wide variety of file formats.

ExifTool supports many different metadata formats including EXIF, GPS, IPTC, XMP, JFIF, GeoTIFF, ICC Profile, Photoshop IRB, FlashPix, AFPCP and ID3, as well as the maker notes of many digital cameras by Canon, Casio, DJI, FLIR, FujiFilm, GE, HP, JVC/Victor, Kodak, Leaf, Minolta/Konica-Minolta, Motorola, Nikon, Nintendo, Olympus/Epson, Panasonic/Leica, Pentax/Asahi, Phase One, Reconyx, Ricoh, Samsung, Sanyo, Sigma/Foveon and Sony.

The task creates a new file under the [AIP_ID]/representation/metadata/other/ExifTool. This information is not yet added to PREMIS or indexed but it can be inspected by downloading the AIP. A PREMIS event is recorded after the task is run.

For a
http

Process

- CREATE
- OBTAIN COMMAND
- CANCEL

Preservation actions

Access...



Catalogue

The catalogue is the inventory of all items or records found in the repository. A record can represent any information entity available in the repository (e.g. book, electronic document, image, database export, etc.). Records are typically aggregated in collections (or fonds) and subsequently organised in subcollections, sections, series, files, etc. This page lists all the top-level aggregations in the repository. You may drill-down to sub-aggregations by clicking on any of the items in the table below.

Search...



<input type="checkbox"/> Level	Title	▲ Dates	
<input type="checkbox"/> Fonds	Digital Information LifeCycle Interoperability Standards Board	2017-01-01 to 2017-01-30	

EXPORT

1-1 of 1

Description levels

- Fonds (1)

Representations

- without files (1)

Search

In this page you can search for Intellectual Entities, Representations or Files (use the down arrow to select the search domain). For each one of these domains you can search in all its properties or in specific properties (use the down arrow to expand the advanced search). For example, if you select Intellectual Entities, you can search in a specific field of the descriptive metadata, or find files of a certain format if the Files advanced search is selected.

The search engine locates only whole words. If you want to search for partial terms you should use the '*' operator. For more information on the available search operators, take a look at the [help page](#).

- Intellectual Entities ^
- Intellectual Entities
- 🔗 Representations
- 📁 Files

Search...

advanced



		▲ Dates	
<input type="checkbox"/>	Series Specifications	2017-01-01 to 2017-01-31	
<input type="checkbox"/>	File Archival Information Package (AIP)	2017-01-27 to 2017-01-27	
<input type="checkbox"/>	File Common Specification for Information Packages	2017-02-01 to 2017-02-01	
<input type="checkbox"/>	File Dissemination Information Package (DIP)	2017-02-01 to 2017-02-01	
<input type="checkbox"/>	File Submission Information Package (SIP)	2017-02-01 to 2017-02-01	

Description levels

- File (5)
- Fonds (1)
- Series (1)

Representations

- without files (2)
- with files (5)

EXPORT

1-7 of 7

RODA - SEARCH x +

demo.roda-community.org/?locale=en#search

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

Search

In this page you can search for Intellectual Entities, Representations or Files (use the down arrow to select the search domain). For each one of these domains you can search in all its properties or in specific properties (use the down arrow to expand the advanced search). For example, if you select Intellectual Entities, you can search in a specific field of the descriptive metadata, or find files of a certain format if the Files advanced search is selected.

The search engine locates only whole words. If you want to search for partial terms you should use the '*' operator. For more information on the available search operators, take a look at the [help page](#).

Intellectual Entities Search... advanced

- Original reference
- Title
- Description
- Origination
- Ingest SIP identifier
- Date

008-04-01 2016-06-20

ADD SEARCH FIELD CLEAR SEARCH

Level	Title	Dates
<input type="checkbox"/> File	CEF eArchiving - Release Management Strategy v.1.0.docx	
<input type="checkbox"/> Fonds	Digital Information LifeCycle Interoperability Standards B	
<input type="checkbox"/> Series	Specifications	

Description levels

- File (5)
- Fonds (1)
- Series (1)

Representations

Search by metadata attributes

RODA - SEARCH x +

demo.roda-community.org/?locale=en#search

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

Search

In this page you can search for Intellectual Entities, Representations or Files (use the down arrow to select the search domain). For each one of these domains you can search in all its properties or in specific properties (use the down arrow to expand the advanced search). For example, if you select Intellectual Entities, you can search in a specific field of the descriptive metadata, or find files of a certain format if the Files advanced search is selected.

The search engine locates only whole words. If you want to search for partial terms you should use the '*' operator. For more information on the available search operators, take a look at the [help page](#).

Files Search... advanced

- Filename
- Format
- Format version
- PRONOM
- Mimetype
- Extension
- Filesize 108 B
- Fulltext

ADD SEARCH FIELD CLEAR SEARCH

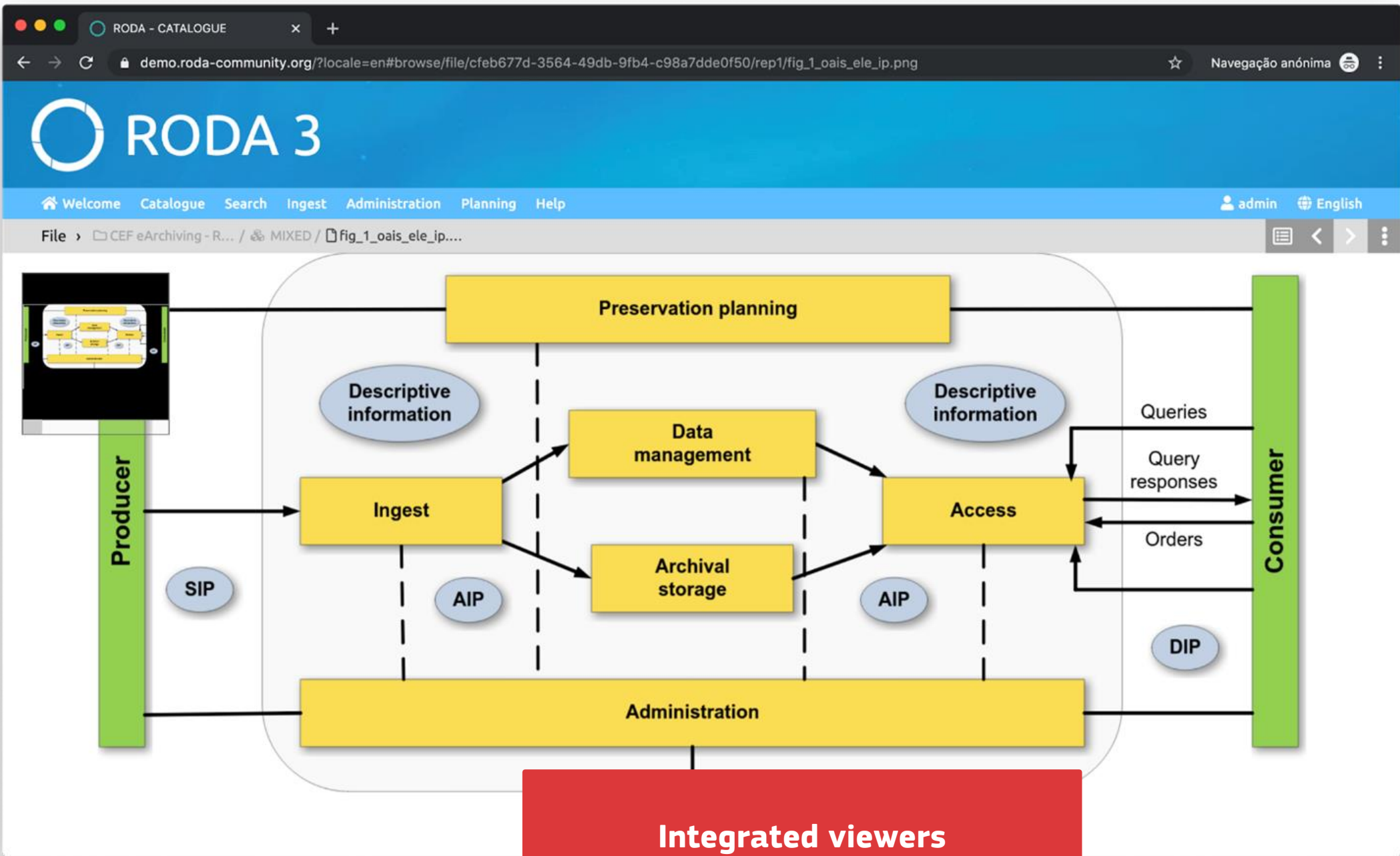
File type
 regular file (5)

Path
CEF eArchiving - Release Management Strategy v.1.0.docx

Search by file attributes

The screenshot displays the RODA 3 web application interface. At the top, there is a blue header with the RODA 3 logo and navigation links: Welcome, Catalogue, Search, Ingest, Administration, Planning, and Help. The user is logged in as 'admin' and the language is set to 'English'. Below the header, a breadcrumb trail shows the current location: File > Common Specific... / Text document / Common_Specific... The main content area is a document viewer showing a document titled 'COMMON SPECIFICATION FOR INFORMATION PACKAGES' by the 'DLM Archival Standards Board'. The document is displayed at 90% zoom. On the left side, there is a sidebar with a list of document thumbnails. A red box at the bottom of the viewer area contains the text 'Integrated viewers'.

Integrated viewers



RODA - CATALOGUE x +

demo.roda-community.org/?locale=en#browse/file/cfeb677d-3564-49db-9fb4-c98a7dde0f50/rep1/fig_1_oais_ele_ip.png

Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

File > CEF eArchiving - R... / MIXED / fig_1_oais_ele_ip...

File

Filename
fig_1_oais_ele_ip.png

Size
159.3 KB

Extension
png

Mimetype
image/png

Format
Portable Network Graphics 1.0

PRONOM
fmt/11

Fixity
4C22A38E5325C102773309C8C0799FAF0769AACA (SHA-1, RODA)
9A403E0D47BD7E48FEC33F18F363A4BC491C10A5EB46846CAA
B30F07729EFAC (SHA-256, RODA)
F9A38AF46ED1FB2DFA0187D1D0E1E110 (MDS, RODA)

Storage path
aip/cfeb677d-3564-49db-9fb4-c98a7dde0f50/representations/rep
1/data/fig_1_oais_ele_ip.png

Risks
0 risk incidences

its
preservation events

Technical info.


RODA - CATALOGUE x +

demo.roda-community.org/?locale=en#browse/file/cfeb677d-3564-49db-9fb4-c98a7dde0f50/rep1/CEF+eArchiving+-+Release+Management+Strategy+v.... ☆ Navegação anónima

RODA 3

Welcome Catalogue Search Ingest Administration Planning Help admin English

File > CEF eArchiving - R... / MIXED / CEF eArchiving - R...



File preview not supported

[DOWNLOAD](#)

Download if no viewer available

Worthwhile mentioning...

API documentation x +

demo.roda-community.org/api-docs/ Navegação anónima

swagger /api/swagger.json Explore

v1 aips

- GET** /v1/aips List AIPs
- POST** /v1/aips Create AIP
- PUT** /v1/aips Update AIP
- GET** /v1/aips/{aip_id} Get AIP
- DELETE** /v1/aips/{aip_id} Delete AIP
- GET** /v1/aips/{aip_id}/descriptive_metadata List descriptive metadata
- GET** /v1/aips/{aip_id}/descriptive_metadata/{metadata_id} Get descriptive metadata
- POST** /v1/aips/{aip_id}/descriptive_metadata/{metadata_id} Create descriptive metadata
- PUT** /v1/aips/{aip_id}/descriptive_metadata/{metadata_id} Update descriptive metadata
- DELETE** /v1/aips/{aip_id}/descriptive_metadata/{metadata_id} Delete descriptive metadata
- GET** /v1/aips/{aip_id}/{part} Download part of the AIP
- GET** /v1/aips/{aip_id}/preservation_metadata Get pr...
- POST** /v1/aips/{aip_id}/preservation_metadata Create...

Well documented REST API

Central Authentication Service

Local user database

CAS v1, v2 and v3

SAML v1 and v2 Protocol

OAuth

OpenID

WS-Federation Passive Requestor

Delegated authentication

... more

RODA

USERNAME

PASSWORD

Remember me

[Forgot password?](#)

LOGIN

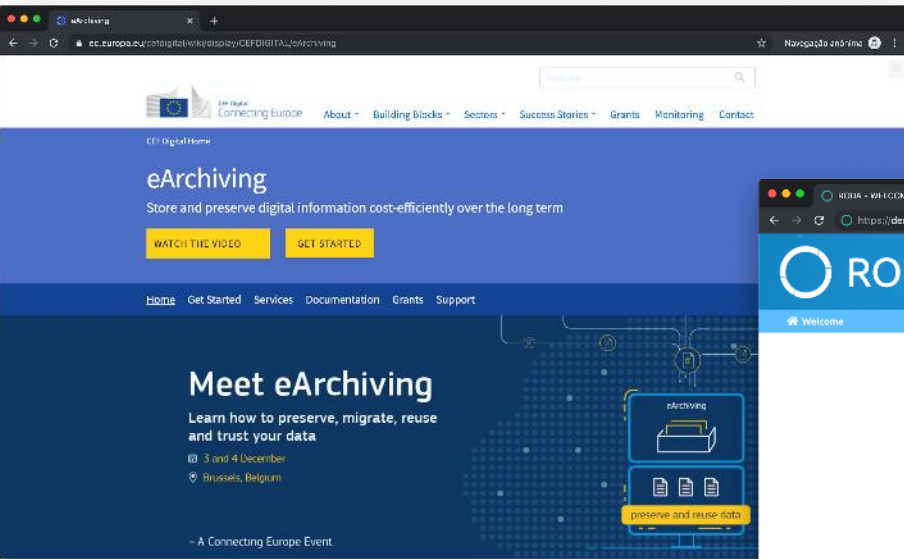
Please fill in your username and password before continuing...

OR

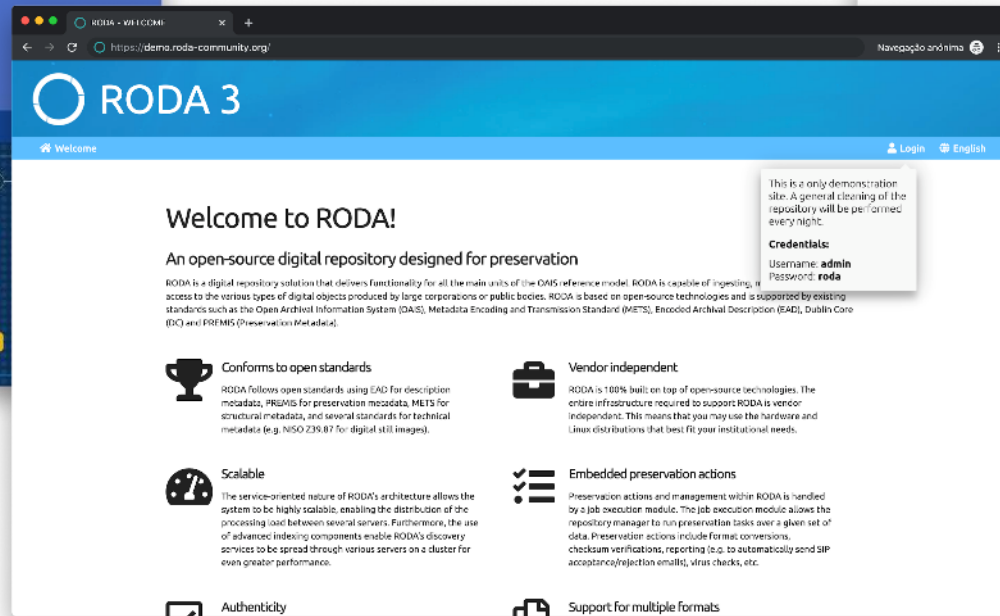
 Sign up with Google

For security reasons, please log out and exit your web browser when you are done accessing services that require authentication!

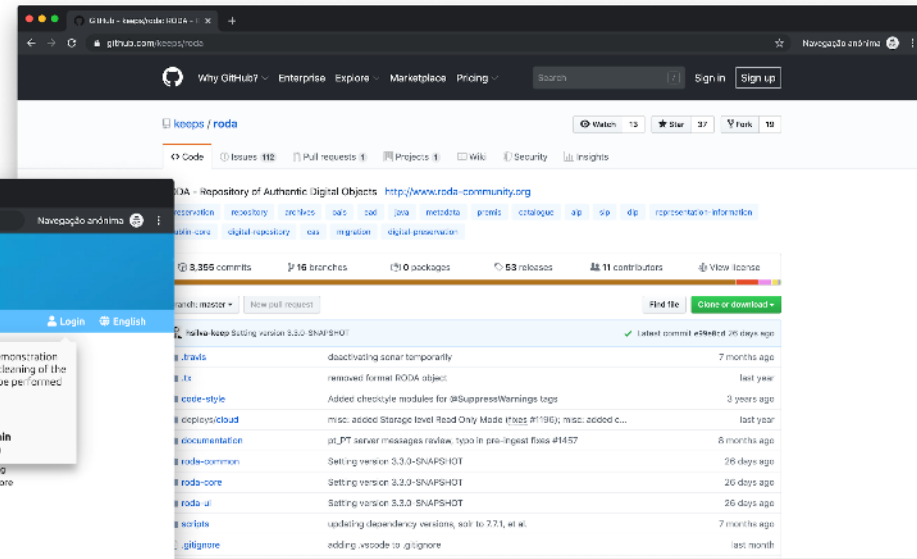
Where can I find it?



<https://ec.europa.eu/cefdigital/wiki/x/FgXvB>
(under Services)



www.roda-community.org



github.com/keeps/roda

Use cases

-
- *High availability for access (Hospital)*
 - *High diversity of file formats (National Archive)*
 - *Seamless integration with live data systems (University)*
 - *High volume of files (EUPO)*
 - *High throughput on ingest (TAXUD)*

Maria Kardami, from the Publications Office of the European Union (EUPO)

Ulrich Kröner, from the DG Taxation and Customs Union (TAXUD)

Thank you for your attention.

Any questions?

Hélder Silva
hsilva@keep.pt

Contact us

 info@keep.pt

© European Union, 2019. All rights reserved. Certain parts are licensed under conditions to the EU.
Reproduction is authorized provided the source is acknowledged.

Software presentation

eArchiving tools for database preservation



Luís Faria

Innovation Director,
KEEP Solutions, Portugal



A common scenario

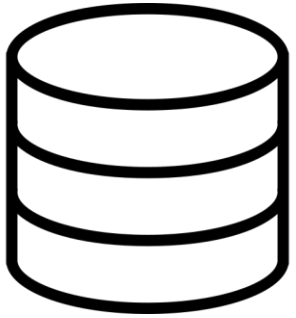
An **information system has been in use for 15 years**

It is about to be **decommissioned** as it has been replaced by a new, more advanced system

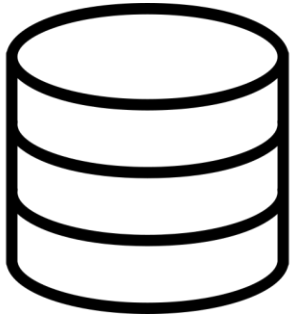
There is an interest in maintaining the data produced by the legacy information system for **legal and historical reasons**

This means that the system's database has been selected for **long-term archival**

Lets see how it works...



Database



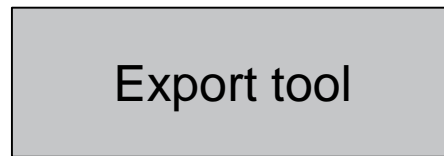
Database



Export tool



Database

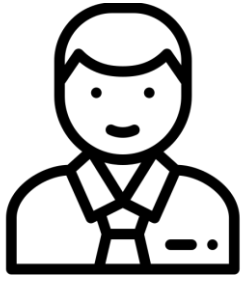


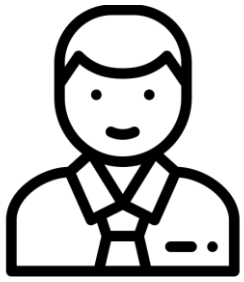
Export tool



Preservation format

*Years later, a user **wants to access** the data ...*

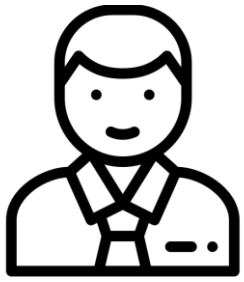




Find



Repository



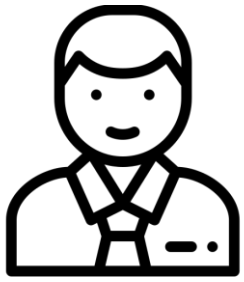
Find



Retrieve



Preservation
format



Find



Repository



Retrieve



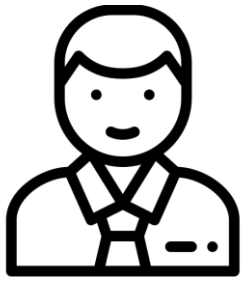
Preservation
format



View



Browse content on
the Web



Find

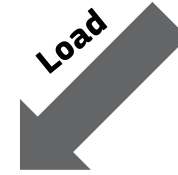
Repository



Retrieve



Preservation
format



Load



View



Live modern database
system



Browse content on
the Web

IT department is happy!

Because it no longer needs to maintain the legacy database system

Management is happy!

Because costs have been greatly reduced

Users are happy!

Has they can access information in modern web browser or desktop app

Not just of obsolete databases but also for continuous delivery

Some national archives mandate delivery of public administration databases every five years.



Tools in the toolkit



DBPTK Desktop

Desktop application to save database to preservation format, validate it, and browse and search the content



DBPTK Enterprise

Web application to browse and search on the content of multiple large preserved databases



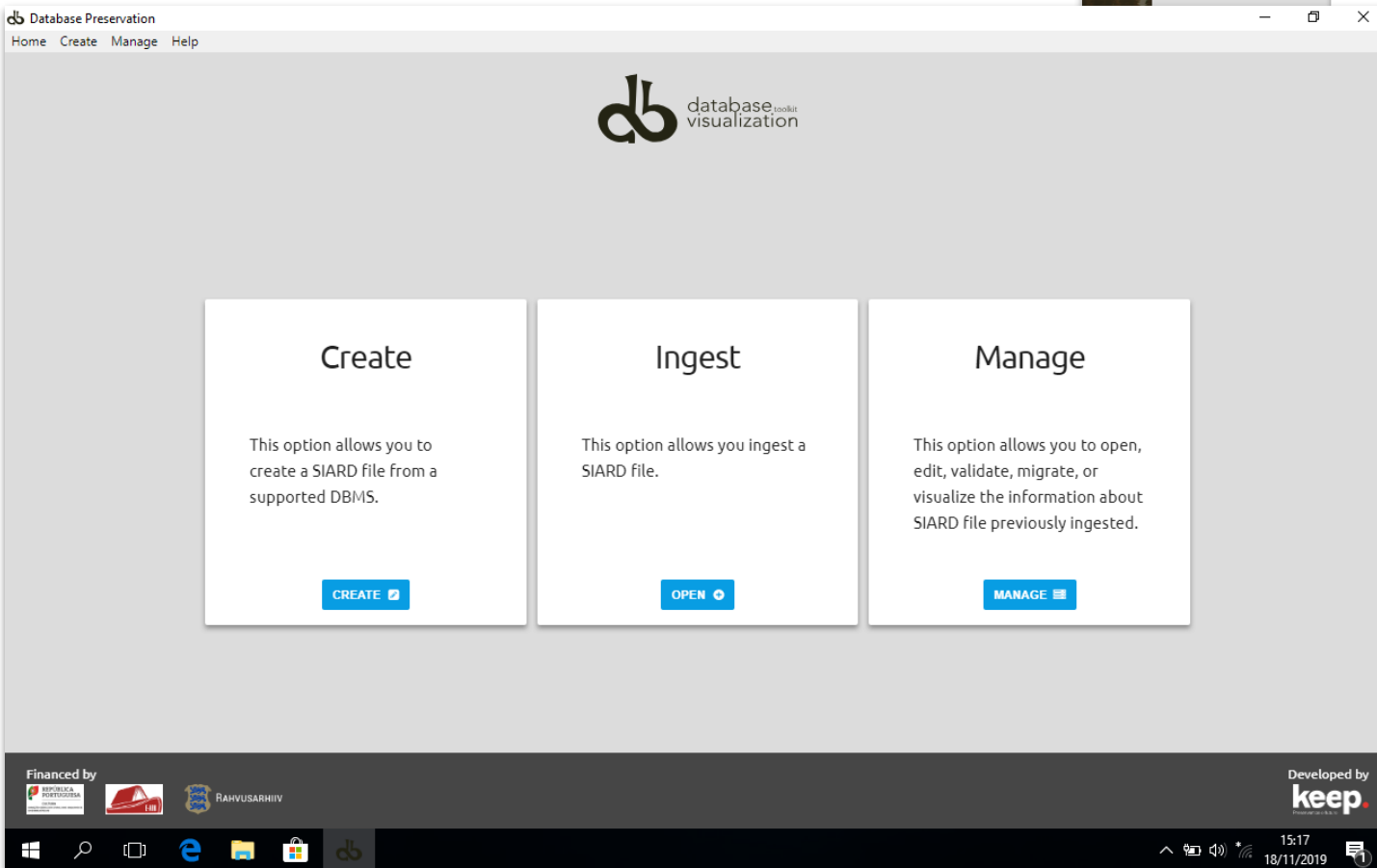
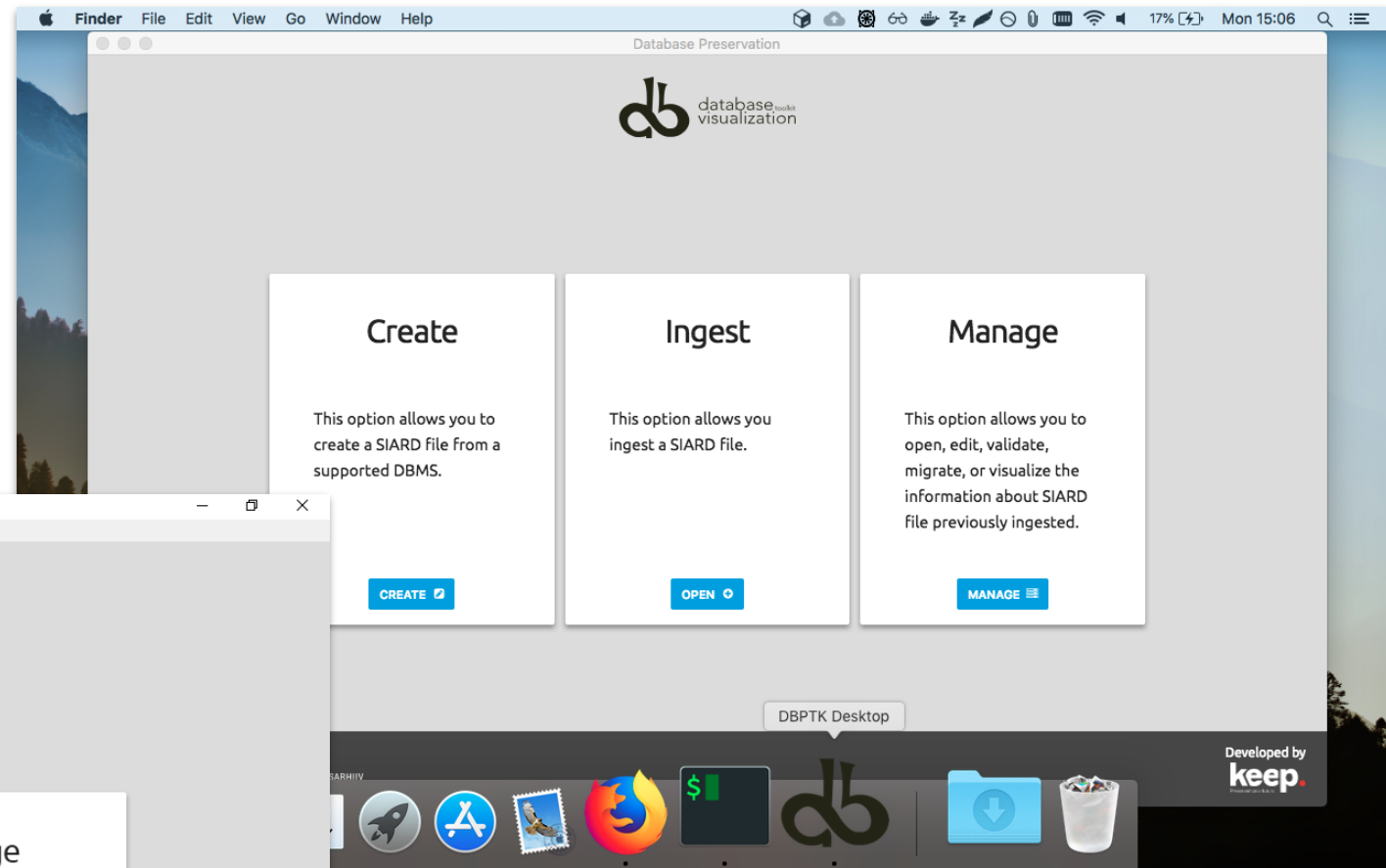
DBPTK Developer

A command-line tool and development library for automation and system integration



DBPTK Desktop

Desktop Application



Also available on Linux



SIARD creation

Export database to a preservation format

Connect to a local or remote database and save all content into a preservation format like SIARD

DBPTK Desktop

Home Create Manage Help

Home > Create SIARD - Connection

DBMS

- JDBC
- Microsoft SQL Server
- MySQL**
- Oracle
- PostgreSQL
- Progress Openedge
- Sybase

General SSH Tunnel

Hostname * localhost
The name of the database server host (e.g. localhost)

Port number 3306
The server port number

Username * root
The name of the user to use in connection

Password *
The password of the user to use in connection

Database * sakila
The name of the database to connect

Disable Encryption
Use to turn off encryption in the connection

Count Rows
Count the number of rows for each table

TEST CONNECTION ⚡

SIARD validation

Validate archived database

Validate SIARD against specification plus many additional checks for a thorough validation

DBPTK Desktop

Home Create Manage Help

Home > Databases > sakila > SIARD Validation

Validator

Database name	sakila	SIARD version:	siard-2.1
Requirements that failed:	3	SIARD specification:	SIARD-2.1
Number of errors:	203	Additional checks specification:	OPEN
Number of passed:	88	Report	OPEN
Number of warnings:	0		
Number of skipped:	12		

Status Validation Failed

Scroll to End of Log

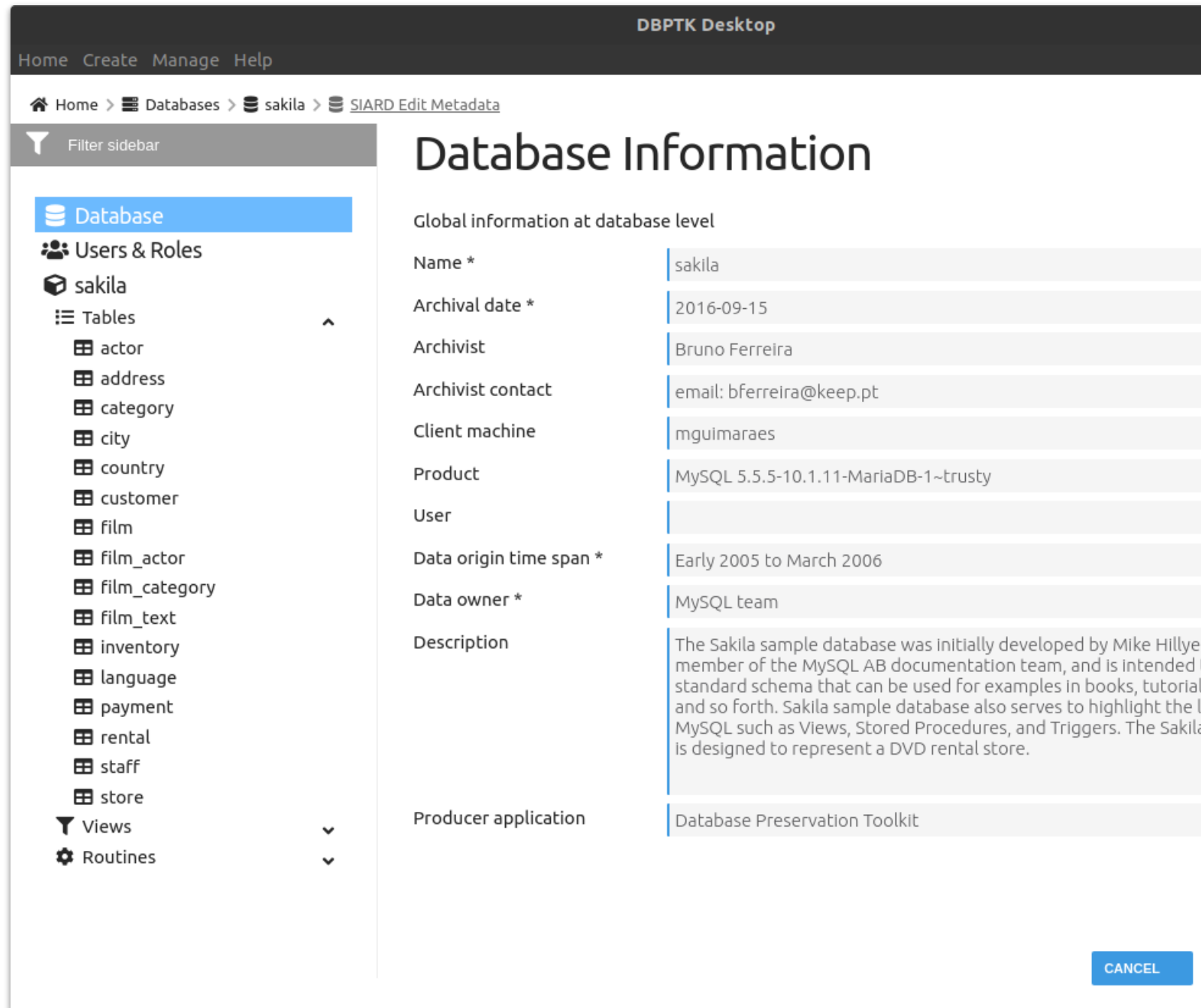
M_5.5-1-1	The name field is mandatory, check if it was provided	OK
A_M_5.5-1-1	Check if the table name element exists, has 3 or more characters and is not composed of blanks	OK
M_5.5-1-2	The folder field is mandatory, check if it was provided	OK
A_M_5.5-1-3	Check if the table description element exists, has 3 or more characters and is not composed of blanks	OK
M_5.5-1-4	The columns field is mandatory, check if it was provided	OK
M_5.5-1-10	The rows field is mandatory, check if it was provided	OK
5.6 Column level metadata		
M_5.6-1-1	The name field is mandatory, check if it was provided	ERROR
A_M_5.6-1-2	The lobFolder should exist if the column is a large object type and the corresponding table.xml column contains date, also check if the file attribute in the table.xml column contains a valid path to lobFolder	OK
M_5.6-1-3		OK
A_M_5.6-1-12	Check if the column description element exists, has 3 or more characters and is not composed of blanks	OK
5.7 Field level metadata		
M_5.7-1	All metadata that are designated as mandatory in the metadata.xsd at field level must be completed appropriately.	SKIPPED
5.8 Primary Key level metadata		
M_5.8-1-1	The name field is mandatory, check if it was provided	OK

[BACK](#) [RUN AGAIN](#)

Edit SIARD metadata

Enrich archived database with descriptions

Add descriptions to database, tables and columns to better understand its contents



The screenshot shows the DBPTK Desktop interface. The top navigation bar includes 'Home', 'Create', 'Manage', and 'Help'. The breadcrumb trail is 'Home > Databases > sakila > SIARD Edit Metadata'. A sidebar on the left, titled 'Filter sidebar', contains a tree view of the database structure: 'Database' (selected), 'Users & Roles', 'sakila' (expanded), 'Tables' (expanded), and 'Views' and 'Routines'. The 'Tables' list includes actor, address, category, city, country, customer, film, film_actor, film_category, film_text, inventory, language, payment, rental, staff, and store. The main area is titled 'Database Information' and displays global information at the database level for 'sakila'. The information is presented in a table-like format with labels and values.

Global information at database level	
Name *	sakila
Archival date *	2016-09-15
Archivist	Bruno Ferreira
Archivist contact	email: bferreira@keep.pt
Client machine	mguimaraes
Product	MySQL 5.5.5-10.1.11-MariaDB-1~trusty
User	
Data origin time span *	Early 2005 to March 2006
Data owner *	MySQL team
Description	The Sakila sample database was initially developed by Mike Hillyer, a member of the MySQL AB documentation team, and is intended as a standard schema that can be used for examples in books, tutorials, and so forth. Sakila sample database also serves to highlight the features of MySQL such as Views, Stored Procedures, and Triggers. The Sakila database is designed to represent a DVD rental store.
Producer application	Database Preservation Toolkit

CANCEL

Search records

Browse and search database content

Google-like search on the database content.

Drill down on specific tables and do advanced search for specific fields to find exactly what you are looking for.

The screenshot shows the DBPTK Desktop application interface. At the top, there is a navigation bar with 'Home', 'Create', 'Manage', and 'Help'. Below it is a breadcrumb trail: 'Home > Databases > sakila > Search'. A 'Filter sidebar' is visible on the left, containing sections for 'Information', 'Search all records' (highlighted), 'Saved searches', and 'Tables'. The 'Tables' list includes: actor, address, category, city, country, customer, film, film_actor, film_category, film_text, inventory, language, payment, rental, staff, store, actor_info, customer_list, and film_list. The main area displays 'Search all records' with a search input containing 'dan'. Below this, two tables are shown:

actor

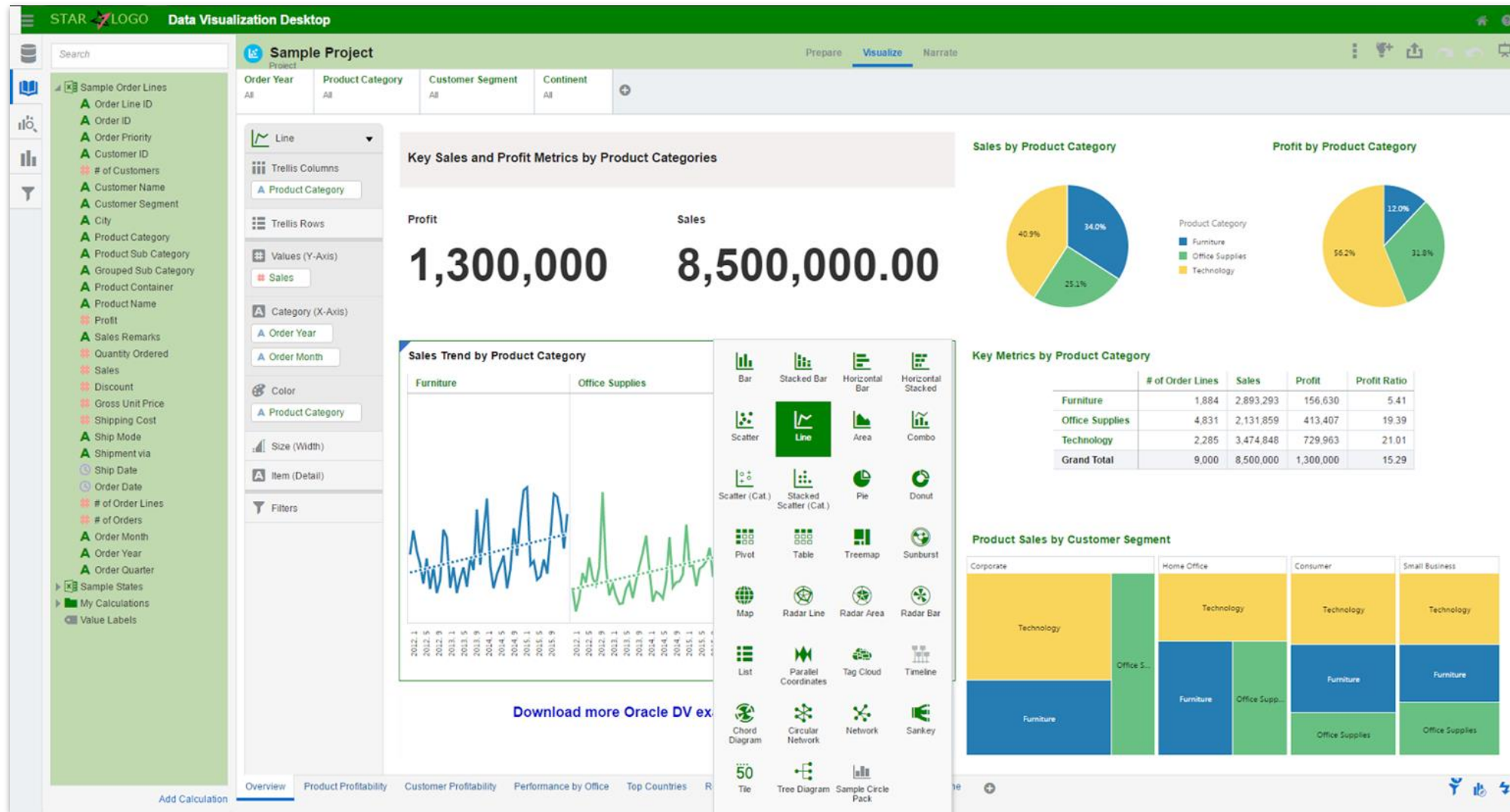
actor_id	first_name	last_name	last_update
18	DAN	TORN	2006-02-15
56	DAN	HARRIS	2006-02-15
116	DAN	STREEP	2006-02-15

1-3 of 3

customer

customer_id	store_id	first_name	last_name	email
477	1	DAN	PAINE	DAN.PAINE@sakilacus

1-1 of 1



Data load

Import archived data into modern database system

Use the full query power of a modern database engine and enable advanced analytics like data mining

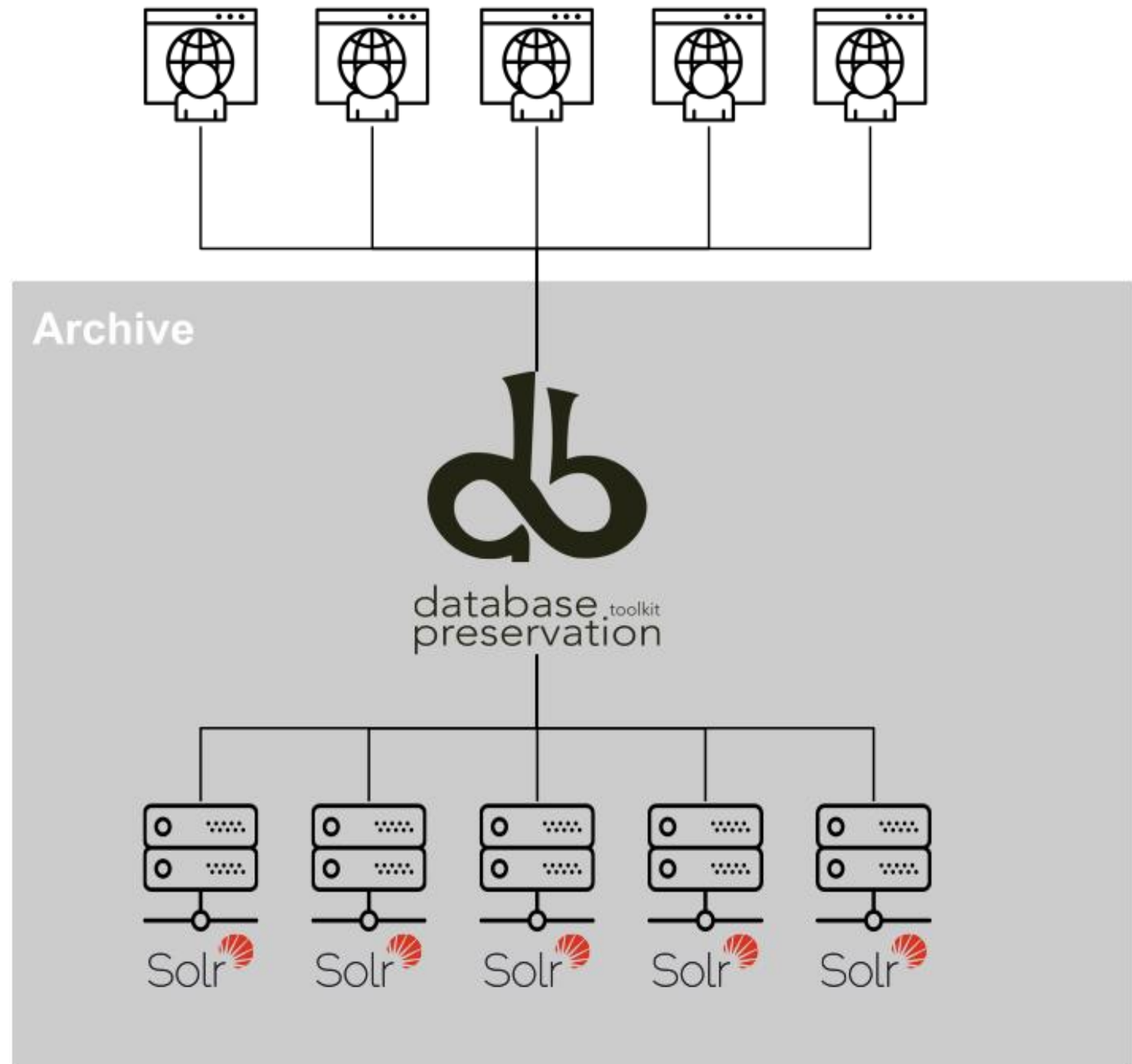


DBPTK Enterprise

DBPTK Enterprise

For large institutions with many databases and users

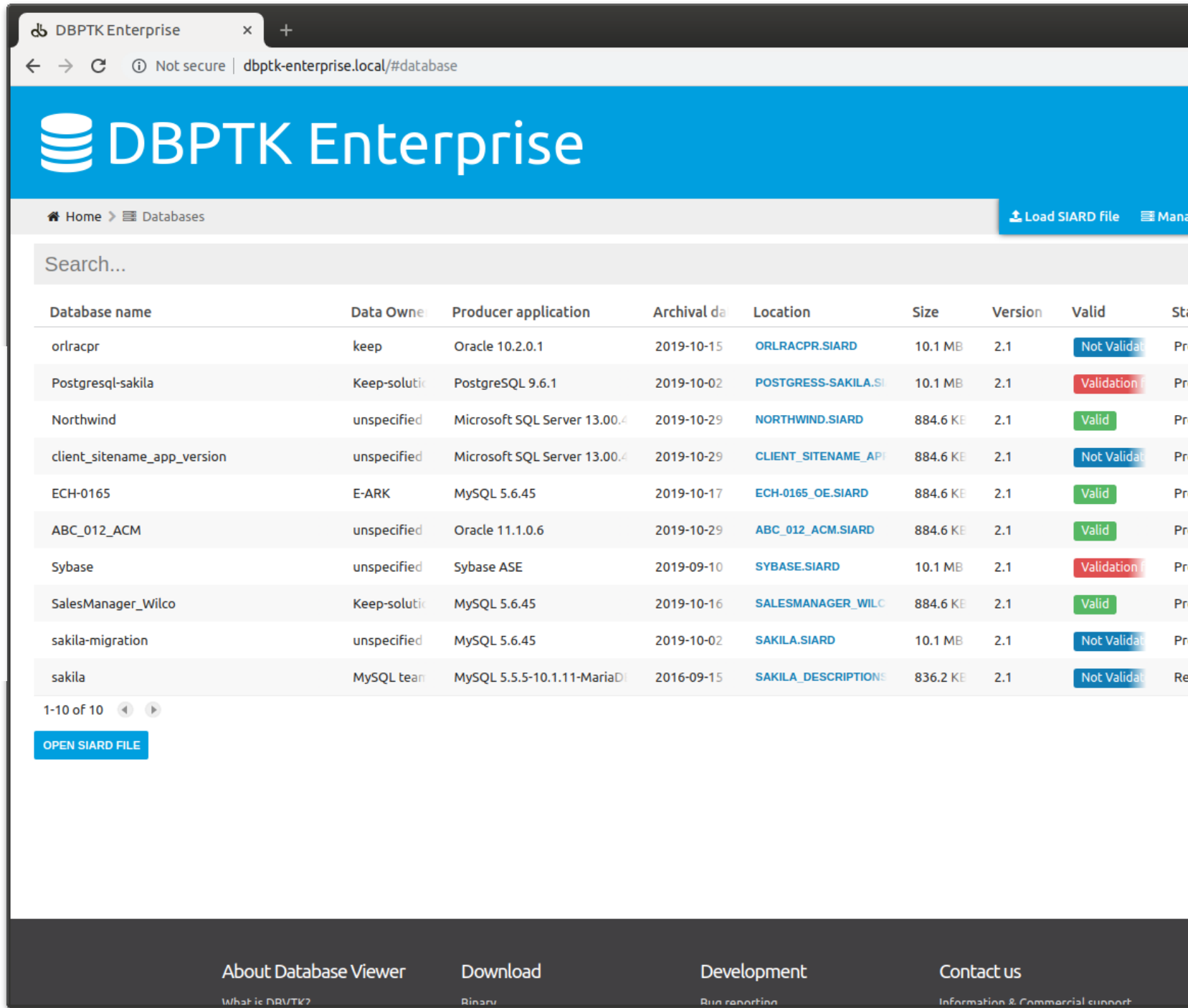
A web application that can be horizontally scaled to support many very large databases being accessed by many users



Manage multiple databases

Single system, multiple databases

Search through the databases, manage their status, enrich their metadata, validate them, make them ready for users to search.



The screenshot shows the DBPTK Enterprise web interface. The browser address bar indicates the URL is `dbptk-enterprise.local/#database`. The page features a blue header with the DBPTK logo and the text "DBPTK Enterprise". Below the header, there is a navigation bar with "Home" and "Databases" links, and a "Load SIARD file" button. A search bar is positioned above a table of databases. The table has columns for Database name, Data Owner, Producer application, Archival date, Location, Size, Version, and Valid status. The status column contains buttons for "Valid", "Not Validated", and "Validation Error".

Database name	Data Owner	Producer application	Archival date	Location	Size	Version	Valid	Status
orlacpr	keep	Oracle 10.2.0.1	2019-10-15	ORLRACPR.SIARD	10.1 MB	2.1	Not Validated	Pr
Postgresql-sakila	Keep-solutio	PostgreSQL 9.6.1	2019-10-02	POSTGRESS-SAKILA.SI	10.1 MB	2.1	Validation Error	Pr
Northwind	unspecified	Microsoft SQL Server 13.00.4	2019-10-29	NORTHWIND.SIARD	884.6 KB	2.1	Valid	Pr
client_sitename_app_version	unspecified	Microsoft SQL Server 13.00.4	2019-10-29	CLIENT_SITENAME_API	884.6 KB	2.1	Not Validated	Pr
ECH-0165	E-ARK	MySQL 5.6.45	2019-10-17	ECH-0165_OE.SIARD	884.6 KB	2.1	Valid	Pr
ABC_012_ACM	unspecified	Oracle 11.1.0.6	2019-10-29	ABC_012_ACM.SIARD	884.6 KB	2.1	Valid	Pr
Sybase	unspecified	Sybase ASE	2019-09-10	SYBASE.SIARD	10.1 MB	2.1	Validation Error	Pr
SalesManager_Wilco	Keep-solutio	MySQL 5.6.45	2019-10-16	SALESMANAGER_WILC	884.6 KB	2.1	Valid	Pr
sakila-migration	unspecified	MySQL 5.6.45	2019-10-02	SAKILA.SIARD	10.1 MB	2.1	Not Validated	Pr
sakila	MySQL team	MySQL 5.5.5-10.1.11-MariaD	2016-09-15	SAKILA_DESCRIPTIONS	836.2 KB	2.1	Not Validated	Re

1-10 of 10

[OPEN SIARD FILE](#)

Footer links: About Database Viewer, Download, Development, Contact us.

Advanced data transformation

Transform content to answer useful questions

De-normalization and table and **column hiding**, to simplify browsing/search and allow **anonymization** of content

The screenshot shows the DBPTK Enterprise web interface. The browser address bar indicates the URL `dbptk-enterprise.local/#database/1547aa41-1800-46b7-a28b-e82fe22f5883`. The interface is divided into several sections:

- Filter sidebar:** Contains an "Information" tab, a search bar for records, saved searches, and a list of tables.
- Tables List:** A scrollable list of tables including: actor, address, category, city, country, customer, film, film_actor, film_category, film_text, inventory, language, payment, rental, staff, store, actor_info, customer_list, film_list, nicer_but_slower_film_list, sales_by_film_category, sales_by_store, and staff_list.
- Schema Description:** A diagram showing relationships between tables. Nodes are represented by circles of varying sizes and colors (blue for primary tables, grey for related tables). Arrows indicate relationships. The "payment" and "rental" tables are the largest and most central nodes.
- Table Description:** A table with two columns: "Table name" and "Description". The first row shows the "payment" table and its description.

```
graph TD
    country((country)) --> city((city))
    city --> address((address))
    address --> customer((customer))
    address --> staff((staff))
    address --> store((store))
    address --> inventory((inventory))
    category((category)) --> film_category((film_category))
    film_category --> film((film))
    film --> actor((actor))
    film --> language((language))
    film --> film_actor((film_actor))
    film --> inventory
    payment((payment)) --> customer
    payment --> staff
    payment --> store
    rental((rental)) --> staff
    rental --> store
    rental --> inventory
```

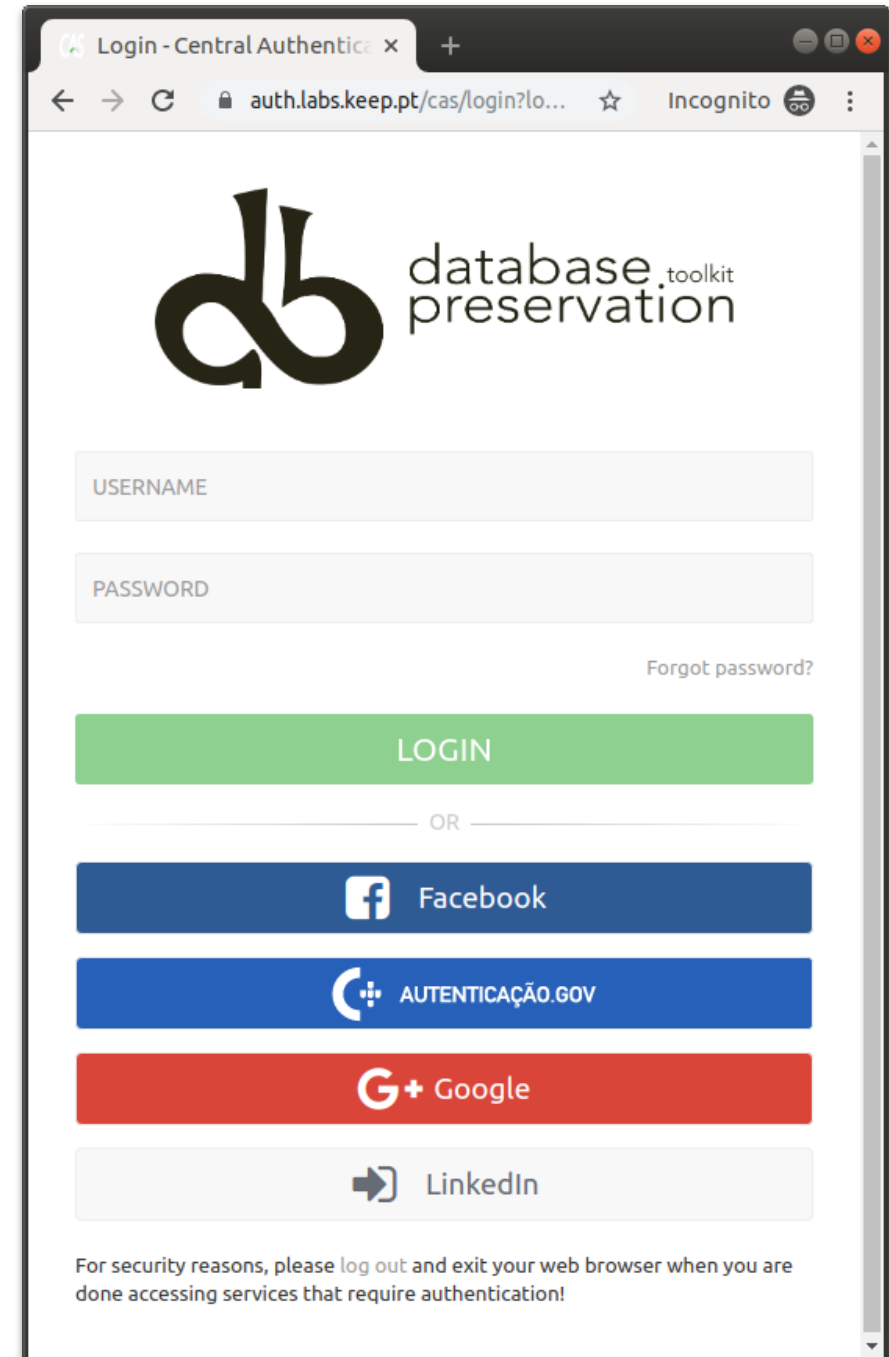
Table name	Description
payment	The payment table records each payment made by a customer, with information such as the amount and the rental being paid for (when applicable). The payment table refers to the customer, rental, and staff tables.

Single sign-on

Support for multiple protocols

LDAP, Active Directory, Database, SAML, ADFS, OAuth2, OpenID, Google, Facebook, Twitter, FIDO U2F, YubiKey, Google Authenticator, Authy, etc.

Supports internal authorization definition or configurable external authorization



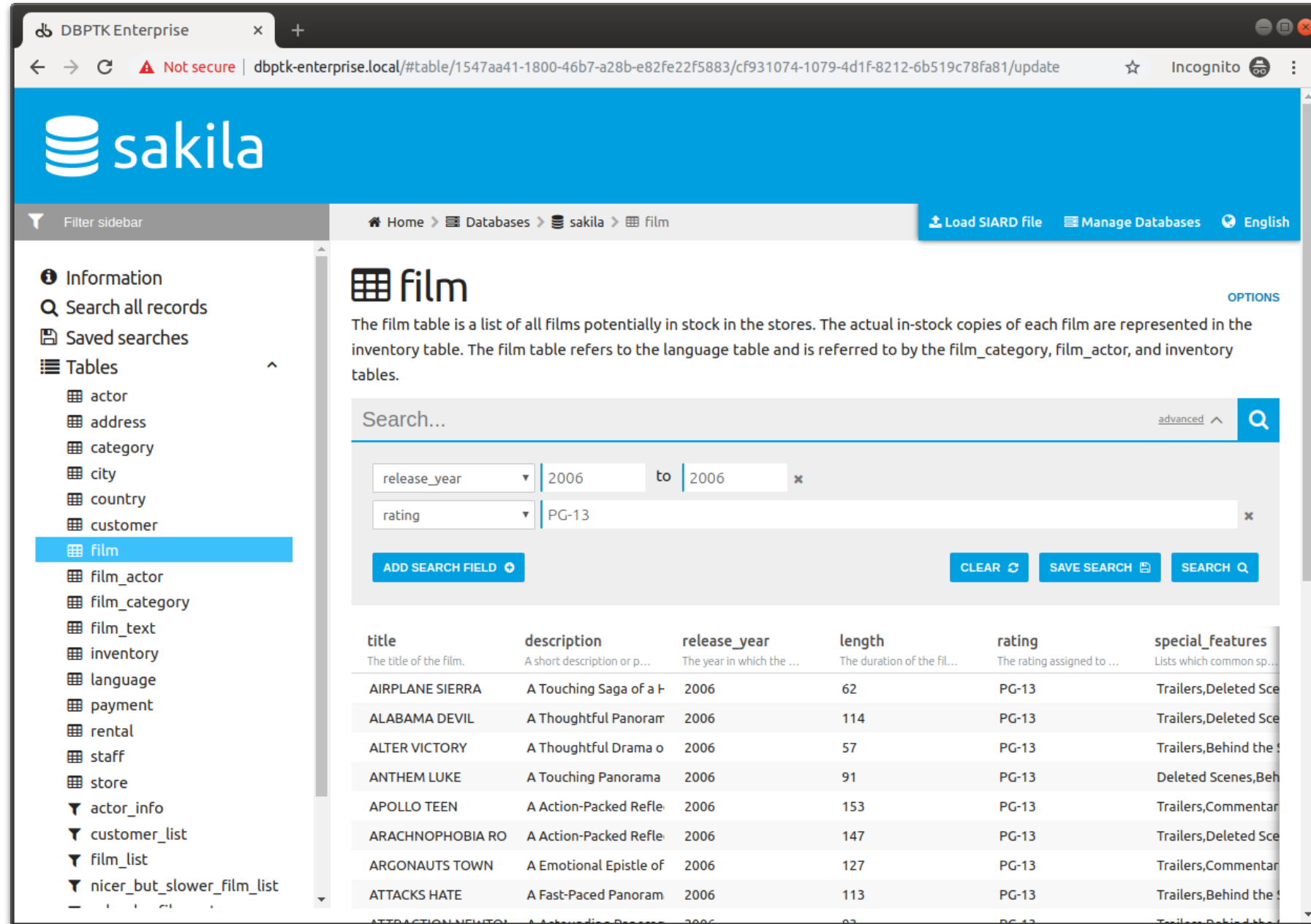
The screenshot shows a web browser window with the following elements:

- Browser title: Login - Central Authentica x
- Address bar: auth.labs.keep.pt/cas/login?lo... Incognito
- Logo: database toolkit preservation
- Form fields: USERNAME, PASSWORD
- Link: Forgot password?
- Button: LOGIN
- Text: OR
- Buttons: Facebook, AUTENTICAÇÃO.GOV, Google, LinkedIn
- Footer: For security reasons, please log out and exit your web browser when you are done accessing services that require authentication!

Browse and search

Allow users to access database content on the Web

Allow them to search on a prepared, de-normalized and anonymized database content.



The screenshot shows the Sakila database web interface. The browser address bar indicates the URL is `dbptk-enterprise.local/#table/1547aa41-1800-46b7-a28b-e82fe22f5883/cf931074-1079-4d1f-8212-6b519c78fa81/update`. The page title is "sakila" and the breadcrumb navigation shows "Home > Databases > sakila > film".

The left sidebar contains a "Filter sidebar" with a list of tables. The "film" table is selected and highlighted in blue. Other tables listed include actor, address, category, city, country, customer, film_actor, film_category, film_text, inventory, language, payment, rental, staff, store, actor_info, customer_list, film_list, and nicer_but_slower_film_list.

The main content area displays the "film" table. It includes a description: "The film table is a list of all films potentially in stock in the stores. The actual in-stock copies of each film are represented in the inventory table. The film table refers to the language table and is referred to by the film_category, film_actor, and inventory tables." Below the description is a search interface with a search bar and filters. The search criteria are: "release_year" from 2006 to 2006, and "rating" PG-13. There are buttons for "ADD SEARCH FIELD", "CLEAR", "SAVE SEARCH", and "SEARCH".

The search results are displayed in a table with the following columns: title, description, release_year, length, rating, and special_features. The table contains 15 rows of film data.

title	description	release_year	length	rating	special_features
AIRPLANE SIERRA	A Touching Saga of a P...	2006	62	PG-13	Trailers,Deleted Sce...
ALABAMA DEVIL	A Thoughtful Panoram	2006	114	PG-13	Trailers,Deleted Sce...
ALTER VICTORY	A Thoughtful Drama o	2006	57	PG-13	Trailers,Behind the s...
ANTHEM LUKE	A Touching Panorama	2006	91	PG-13	Deleted Scenes,Beh...
APOLLO TEEN	A Action-Packed Refle	2006	153	PG-13	Trailers,Commentar...
ARACHNOPHOBIA RO	A Action-Packed Refle	2006	147	PG-13	Trailers,Deleted Sce...
ARGONAUTS TOWN	A Emotional Epistle of	2006	127	PG-13	Trailers,Commentar...
ATTACKS HATE	A Fast-Paced Panoram	2006	113	PG-13	Trailers,Behind the s...
ATTRACTION NEWFO	A Thoughtful Drama o	2006	92	PG-13	Trailers,Behind the s...

Export features

Export data into tabular data

Allow users to save search results in Microsoft Excel or other spreadsheet software format for easy analytics and diagrams

The screenshot shows a Microsoft Excel spreadsheet with a data table and a bar chart. The data table has columns for title, description, release_year, length, rating, and special_features. The bar chart, titled "The duration of the film", shows the length of each film in minutes. The chart is a horizontal bar chart with the x-axis representing length in minutes (0 to 200) and the y-axis listing the film titles. The bars are blue and extend to the right, indicating the duration of each film. The spreadsheet interface includes the ribbon (Home, Insert, Page Layout, Formulas, Data, Review, View, Developer) and the status bar at the bottom.

	A	B	C	D	E	
1	title	description	release_year	length	rating	special_features
2	The title of the film.	A short description or plot summary of the film.	The year in which the movie was released.	The duration of the film	The rating as	Lists which common sp
3	GOLDFINGER SENSIBILITY	A Insightful Drama of a Mad Scientist And a Hunter who must Defeat a Pastry Chef in New Orleans		2006	93 G	Trailers,Commentaries,
4	WOLVES DESIRE	A Fast-Paced Drama of a Squirrel And a Robot who must Succumb a Technical Writer in A Manhat		2006	55 NC-17	Behind the Scenes
5	CREEPERS KANE	A Awe-Inspiring Reflection of a Squirrel And a Boat who must Outrace a Car in A Jet Boat		2006	172 NC-17	Trailers,Behind the Sce
6	GUNFIGHT MOON	A Epic Reflection of a Pastry Chef And a Explorer who must Reach a Dentist in The Sahara				d Scenes,Behind
7	TRIP NEWTON	A Fanciful Character Study of a Lumberjack And a Car who must Discover a Cat in An Abar				entaries,Deleted
8	PERDITION FARGO	A Fast-Paced Story of a Car And a Cat who must Outgun a Hunter in Berlin				s,Behind the Sce
9	INDEPENDENCE HOTEL	A Thrilling Tale of a Technical Writer And a Boy who must Face a Pioneer in A Monastery				entaries,Deleted
10	ROSES TREASURE	A Astounding Panorama of a Monkey And a Secret Agent who must Defeat a Woman in T				entaries,Deleted
11	KENTUCKIAN GIANT	A Stunning Yarn of a Woman And a Frisbee who must Escape a Waitress in A U-Boat				s,Commentaries,
12	PUNK DIVORCE	A Fast-Paced Tale of a Pastry Chef And a Boat who must Face a Frisbee in The Canadian I				s,Commentaries,
13	KNOCK WARLOCK	A Unbelievable Story of a Teacher And a Boat who must Confront a Moose in A Baloon				s,
14	UPTOWN YOUNG	A Fateful Documentary of a Dog And a Hunter who must Pursue a Teacher in An Abandon				entaries
15	MAGUIRE APACHE	A Fast-Paced Reflection of a Waitress And a Hunter who must Defeat a Forensic Psycholc				s,Commentaries
16	WYOMING STORM	A Awe-Inspiring Panorama of a Robot And a Boat who must Overcome a Feminist in A U-				d Scenes
17	CENTER DINOSAUR	A Beautiful Character Study of a Sumo Wrestler And a Dentist who must Find a Dog in Ca				d Scenes
18	DIVIDE MONSTER	A Intrepid Saga of a Man And a Forensic Psychologist who must Reach a Squirrel in A Mo				s,Commentaries
19	SPIRIT FLINTSTONES	A Brilliant Yarn of a Cat And a Car who must Confront a Explorer in Ancient Japan				entaries,Deleted
20	INTOLERABLE INTENTIONS	A Awe-Inspiring Story of a Monkey And a Pastry Chef who must Succumb a Womanizer in				entaries,Behind
21	HOOK CHARIOTS	A Insightful Story of a Boy And a Dog who must Redeem a Boy in Australia				s,Commentaries,
22	ENCINO ELF	A Astounding Drama of a Feminist And a Teacher who must Confront a Husband in A Balc				s,Behind the Sce
23	CURTAIN VIDEOTAPE	A Boring Reflection of a Dentist And a Mad Cow who must Chase a Secret Agent in A Sha				s,Commentaries,
24	LAMBS CINCINNATI	A Insightful Story of a Man And a Feminist who must Fight a Composer in Australia				s,Behind the Sce
25	MAGNOLIA FORRESTER	A Thoughtful Documentary of a Composer And a Explorer who must Conquer a Dentist in				s,Commentaries,
26	BACKLASH UNDEFEATED	A Stunning Character Study of a Mad Scientist And a Mad Cow who must Kill a Car in A M				s,Behind the Sce
27	CLEOPATRA DEVIL	A Fanciful Documentary of a Crocodile And a Technical Writer who must Fight a A Shark i				s,Deleted Scenes
28	HOCUS FRIDA	A Awe-Inspiring Tale of a Girl And a Madman who must Outgun a Student in A Shark Tan				s,Deleted Scenes
29	STAGE WORLD	A Lacklustre Panorama of a Woman And a Frisbee who must Chase a Crocodile in A Jet I				entaries,Behind
30	CHAINSAW UPTOWN	A Beautiful Documentary of a Boy And a Explorer who must Discover a Squirrel in Australia				d Scenes,Behind
31	PILOT HOOSIERS	A Awe-Inspiring Reflection of a Crocodile And a Sumo Wrestler who must Meet a Forensi				s,Deleted Scenes
32	REMEMBER DIARY	A Insightful Tale of a Technical Writer And a Waitress who must Conquer a Monkey in Ancient Indi		2006	110 R	Trailers,Commentaries
33	JAPANESE RUN	A Awe-Inspiring Epistle of a Feminist And a Girl who must Sink a Girl in The Outback		2006	135 G	Deleted Scenes
34	RAINBOW SHOCK	A Action-Packed Story of a Hunter And a Boy who must Discover a Lumberjack in Ancient India		2006	74 PG	Trailers,Commentaries,
35	MAIDEN HOME	A Lacklustre Saga of a Moose And a Teacher who must Kill a Forensic Psychologist in A MySQL Cc		2006	138 PG	Behind the Scenes




DBPTK Developer

Command line interface

Automation of periodic preservation tasks

Command line interface allows easy automation of periodic tasks like saving database to preservation format, validating, and editing metadata.



```
2/2 + [ ] [ ] Tmux: Default [ ] [ ] [ ] [ ] [ ] [ ]
~ $ java -jar dbptk-app-2.6.3.jar
Database Preservation ToolkitDatabase Preservation Toolkit (version 2.6.3)
More info: http://www.database-preservation.com

Usage: dbptk COMMAND [OPTIONS]

Commands:

    migrate      Migrates data and metadata from an import module to an export module.
    edit         Edit the metadata information from a SIARD 2 archive.
    validate     Validate a SIARD 2 archive.

Run 'dbptk -h|help COMMAND' for more information on a command.

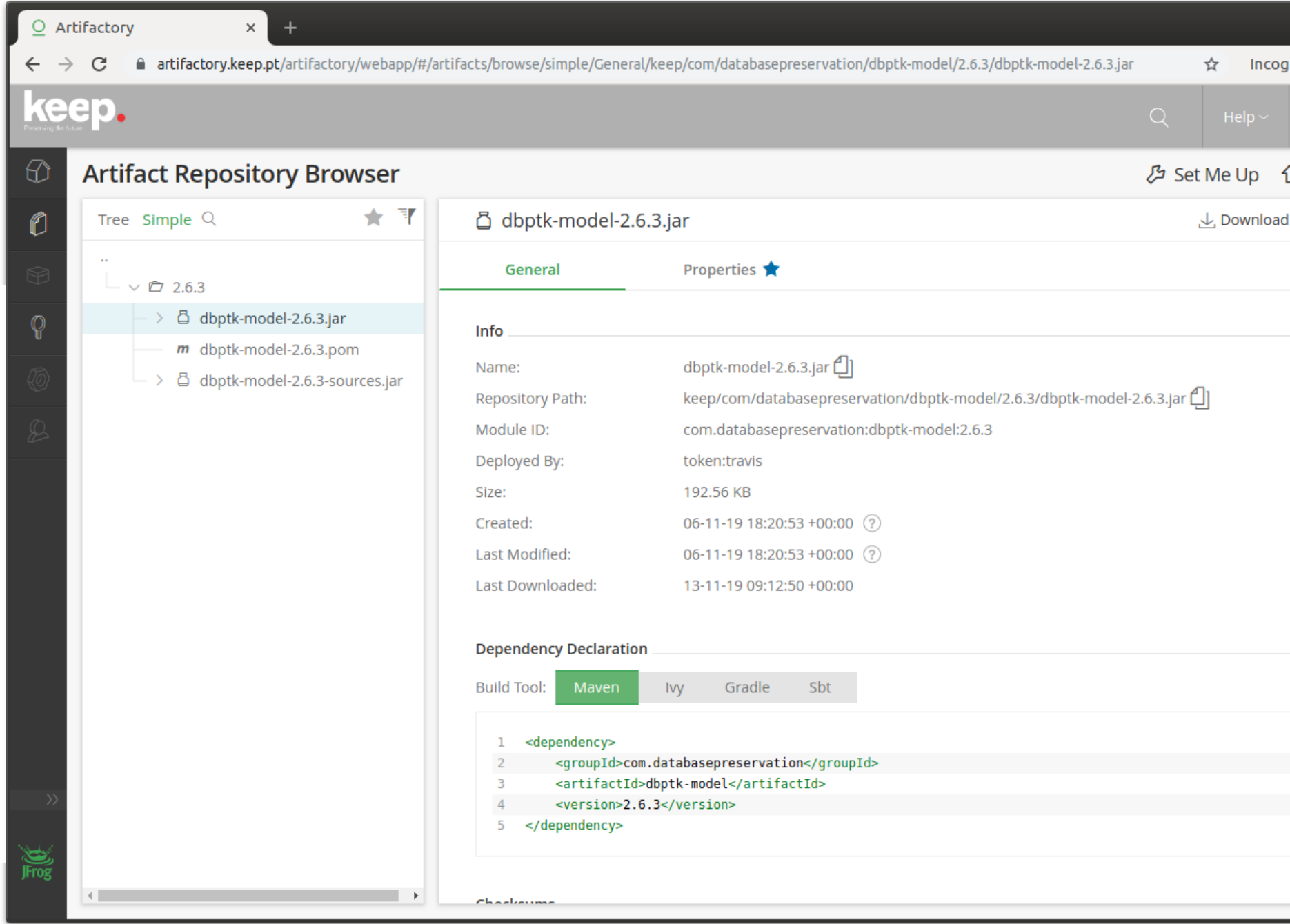
    Log files and migration reports were saved in /home/mguimaraes
    Troubleshooting information can be found at http://www.database-preservation.com/#troubleshooting
    Please report any problems at https://github.com/keeps/db-preservation-toolkit/issues/new

~ $ _
```

Systems integration

Java library

Library to allow integration of production systems to directly use database preservation features.



Open source

For custom development

Code base that allows custom development of new features or specialized support for new or legacy database systems.

The screenshot shows the GitHub repository page for 'db-preservation-toolkit'. At the top, the navigation bar includes 'Why GitHub?', 'Enterprise', 'Explore', 'Marketplace', 'Pricing', a search bar, and 'Sign in' / 'Sign up' buttons. The repository name 'keeps / db-preservation-toolkit' is displayed, along with 'Watch 14', 'Star 27', and 'Fork 10' buttons. Below this, there are tabs for 'Code', 'Issues 61', 'Pull requests 0', 'Projects 1', 'Wiki', 'Security', and 'Insights'. The repository description is 'Database Preservation Toolkit' with a link to 'http://www.database-preservation.com'. There are also tags for 'preservation', 'database', 'relational-databases', 'siard', and 'preservation-formats'. A progress bar shows '1,175 commits', '8 branches', '0 packages', '45 releases', and '11 contributors'. Below the progress bar, there are buttons for 'Branch: master', 'New pull request', 'Find file', and 'Clone or download'. The main content area shows a list of files and folders with their commit history. The most recent commit is by 'hmiguim' with the message 'Change java version to 8 in order to use Instant class for better dat...' and a timestamp of '12 days ago'. The file list includes folders like '.github', '.travis', 'code-style', 'dbptk-bindings', 'dbptk-core', 'dbptk-model', 'dbptk-modules', 'dbptk-plugin-example', 'doc', 'scripts', 'testing' and files like '.gitattributes', '.gitignore', '.grenrc.yml', '.travis.yml', 'CHANGELOG.md', 'DEV_NOTES.md', 'LICENSE', 'LICENSE_HEADER.txt', 'README.md', and 'pom.xml'. At the bottom, there is a section for 'README.md' with the heading 'Database Preservation'.

How to choose?



database_{toolkit}
preservation



DESKTOP



Producers



Archivists



database_{toolkit}
preservation



ENTERPRISE



Producers



Admin



Archivists



Applications



database_{toolkit}
preservation



DEVELOPER



Developers



Applications



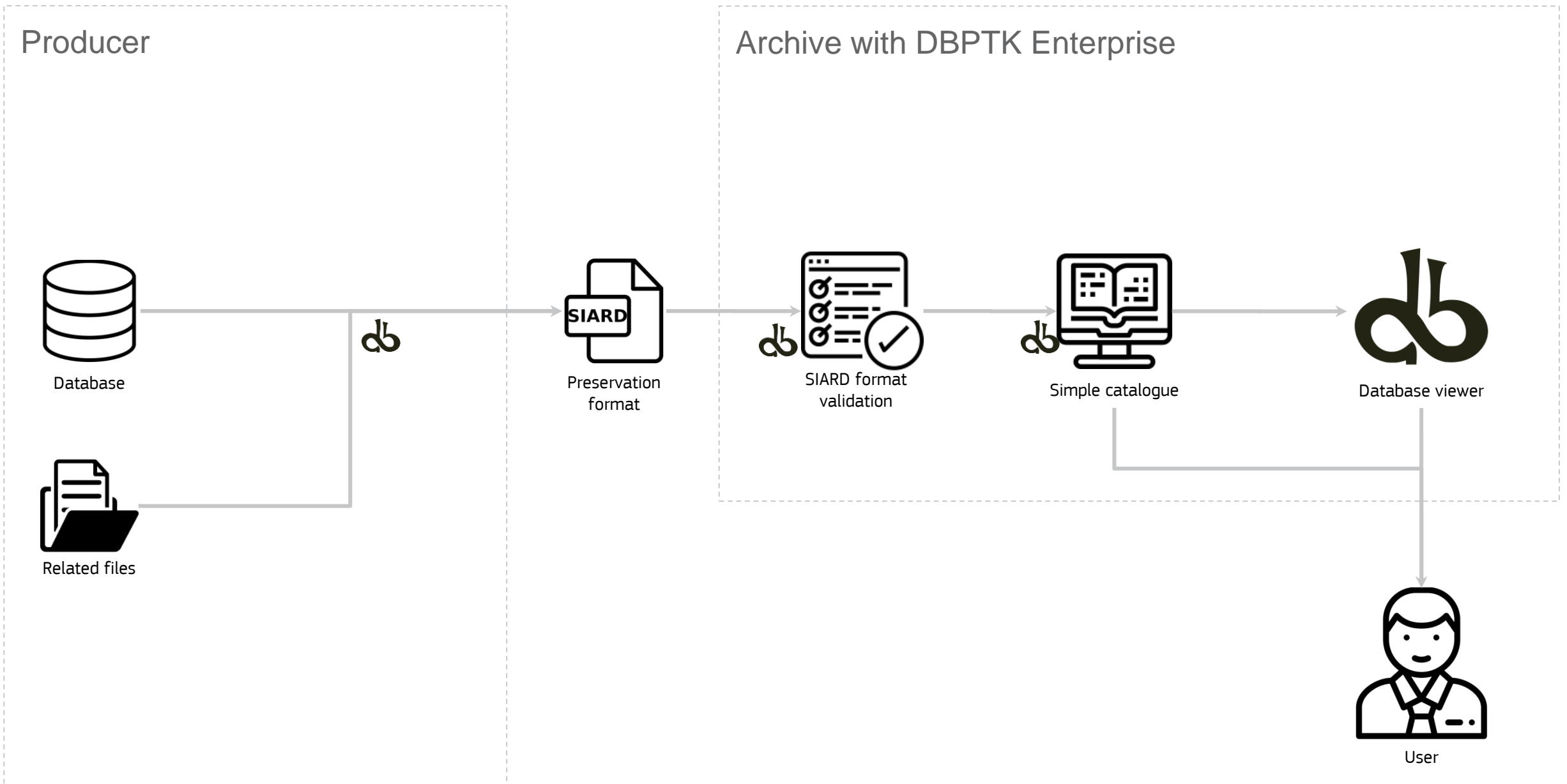
Producers

	Desktop	Enterprise	Developer
Save to preservation format	✓	✓*	✓
Validation	✓	✓	✓
Enrich descriptions	✓	✓	✓
Browse and search	✓	✓	X
Transform (de-normalization)	✓**	✓	X
Export to live databases	✓	✓*	✓
Authentication	X	✓	X
Number of users	one	many	one
Number of loaded databases	few	many	N/A
Integration with repositories	X	✓	N/A
Embeddable in Web portals	X	✓	N/A

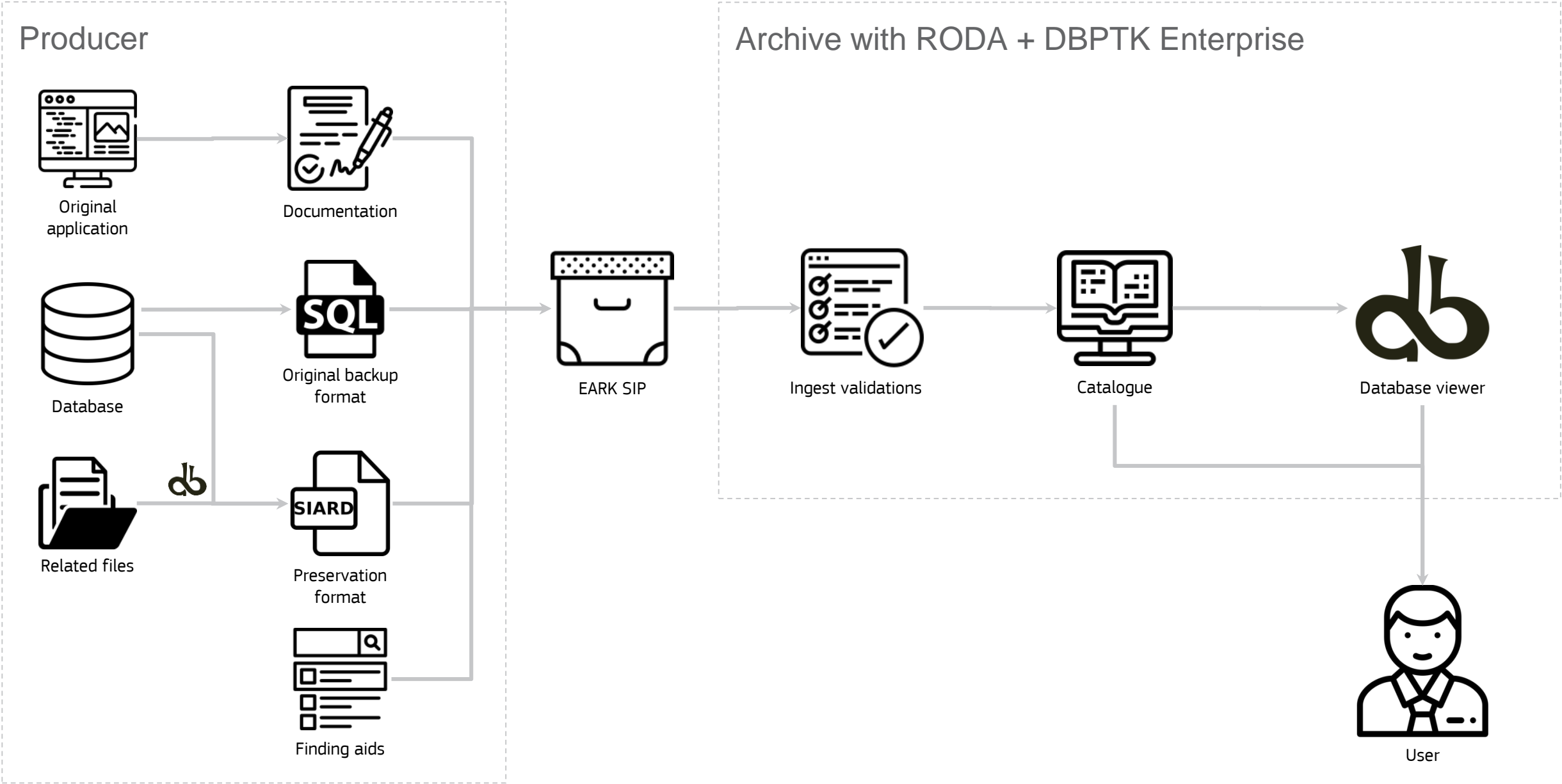
* Enterprise feature done via the upload/download of SIARD and usage of related tools

** Limited functionally, may need very high requirements for a Desktop

Simple database archival flow



Full database archival ecosystem flow



Download at

<https://www.database-preservation.com>

Thank you for your attention.

Any questions?

Luis Faria

lfaria@keep.pt

www.database-preservation.com

Contact us



cef-building-blocks@ec.europa.eu

© European Union, 2019. All rights reserved. Certain parts are licensed under conditions to the EU.

Reproduction is authorized provided the source is acknowledged.

Coffee break

We will resume at **15:15**

