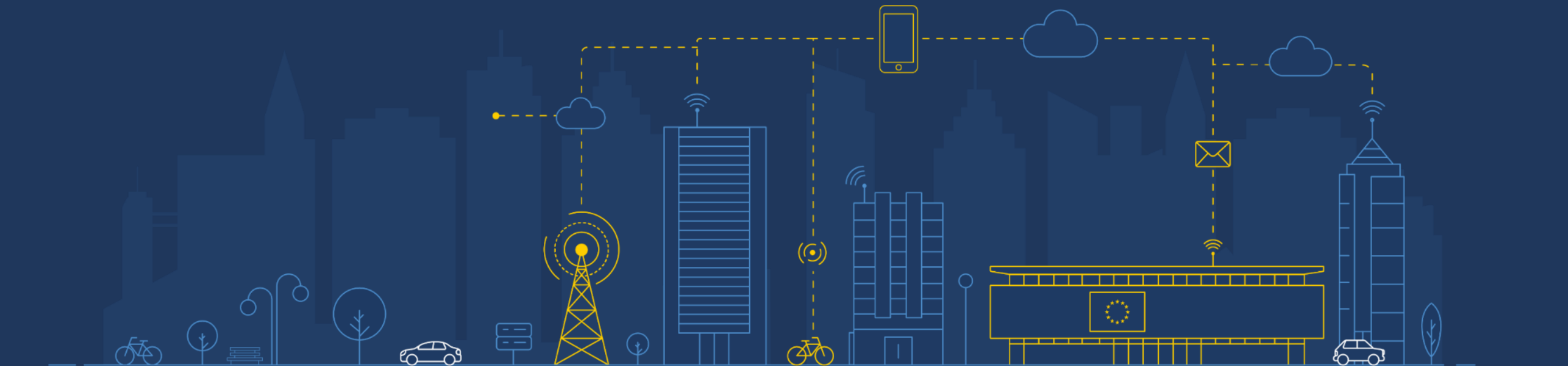
Agenda Day 1 - afternoon

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**	

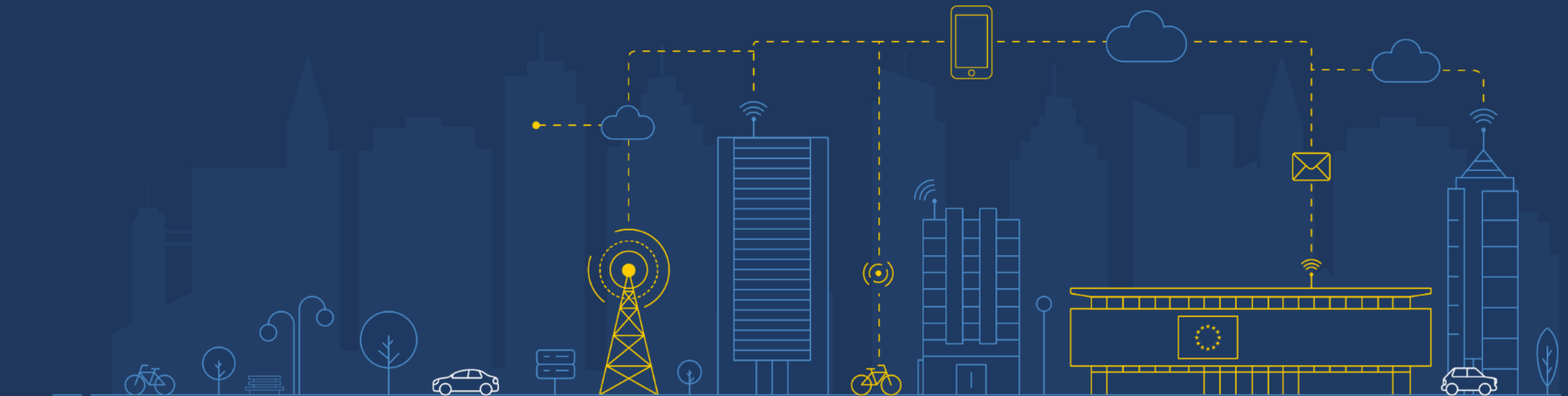
Software presentation



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

European Commission

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

MIME

EAD

Migration

WARC

JPEG

TIFF

OAIS

OPF

METS

RODA

AIP

DC

ISO

SIP

JHove

PREMIS

Metadata

NDSA

Refreshing

Authentication

PAIMAS

PDF/A

MPEG

DCC

Digitisation

DPC

Formats

DRAMBORA

Authenticity

DDI

XML

PRONOM

DIP

SGML

Checksum

CCSDS

DROID

HTML

DOI

Emulation

TRAC

European Commission

End-to-end eArchiving solutions for digital preservation

European Commission

20180912_eArchiving_Flyer.pdf (página 1 de 2)

Pesquisa

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

European Commission

20180912_eArchiving_Flyer.pdf (página 1 de 2)

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 eAr
informa
referenc
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

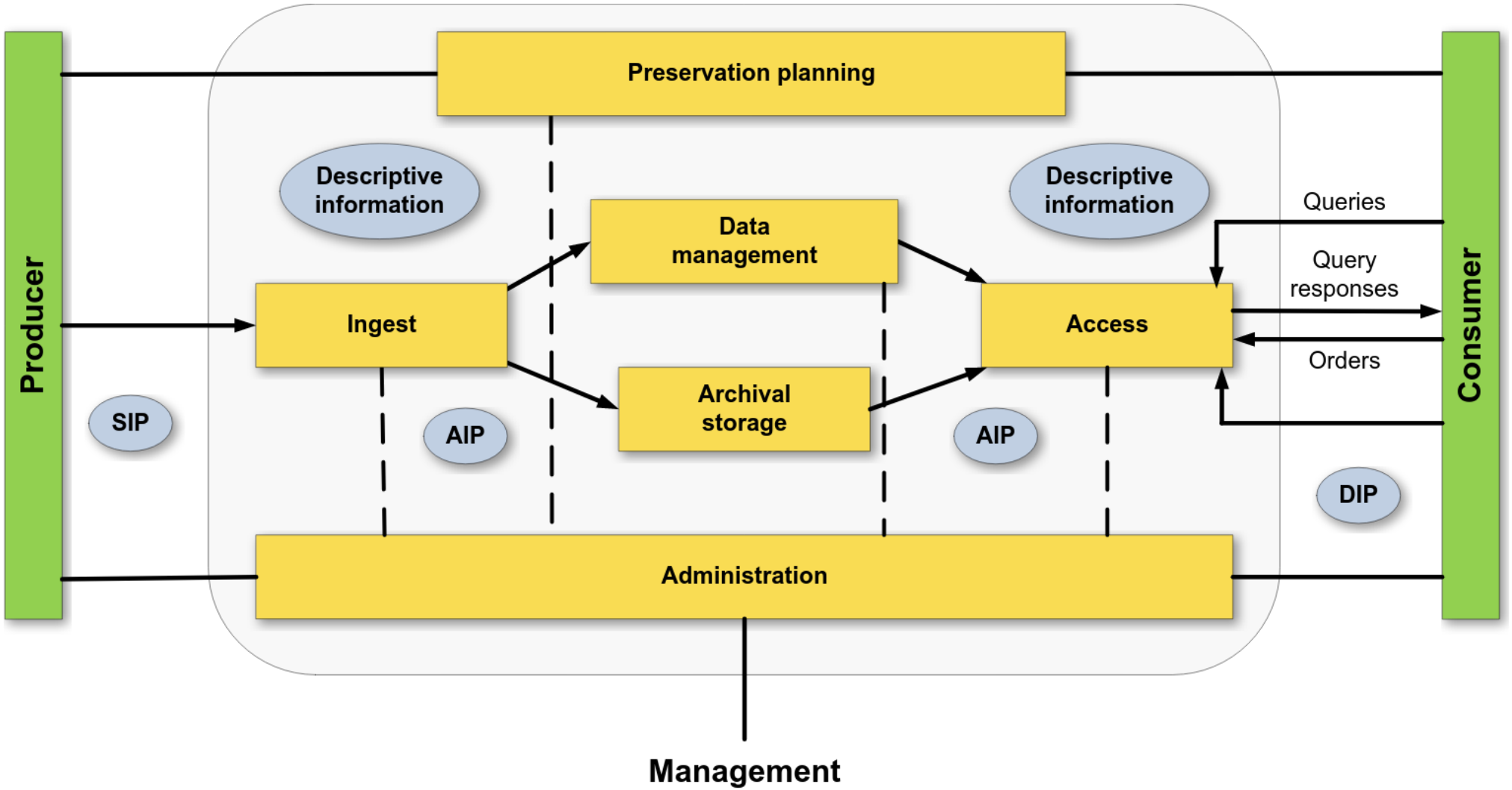
Package specifications that provide interoperability across borders, types of institution and user communities. These are supported by the Common Specification for Information Packages (CS IP), expressed with the Metadata Encoding & Transmission Standard (METS), with specialisations for:

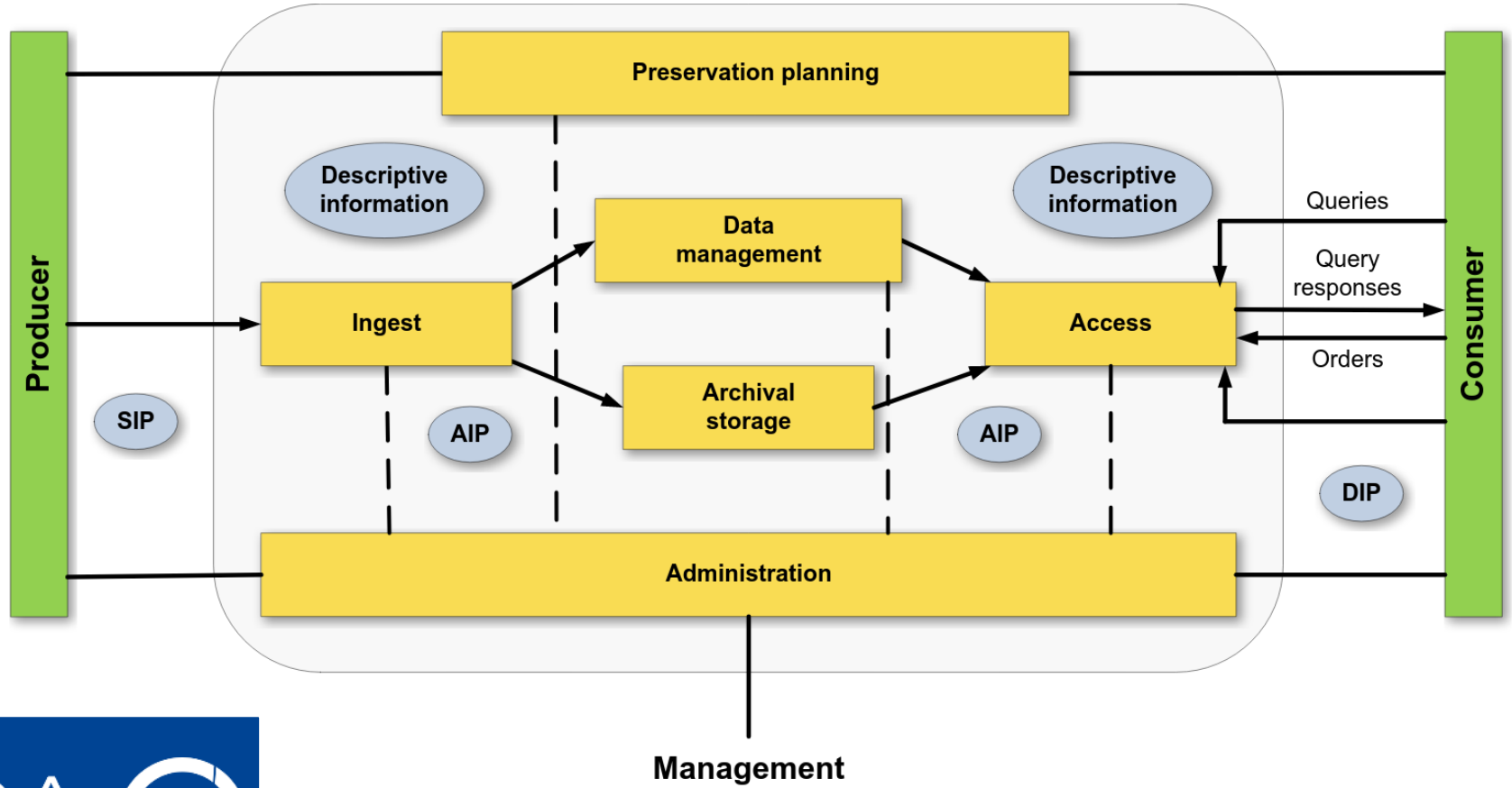
- Submission Information Package

European Commission

End-to-end eArchiving solutions for digital preservation

European Commission


Preservation planning
Descriptive information
Descriptive information
Data management
Producer
Ingest
Access
Consumer
Queries
Query responses
Orders
SIP
AIP
Archival storage
AIP
DIP
Administration
Management



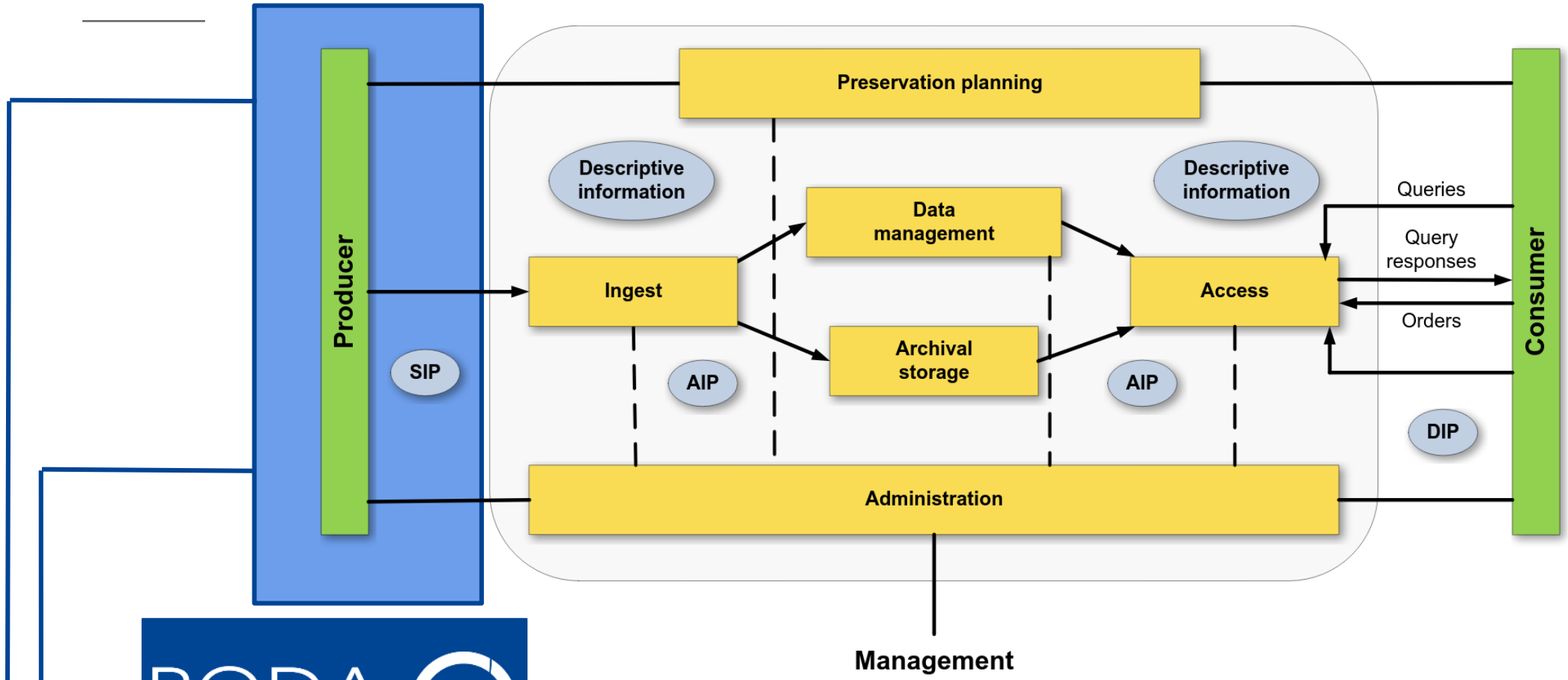
RODA SIP creation tool

Preservation Toolkit

RODA

Preservation Toolkit

European Commission



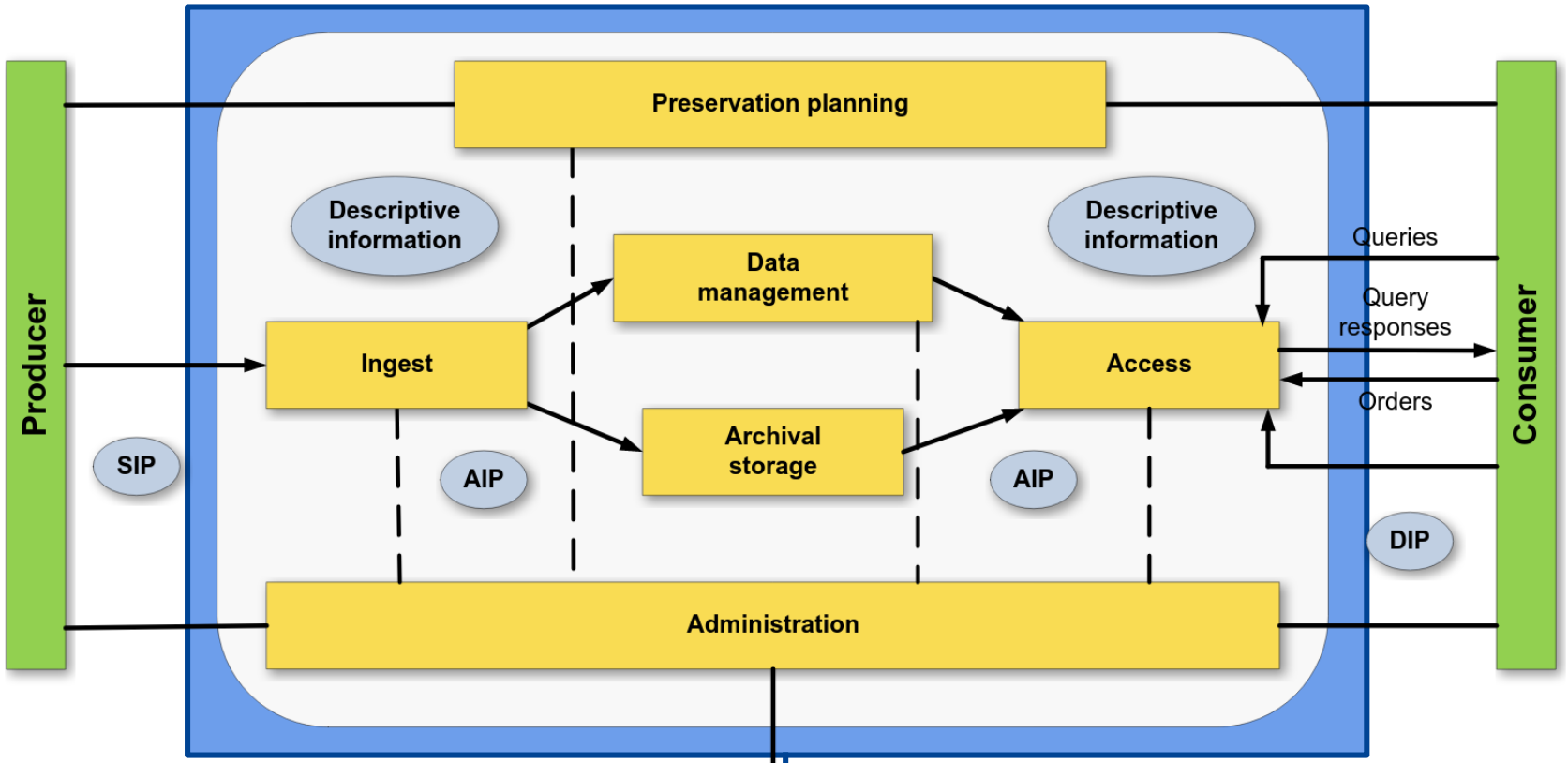
RODA
SIP creation tool

Preservation Toolkit

RODA

Preservation Toolkit

European Commission

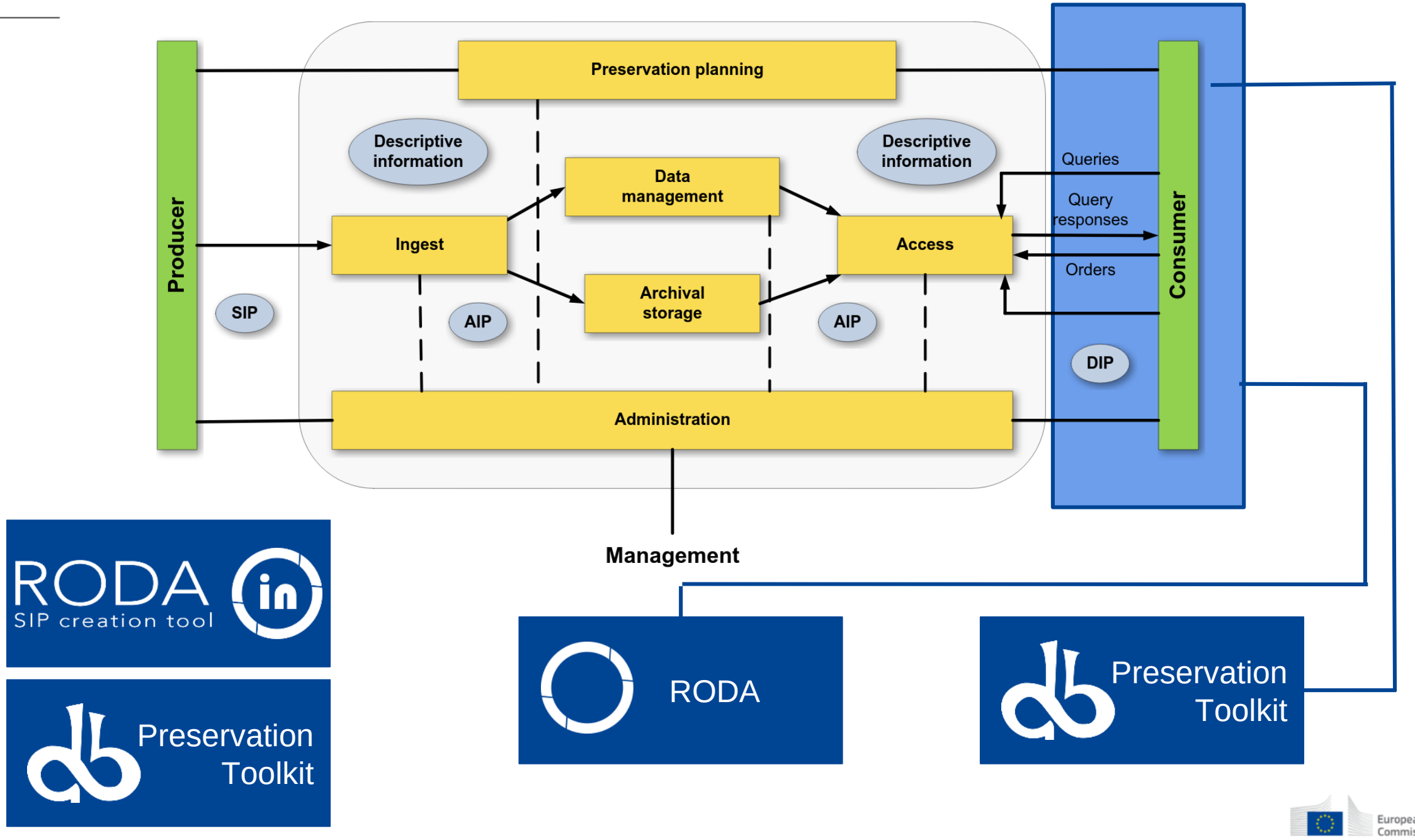


RODA
SIP creation tool

Preservation Toolkit

RODA

Preservation Toolkit


Preservation planning
Descriptive information
Descriptive information
Data management
Producer
Ingest
Access
Consumer
Queries
Query responses
Orders
SIP
AIP
Archival storage
AIP
DIP
Administration
Management
RODA
SIP creation tool
Preservation Toolkit
RODA
Preservation Toolkit

eArchiving tools

RODA

SIP creation tool

European Commission

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

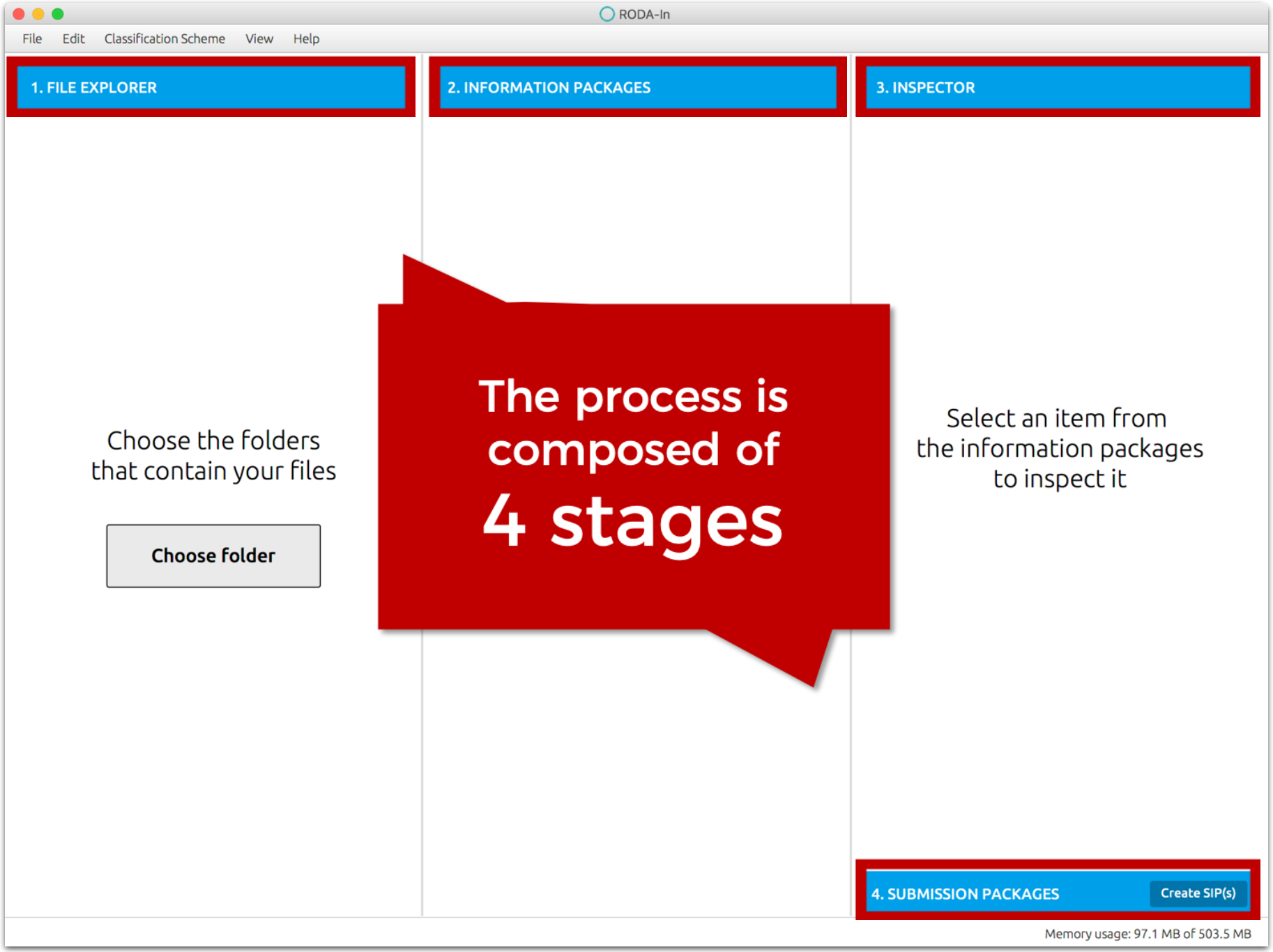
RODA SIP creation tool

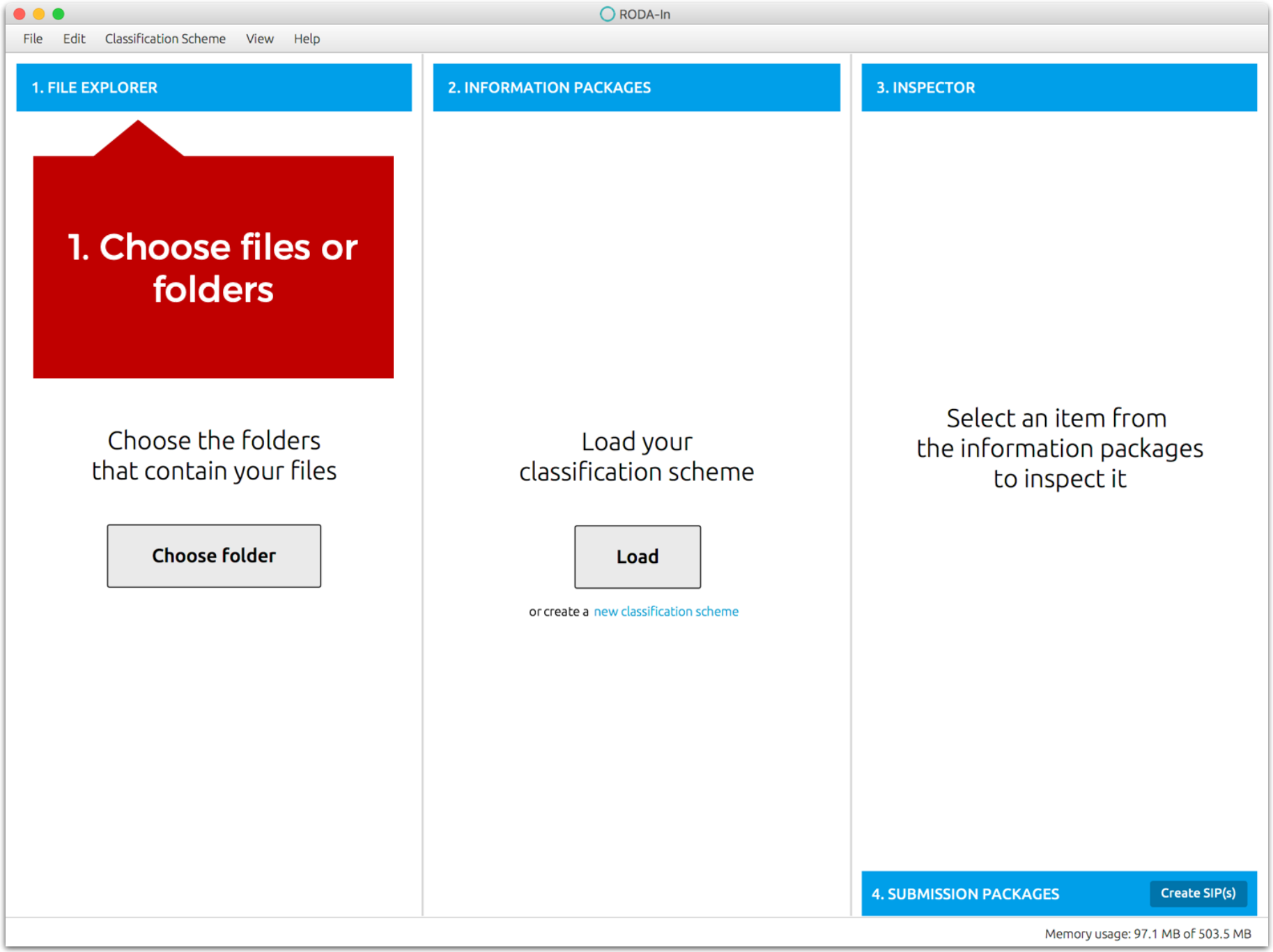
Packaging

Transfer & ingest

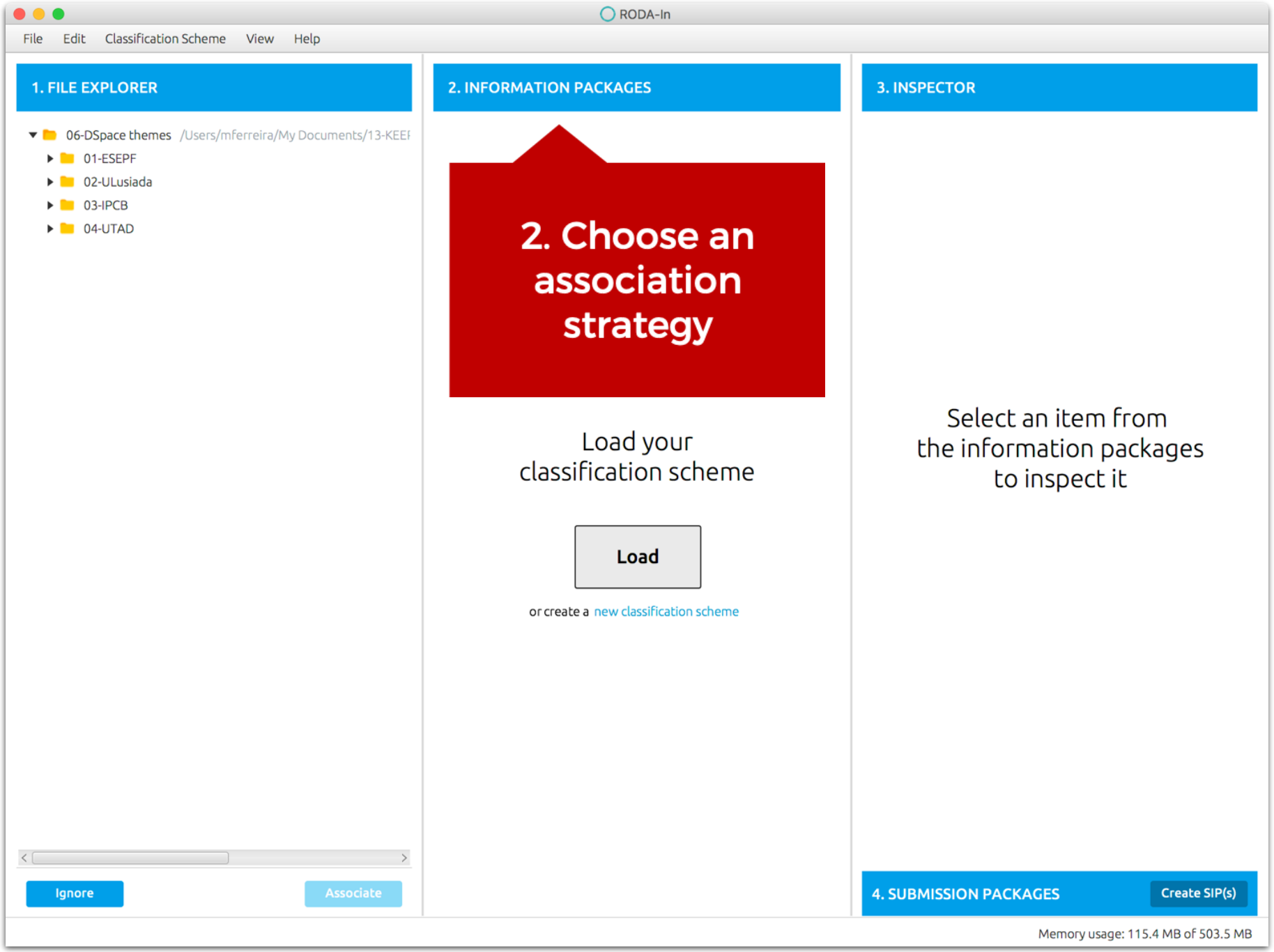
EARK compatible
Digital Repository

How does it work?

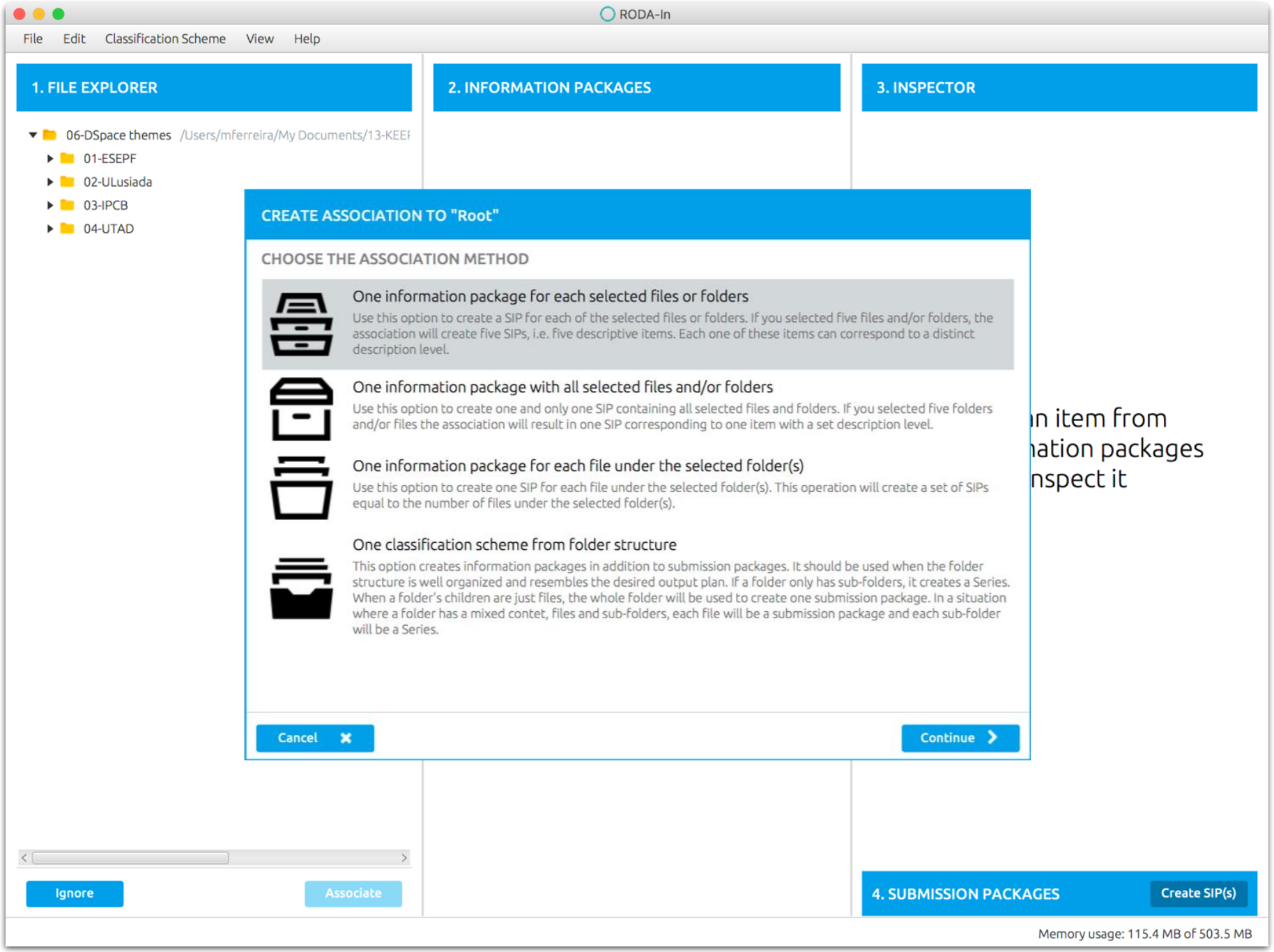




European Commission

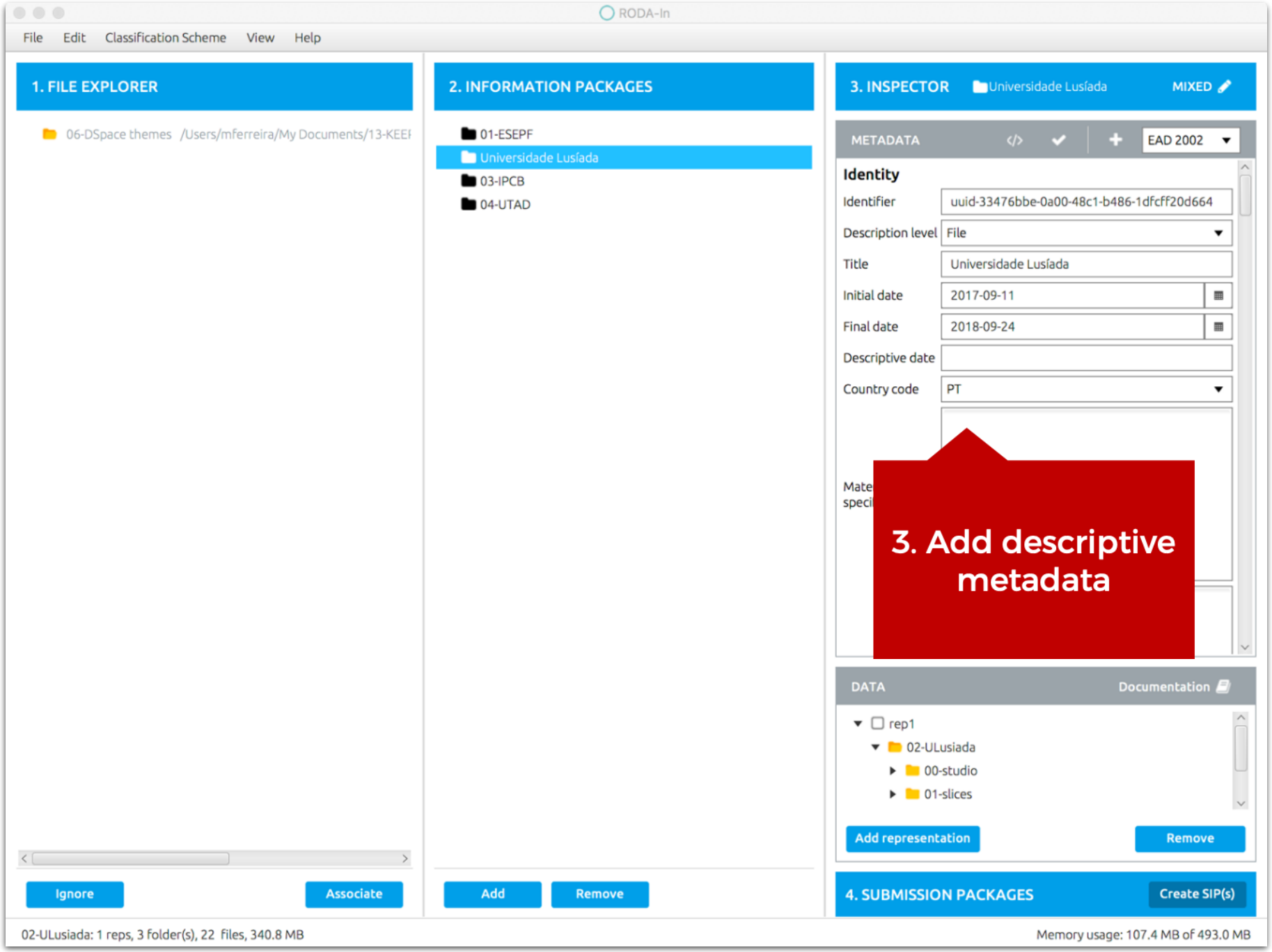


European Commission

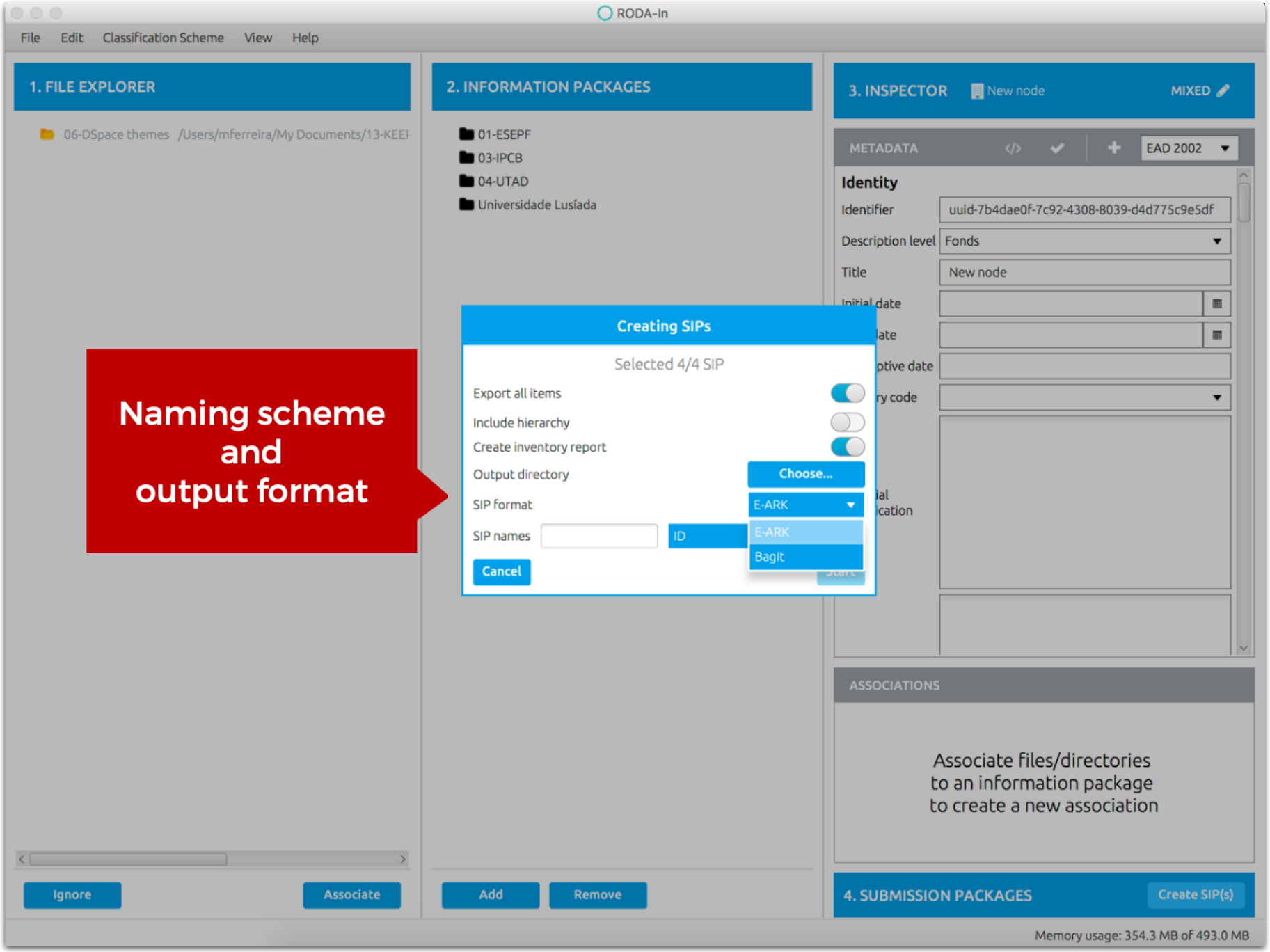


n item from
nation packages
nspect it

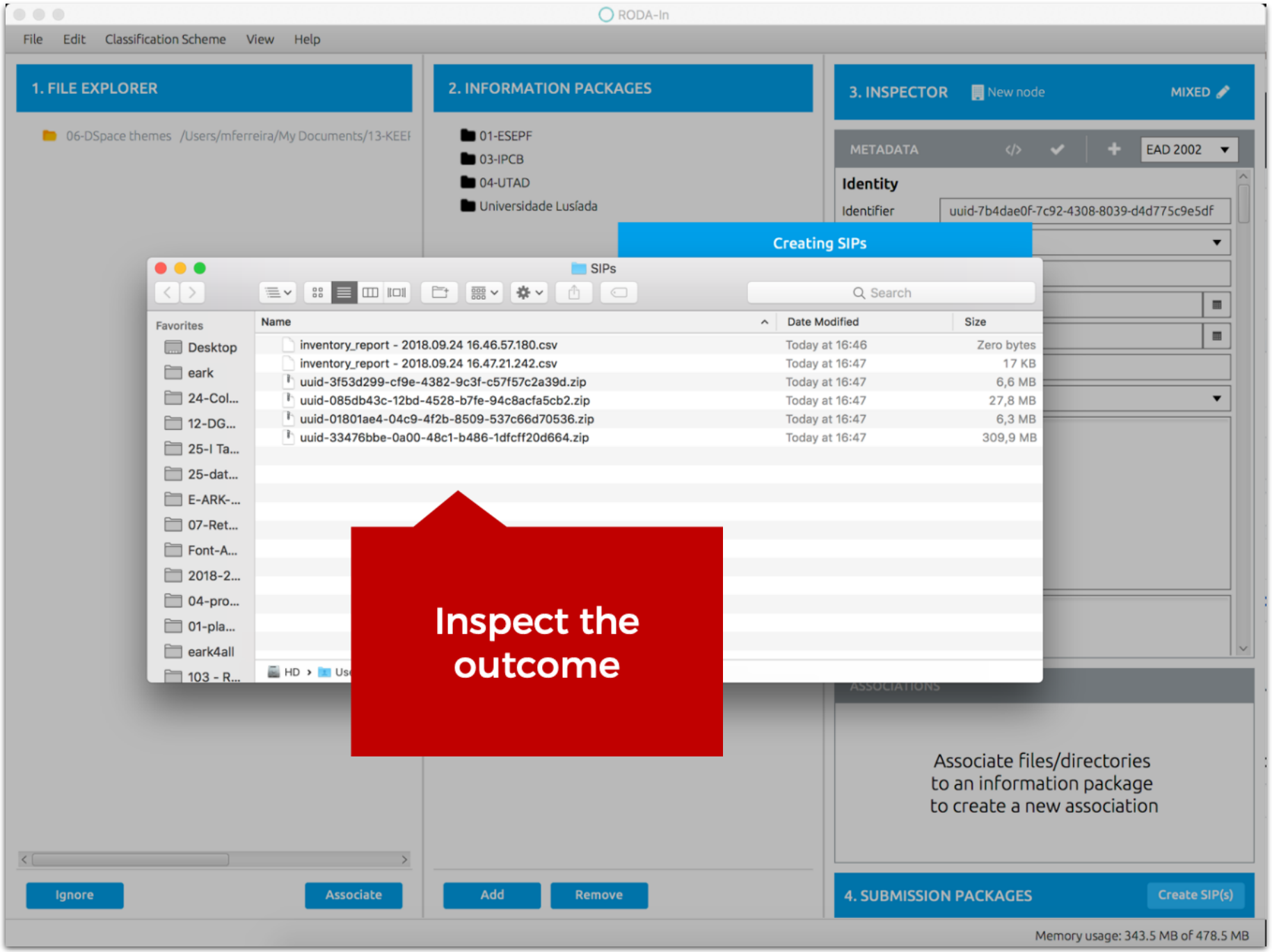
European Commission



European Commission

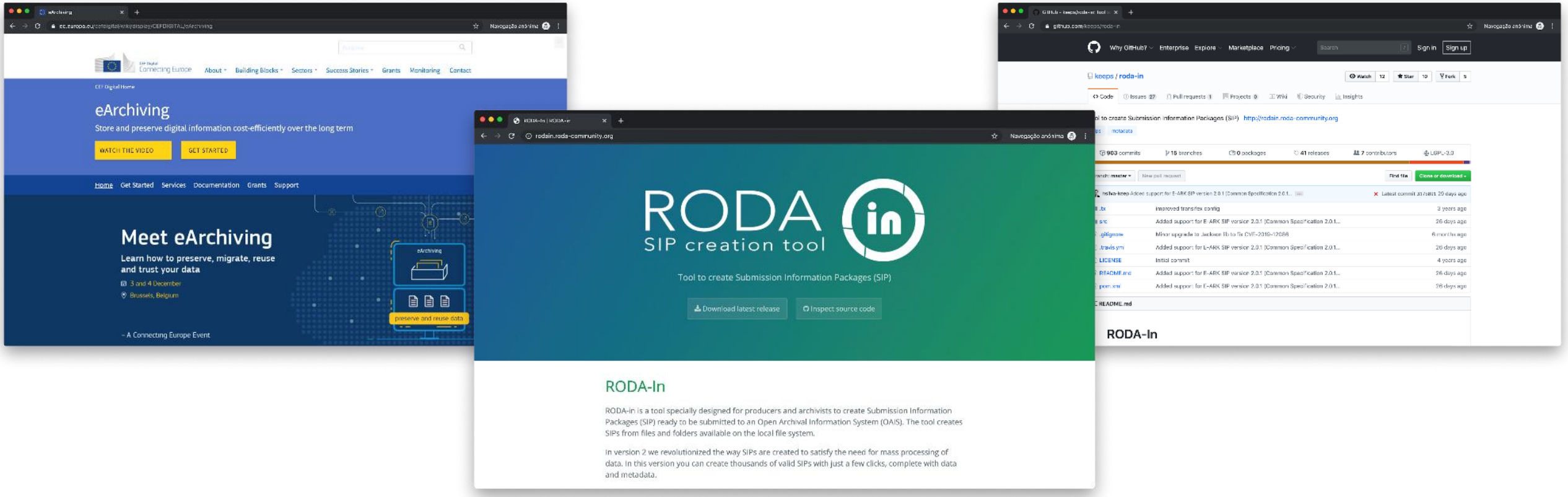


European Commission



European Commission

Where can I find it?



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

rodain.roda-community.org

github.com/keeps/roda-in

European Commission

RODA

European Commission

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

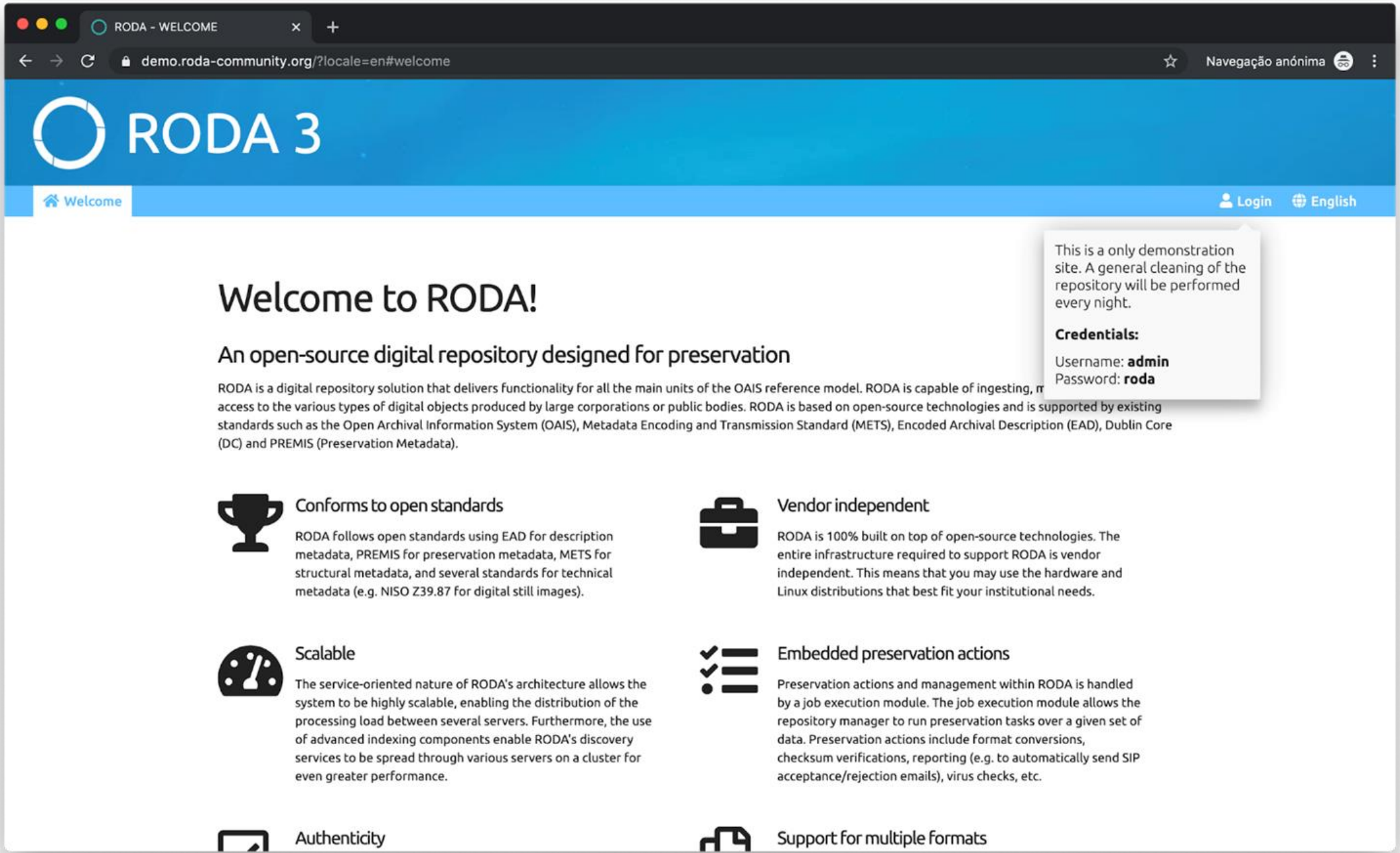
Auditable

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

Integrations with 3rd party systems

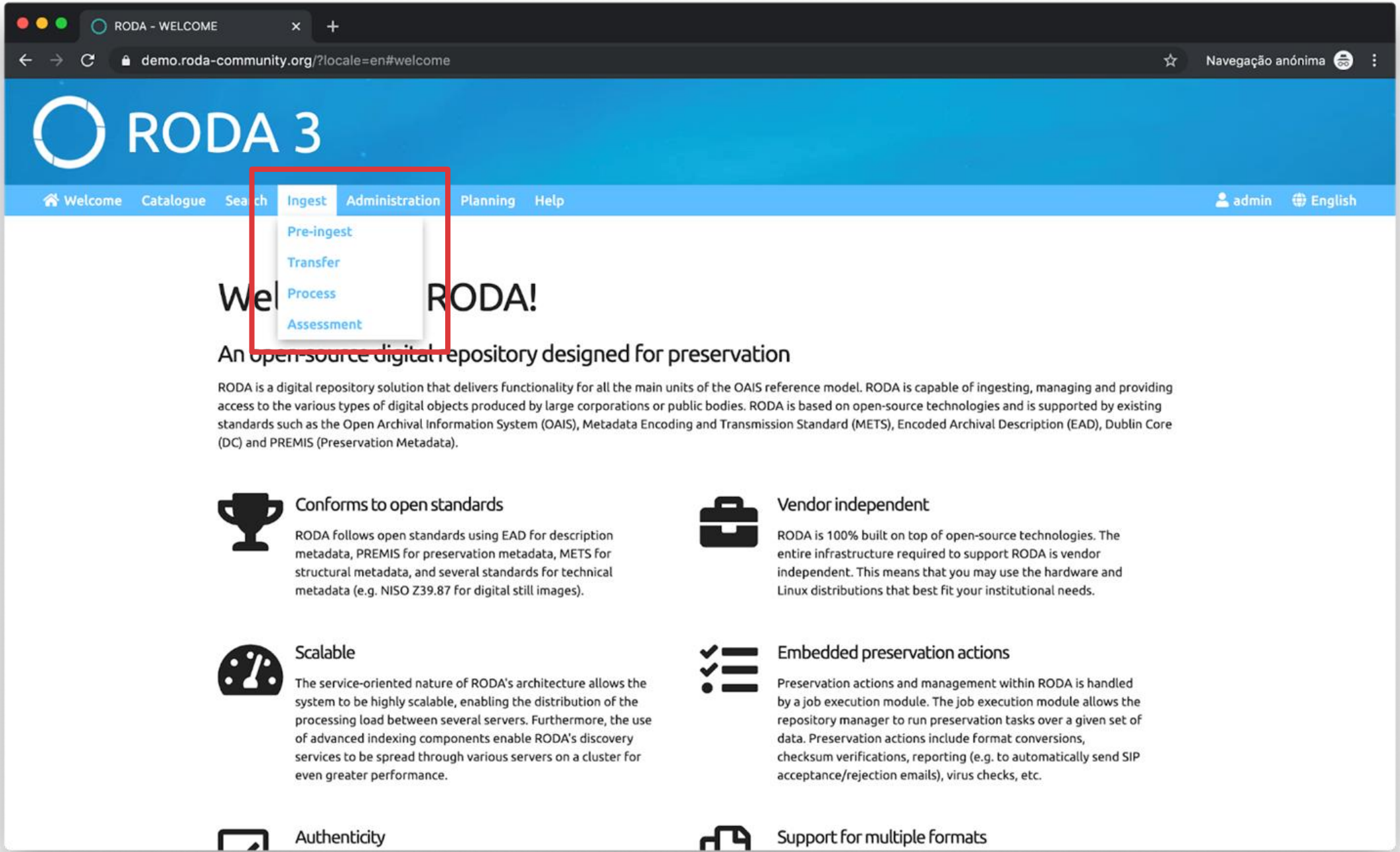
Well documented REST API and convenient Java libraries available on Github

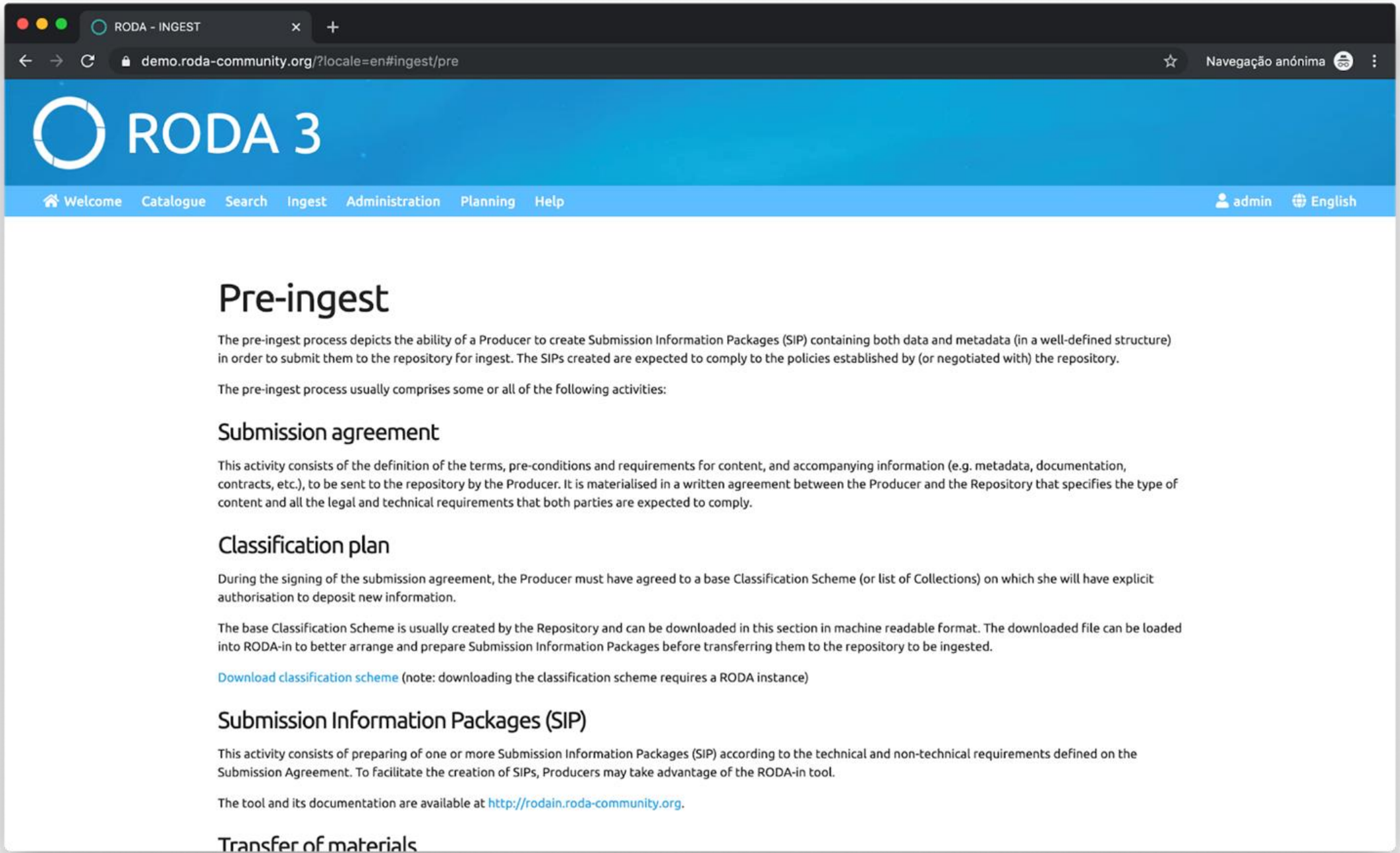
How does it work?

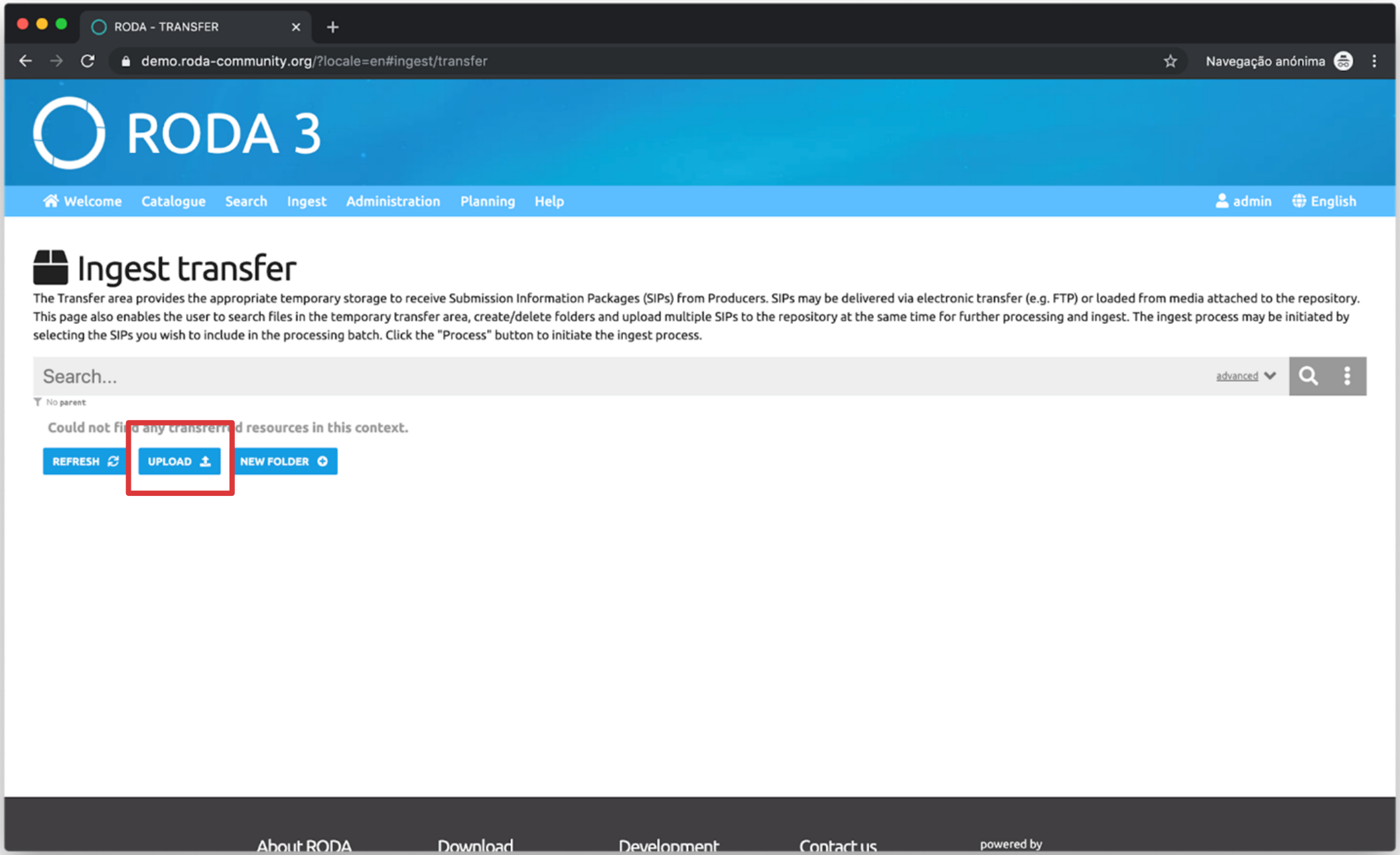


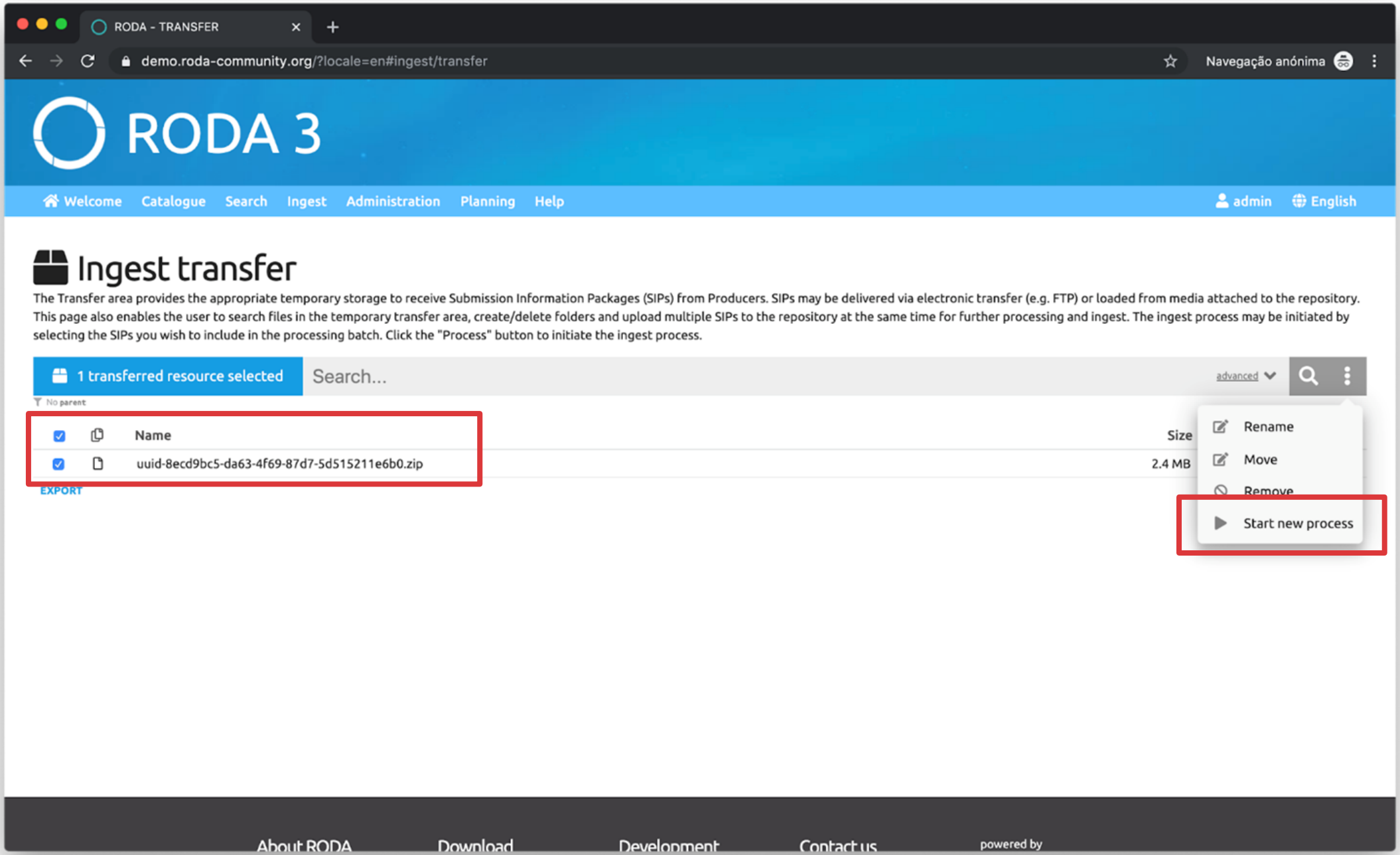
European Commission

Ingest...

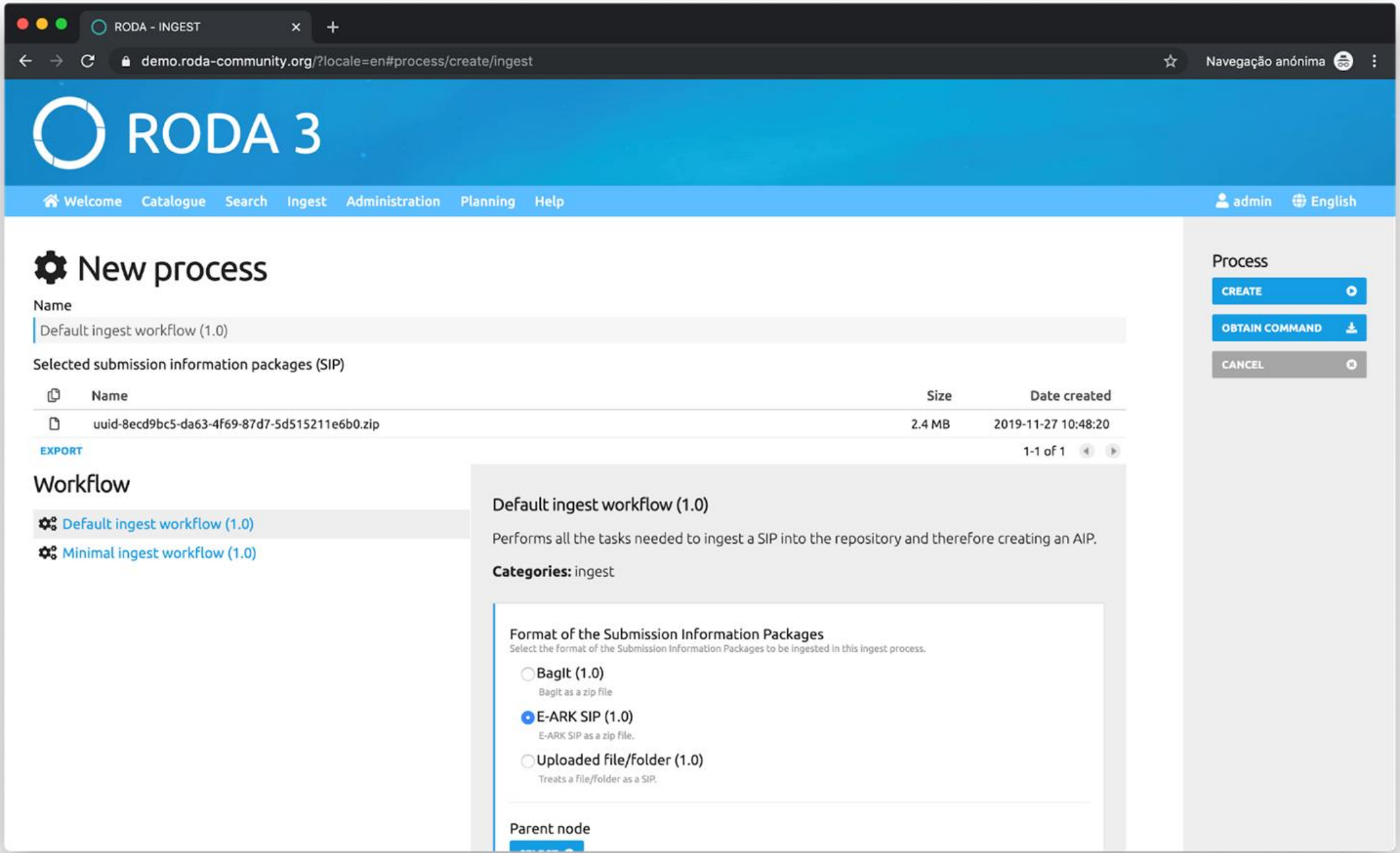




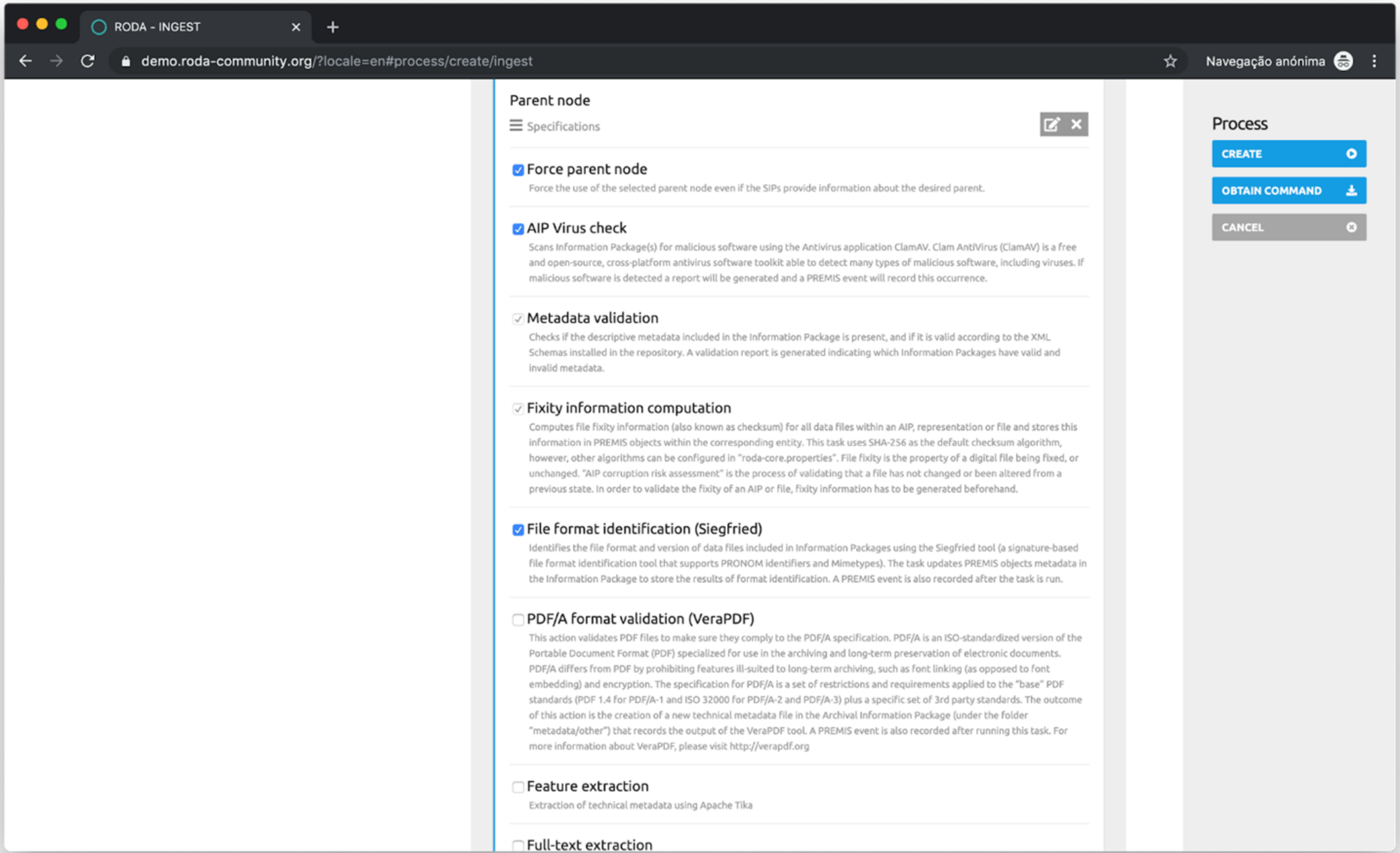


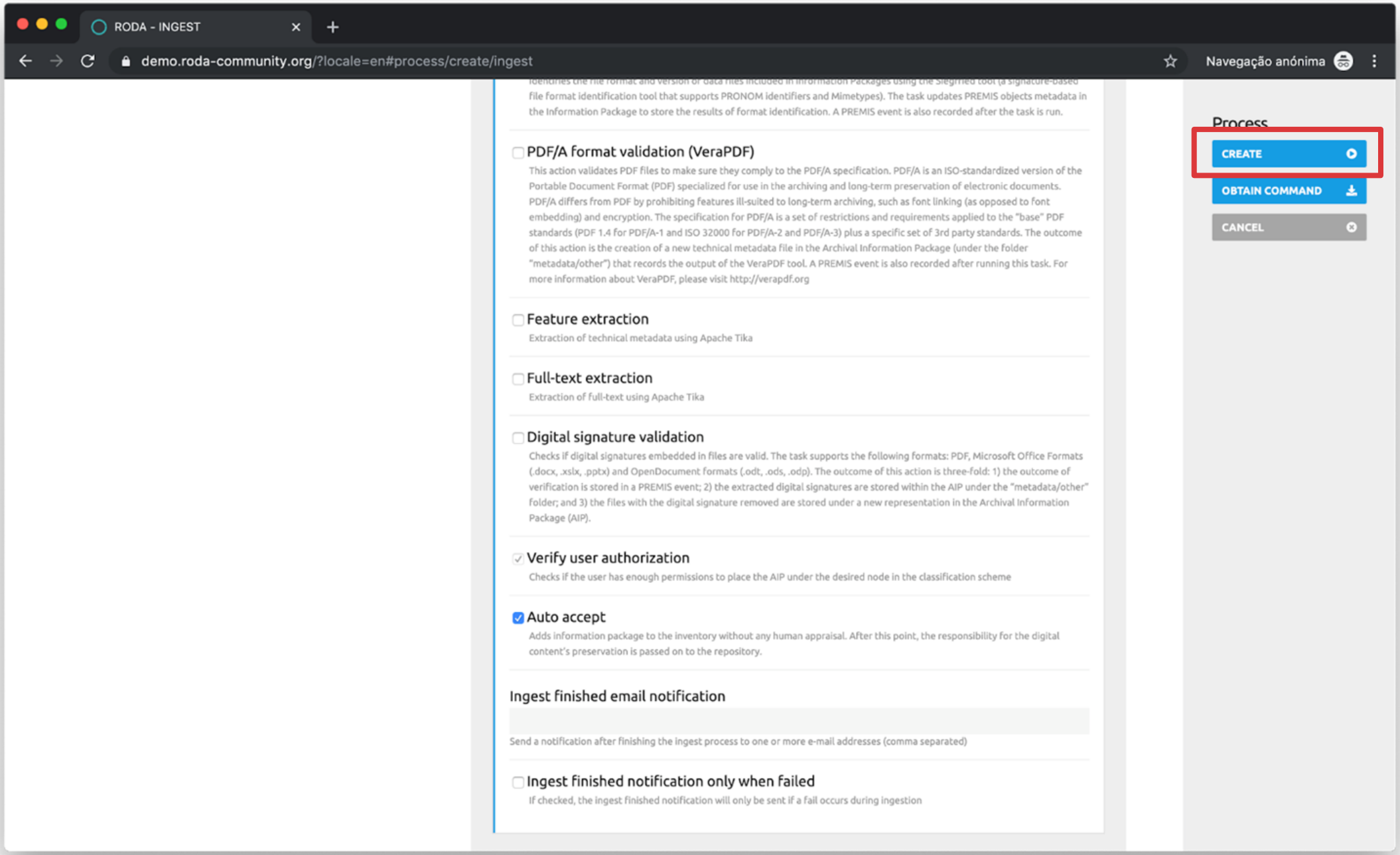


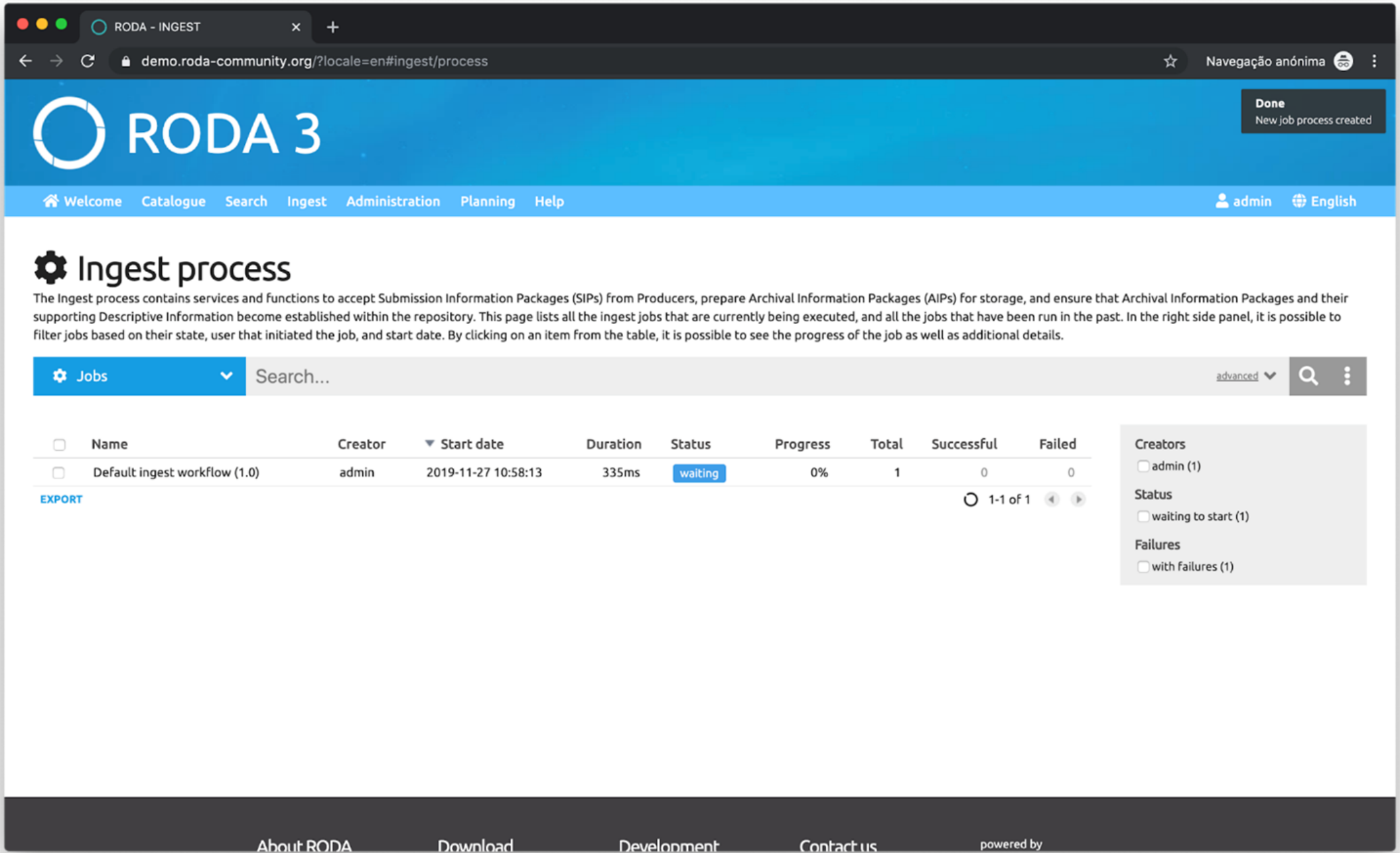
European Commission



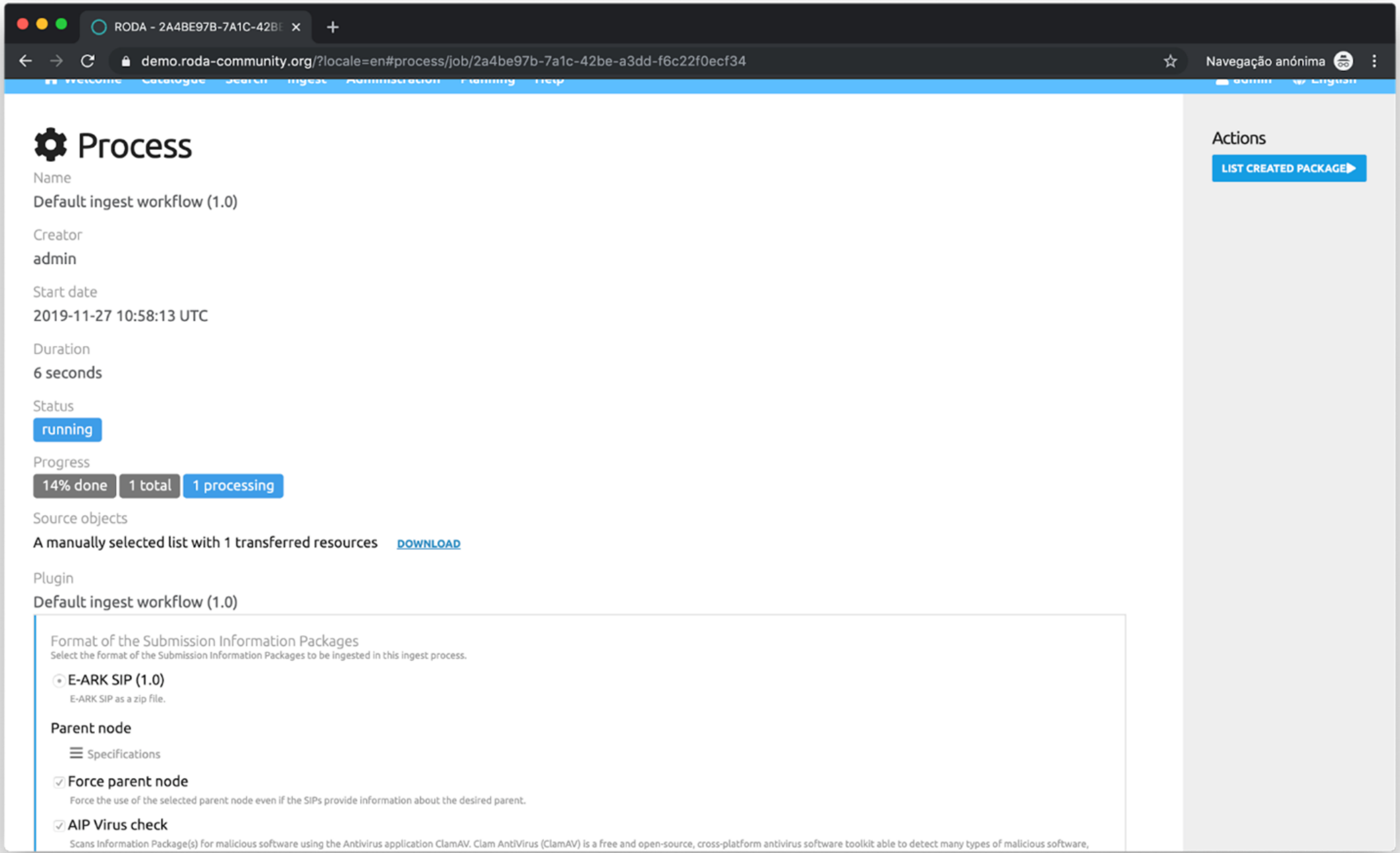
European Commission



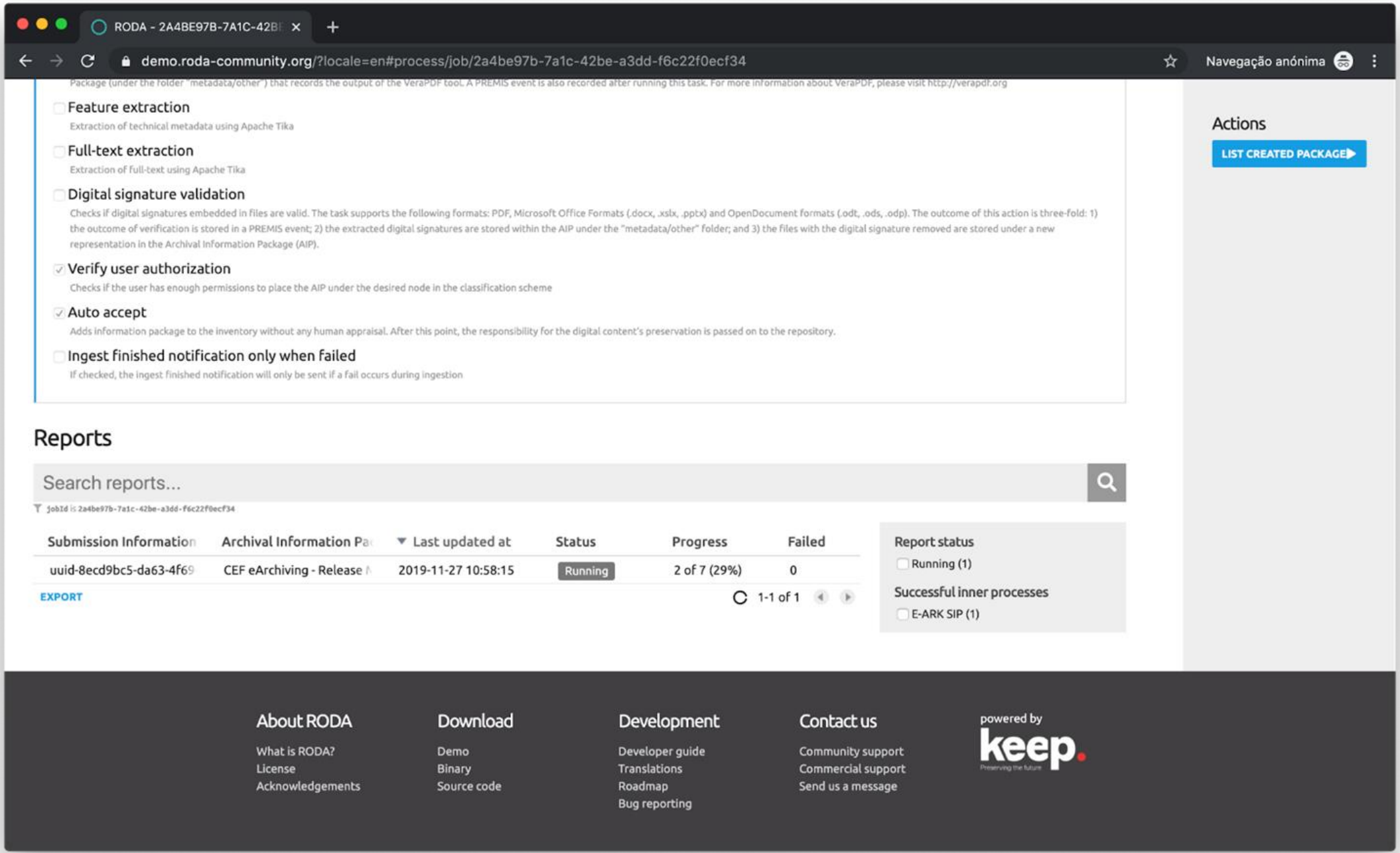




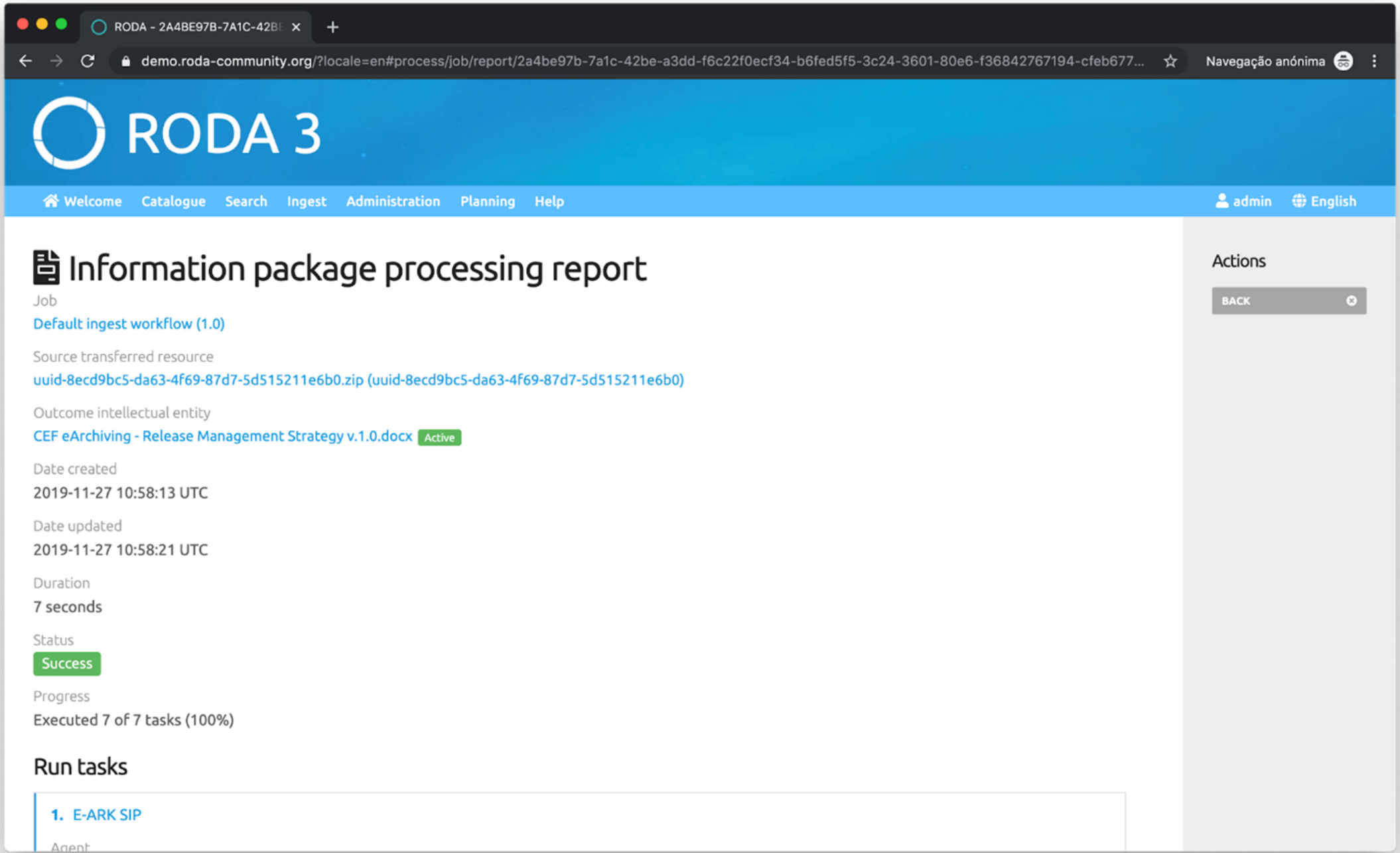
European Commission

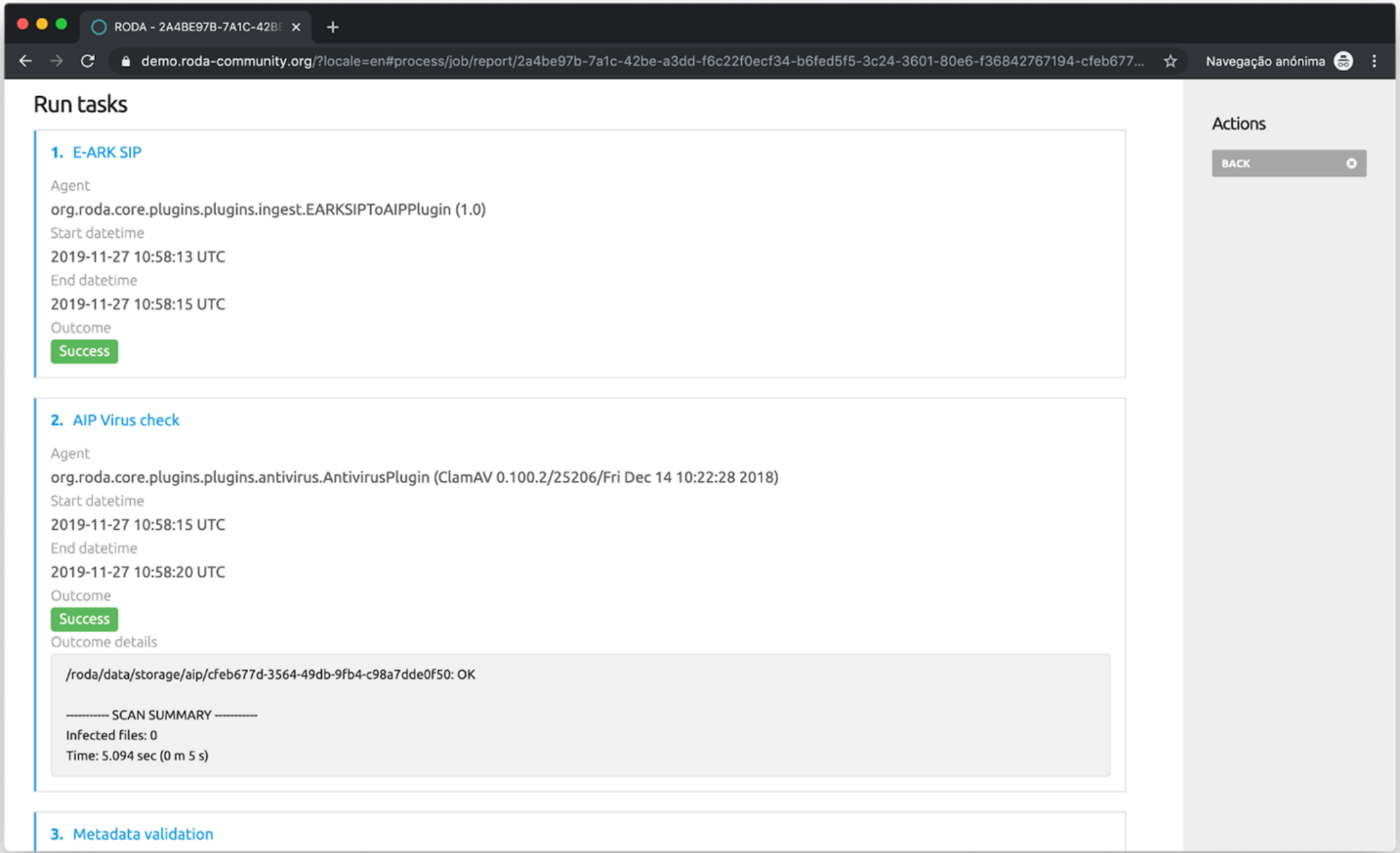


European Commission



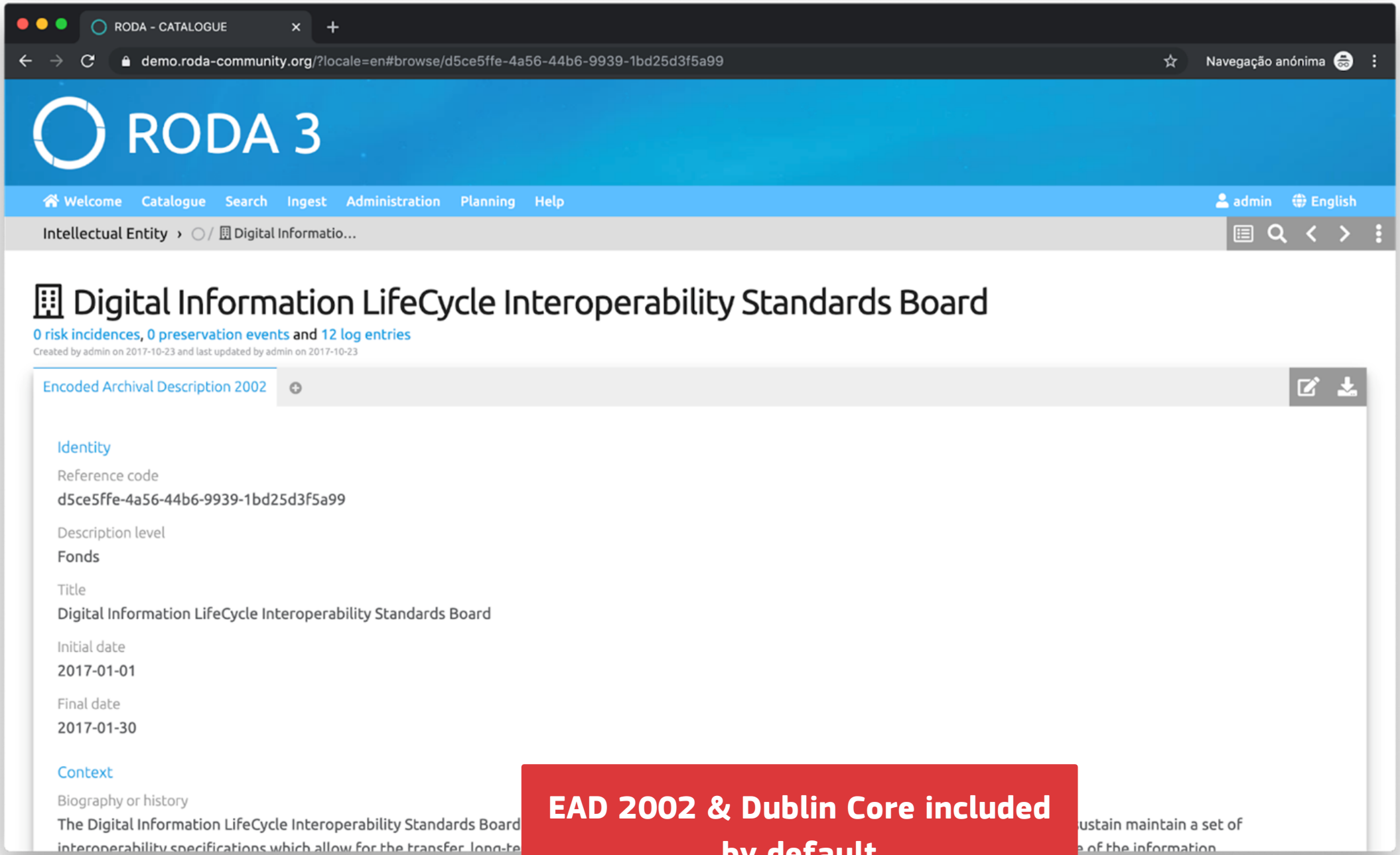
European Commission





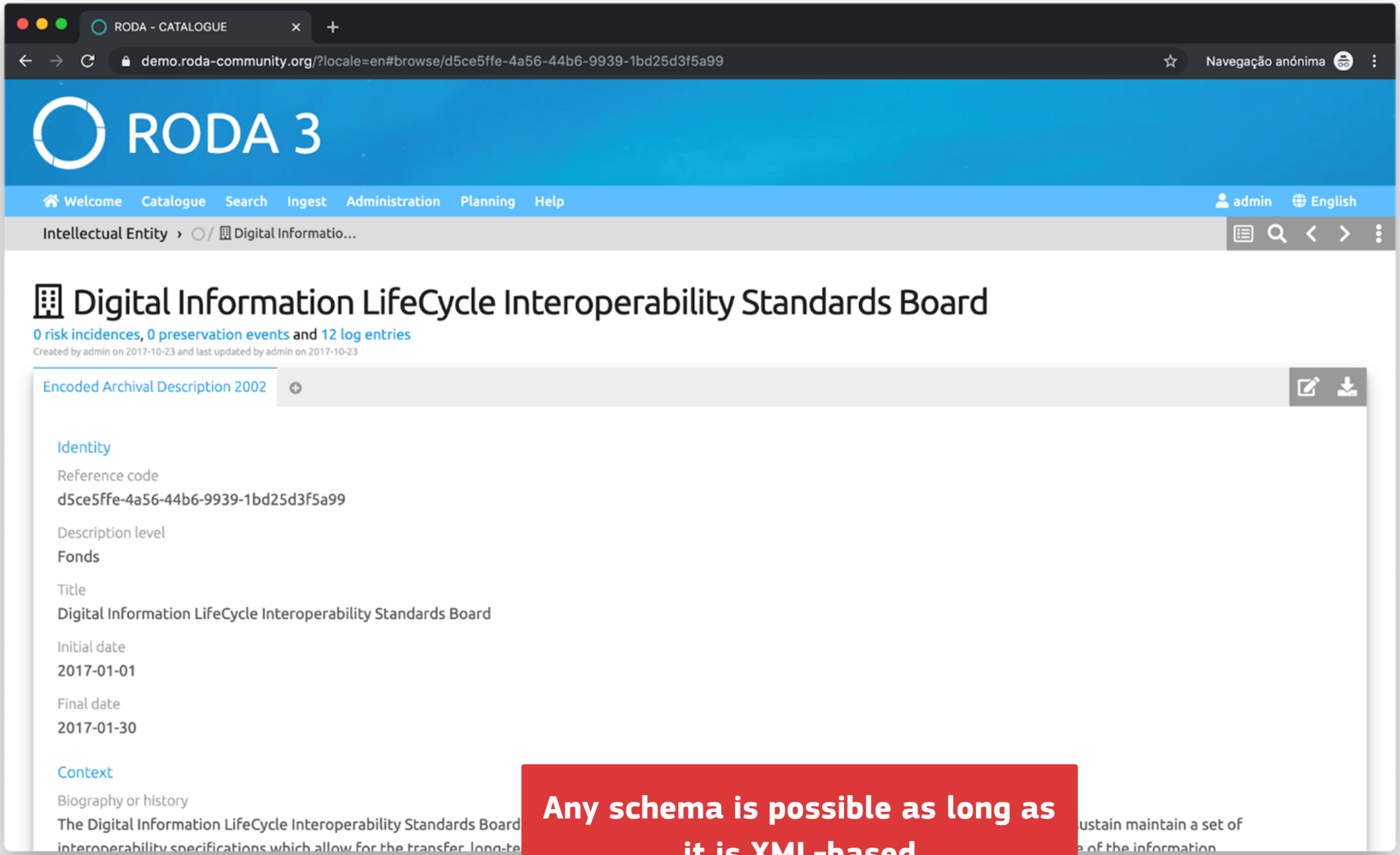
European Commission

Data management…



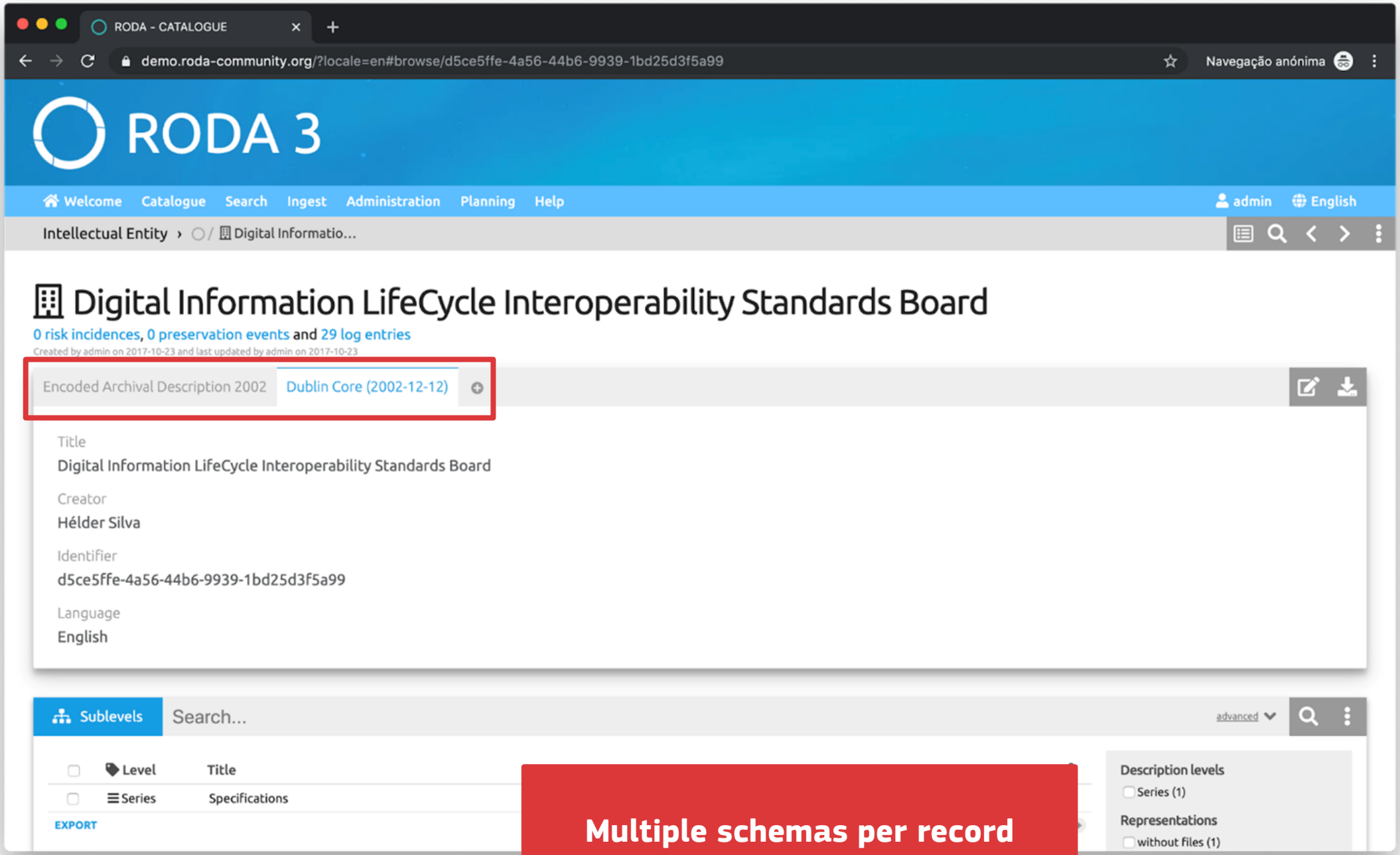
EAD 2002 & Dublin Core included by default

European Commission

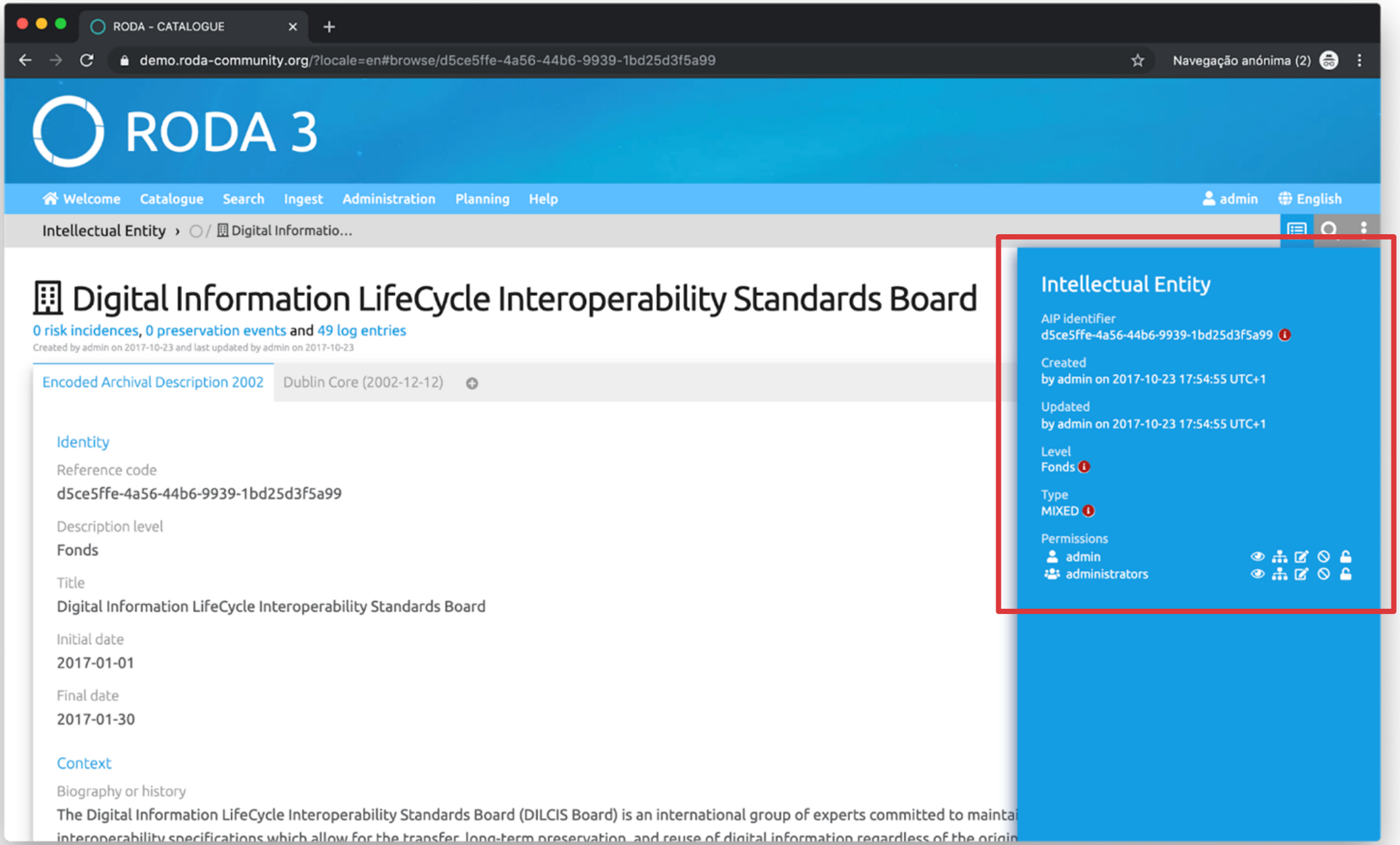


Any schema is possible as long as it is XML-based

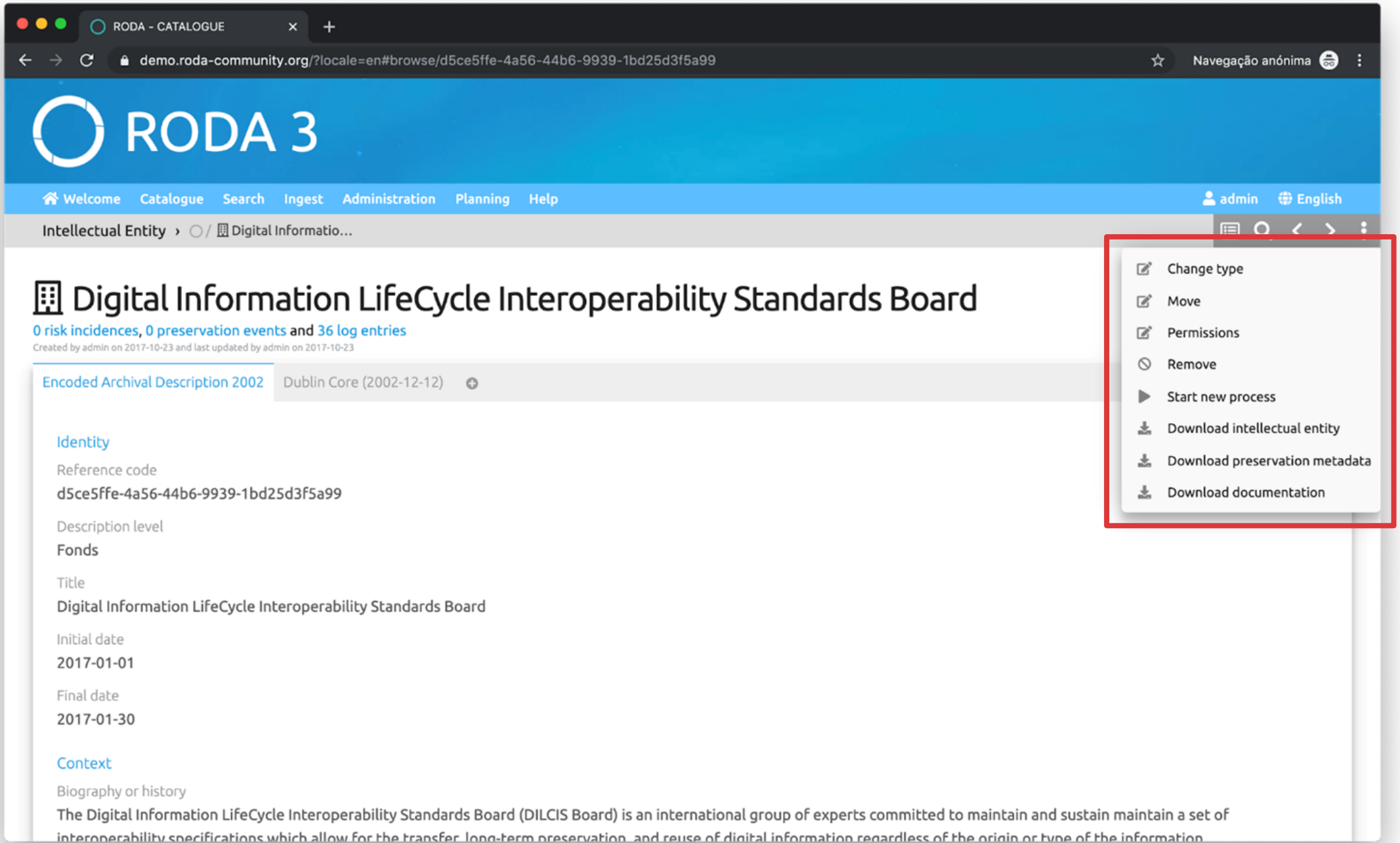
European Commission



Multiple schemas per record



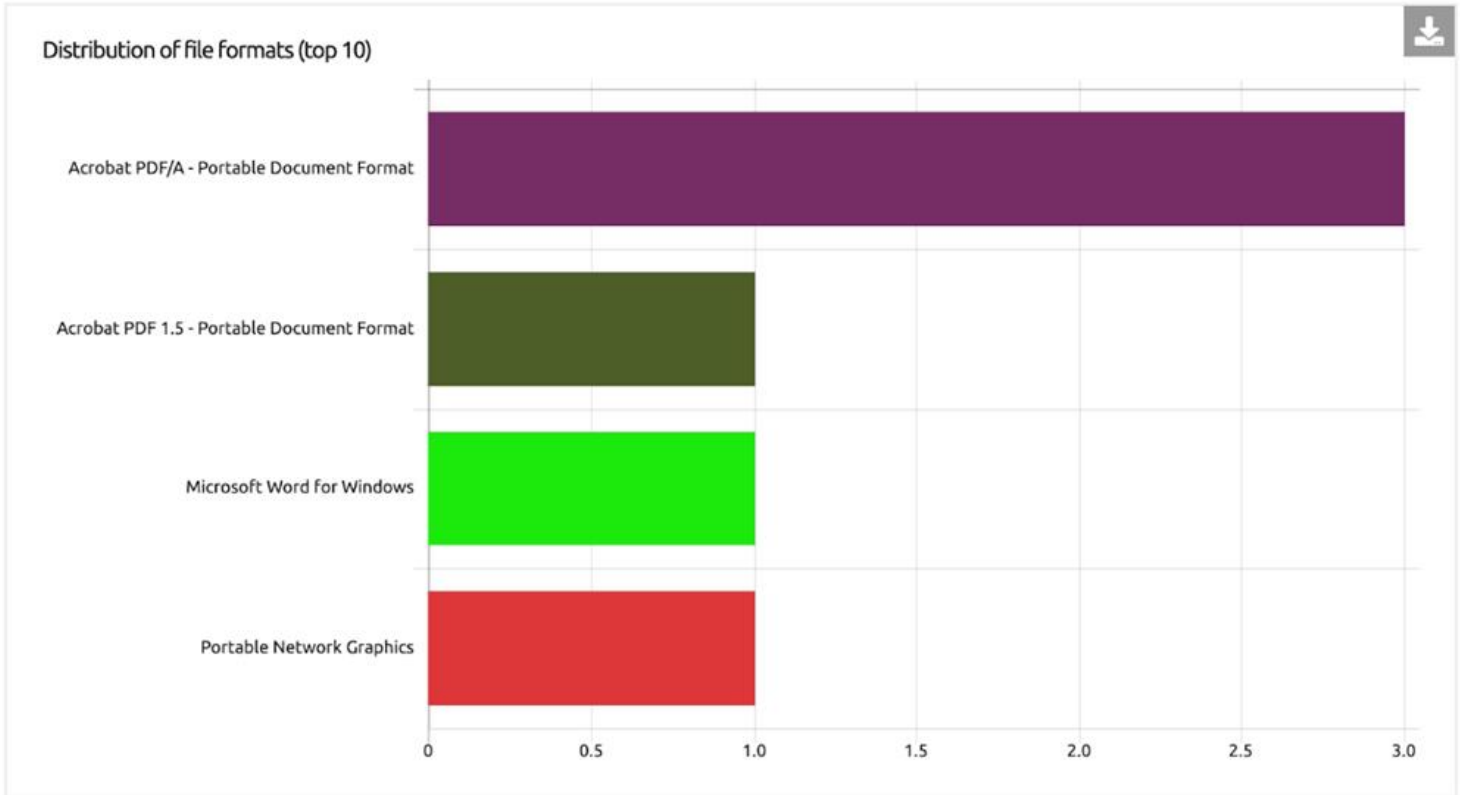
European Commission



RODA - STATISTICS

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

Navegação anónima



European Commission

Archival storage...

AIPs are stored directly on the file system

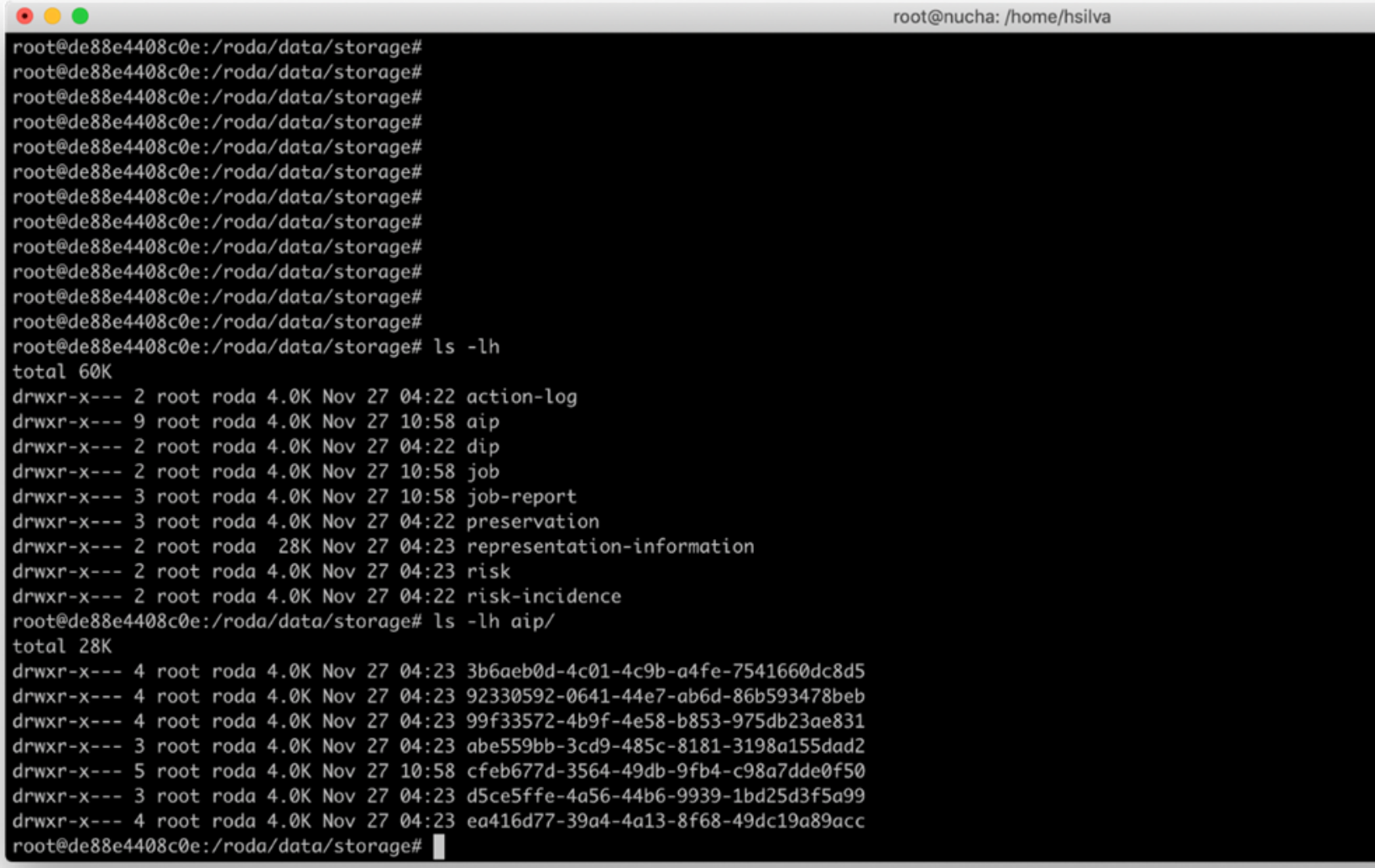
Better transparency

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



AIPs are stored directly on the file system

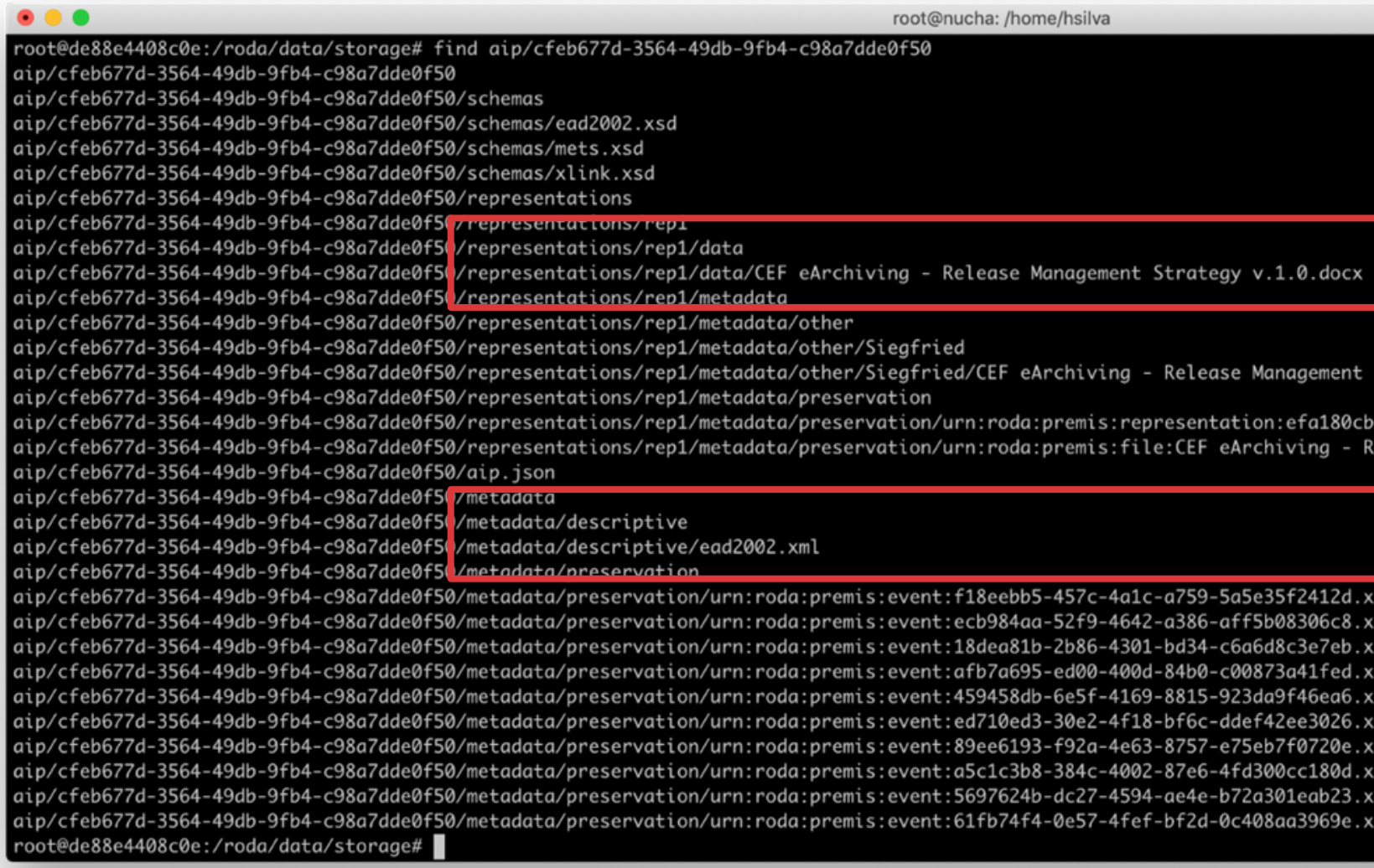
Better transparency

Easier backups and replication

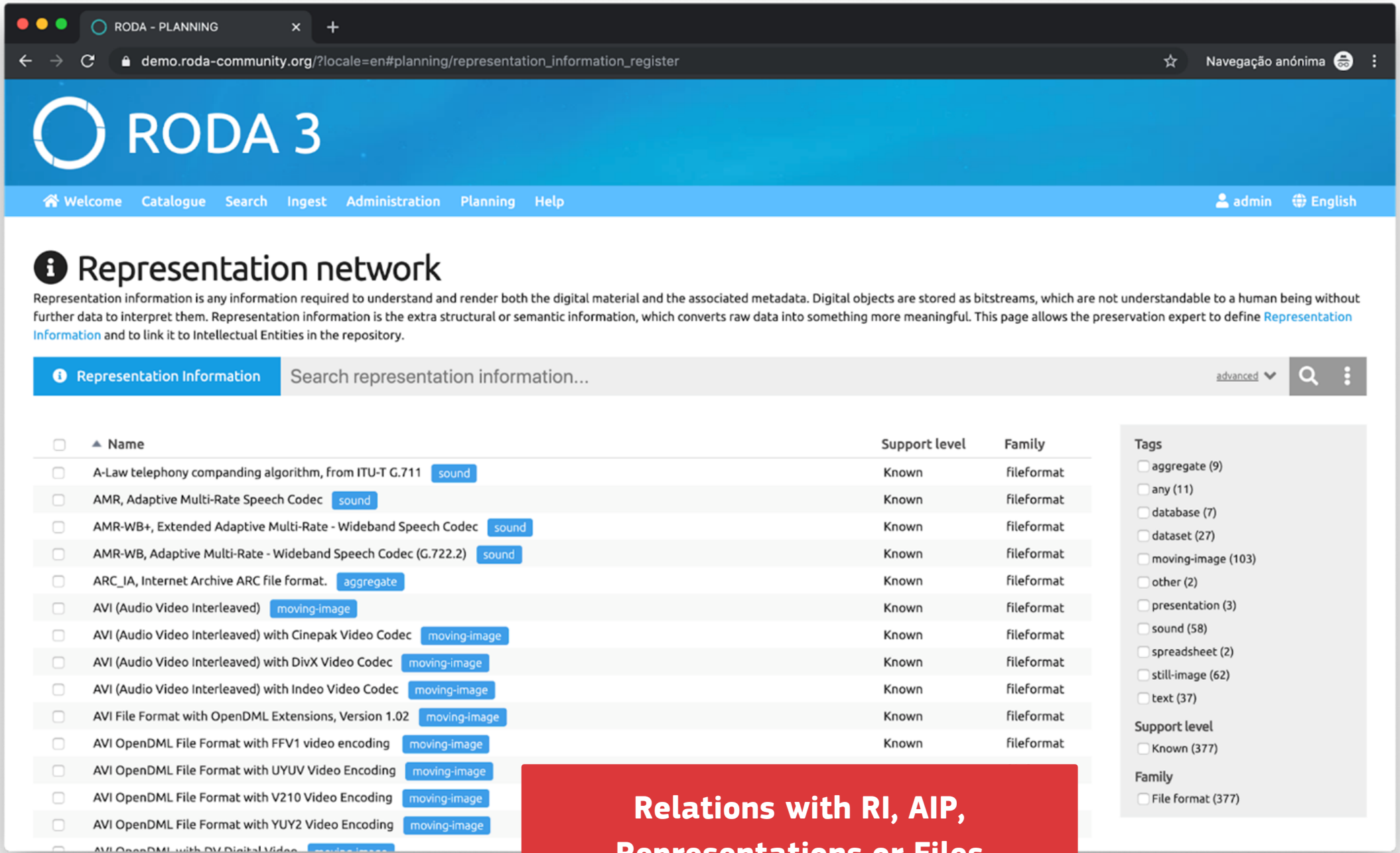
Well defined specifications for SIP, AIP, DIP

Managed by the DILCIS Board - a subgroup of DLM Forum

Part of the eArchiving CEF Building Block to promote European adoption

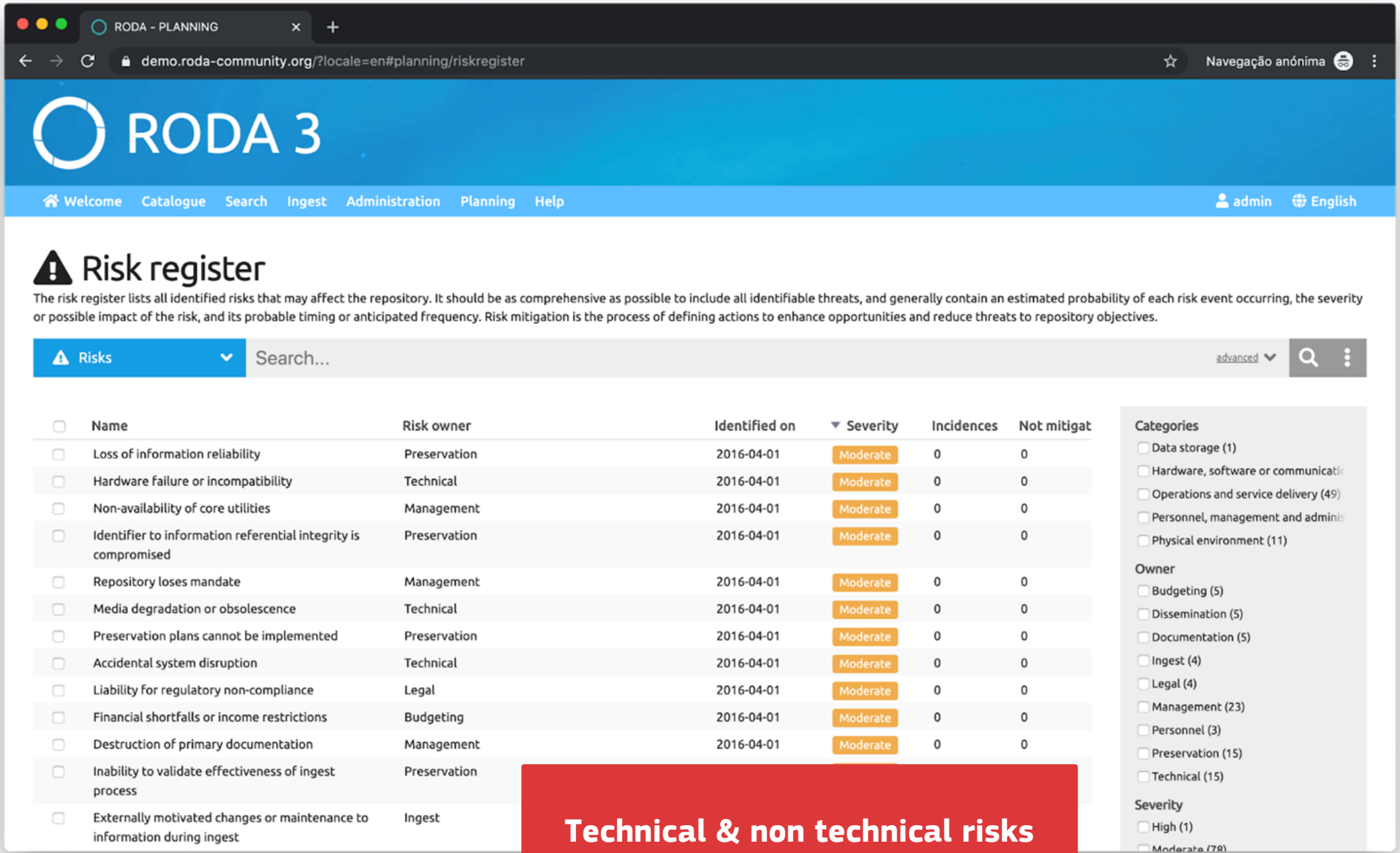


Preservation planning...



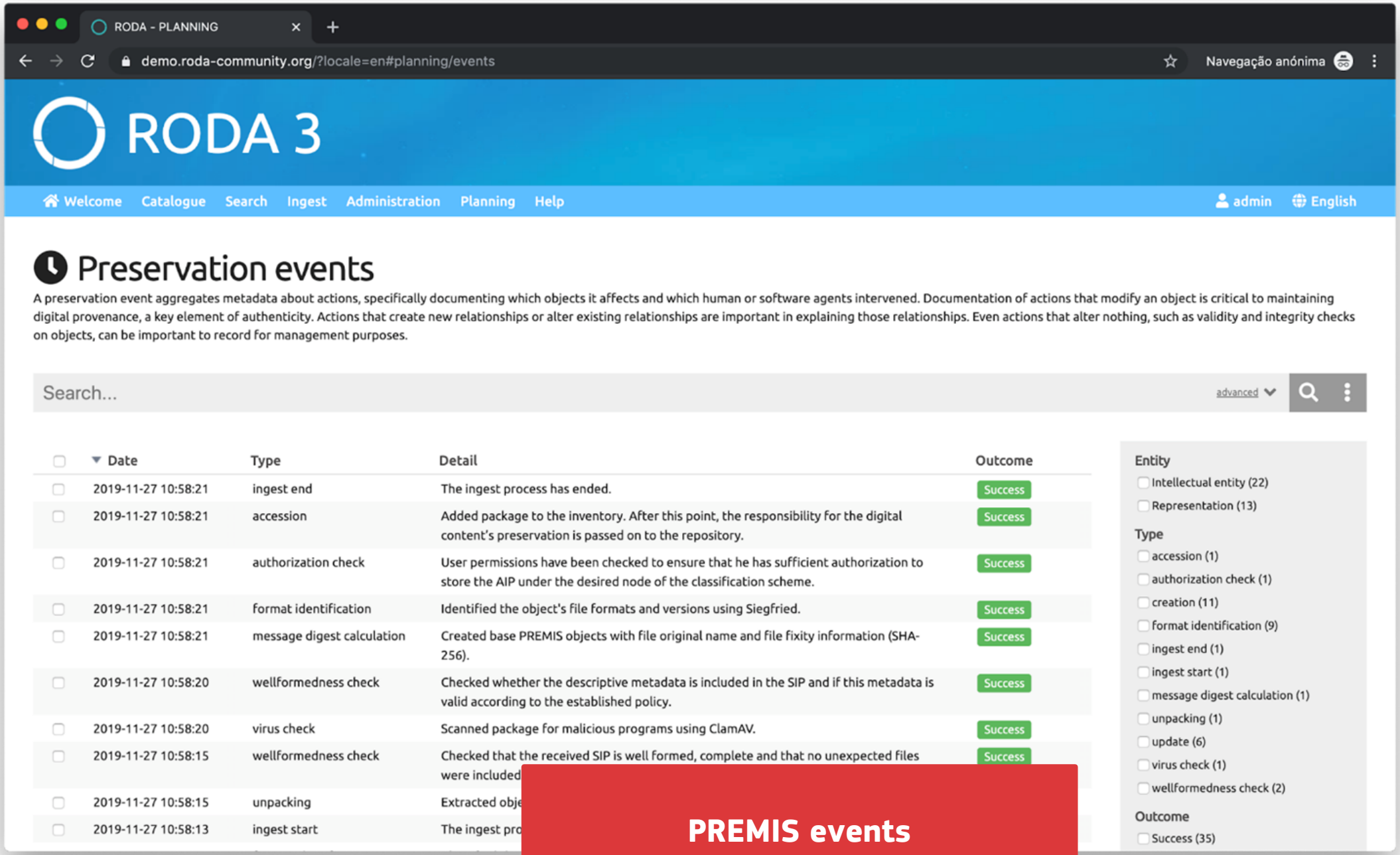
Relations with RI, AIP, Representations or Files

European Commission

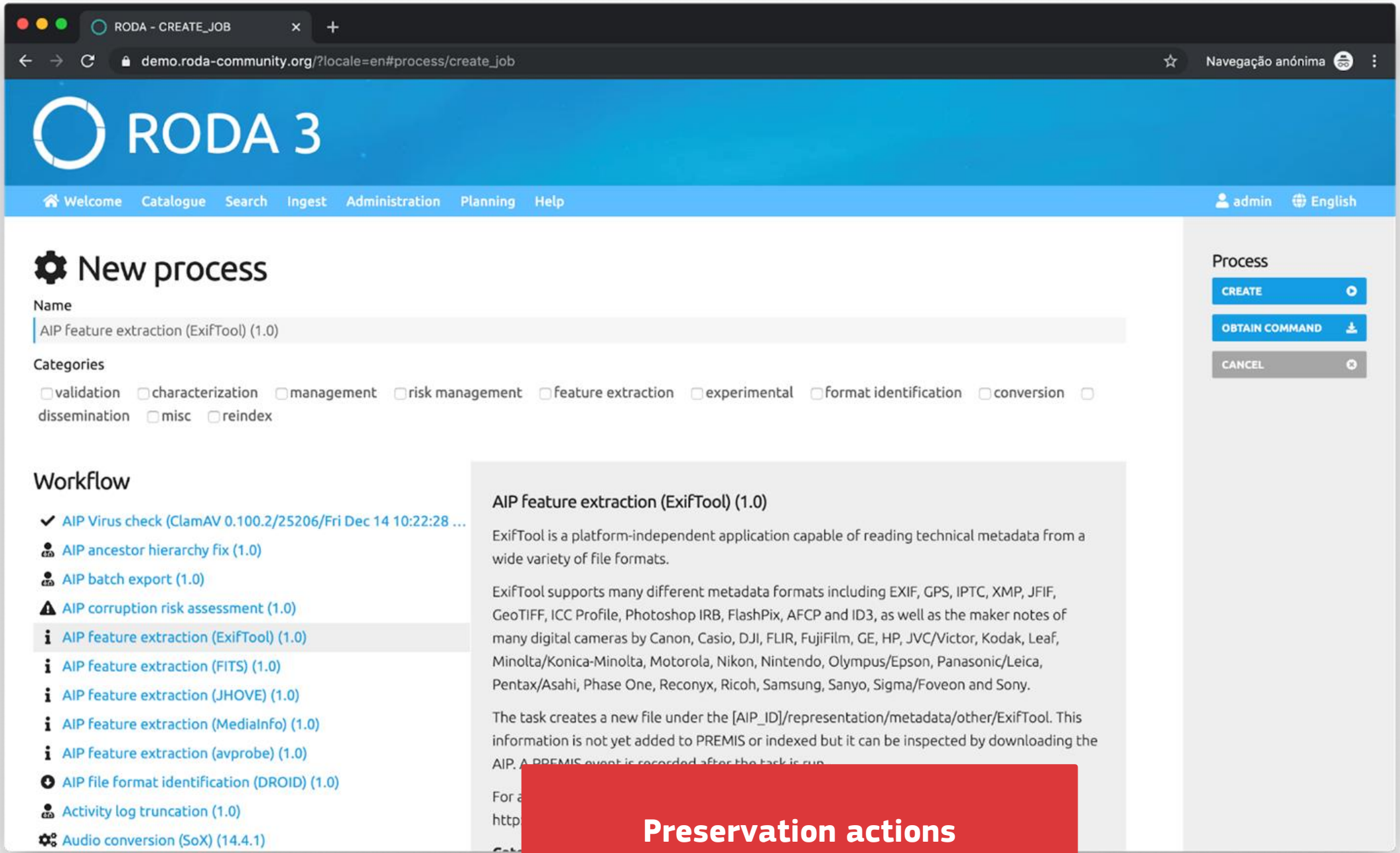


Technical & non technical risks

European Commission

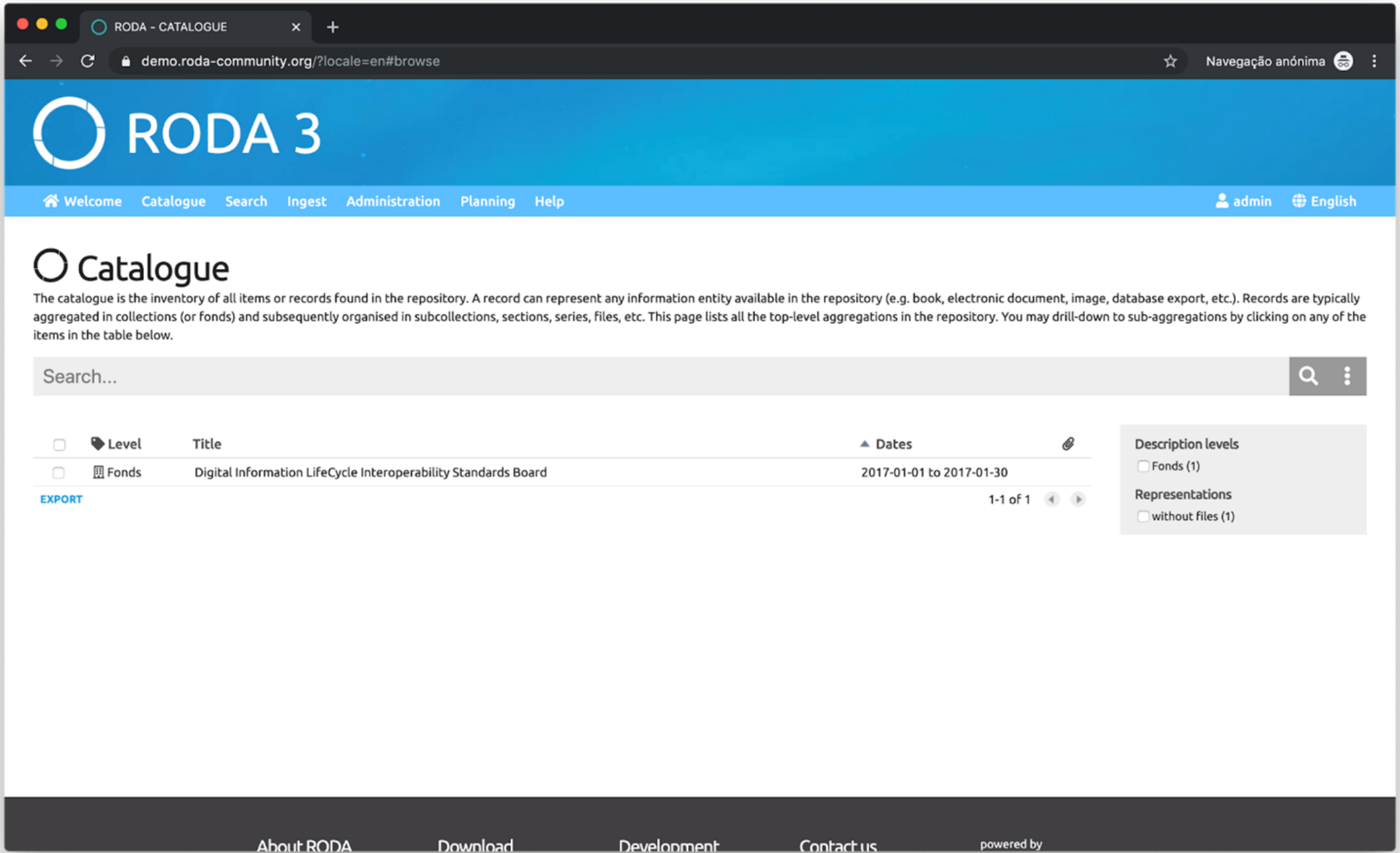


PREMIS events

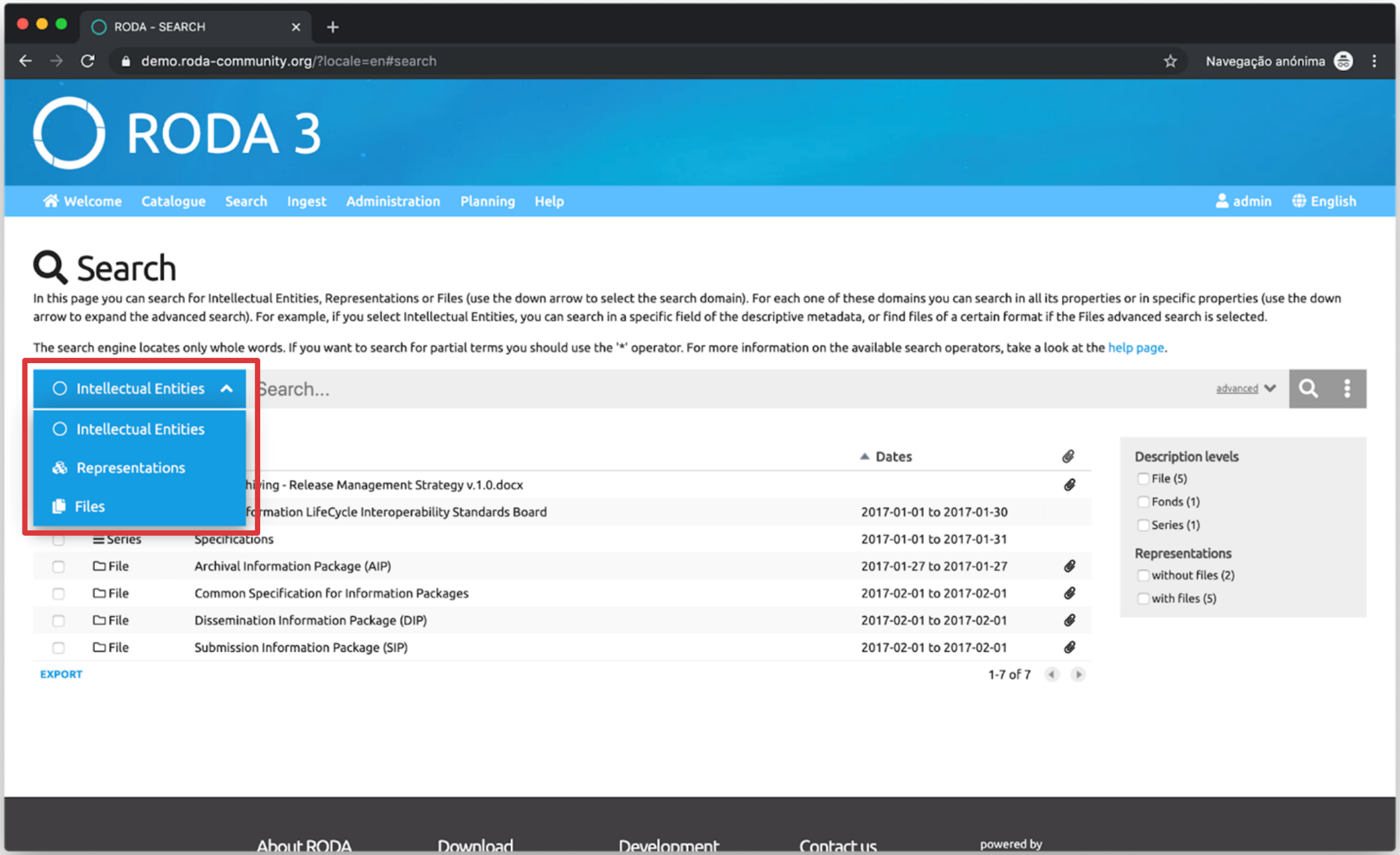


Preservation actions

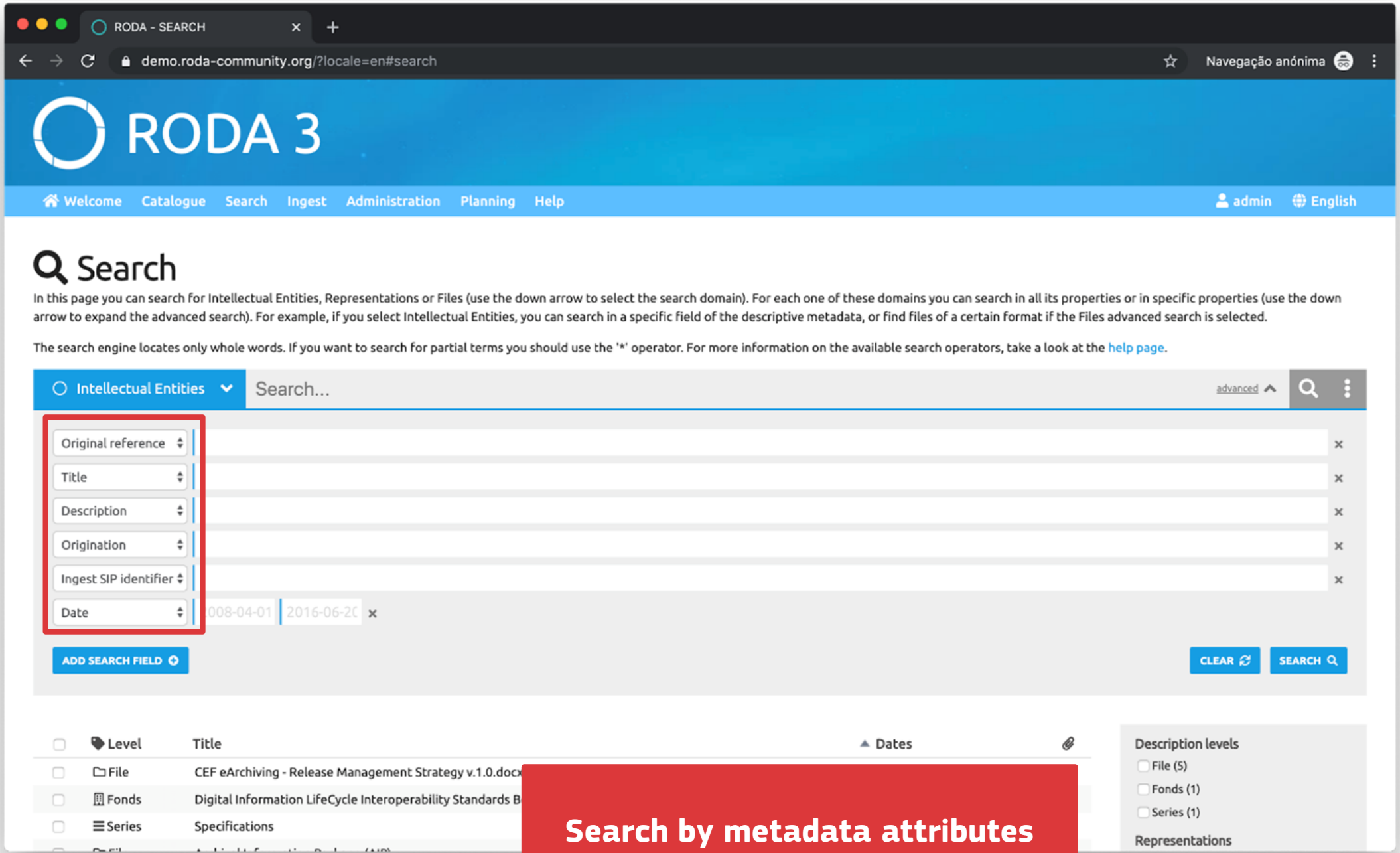
Access...



European Commission

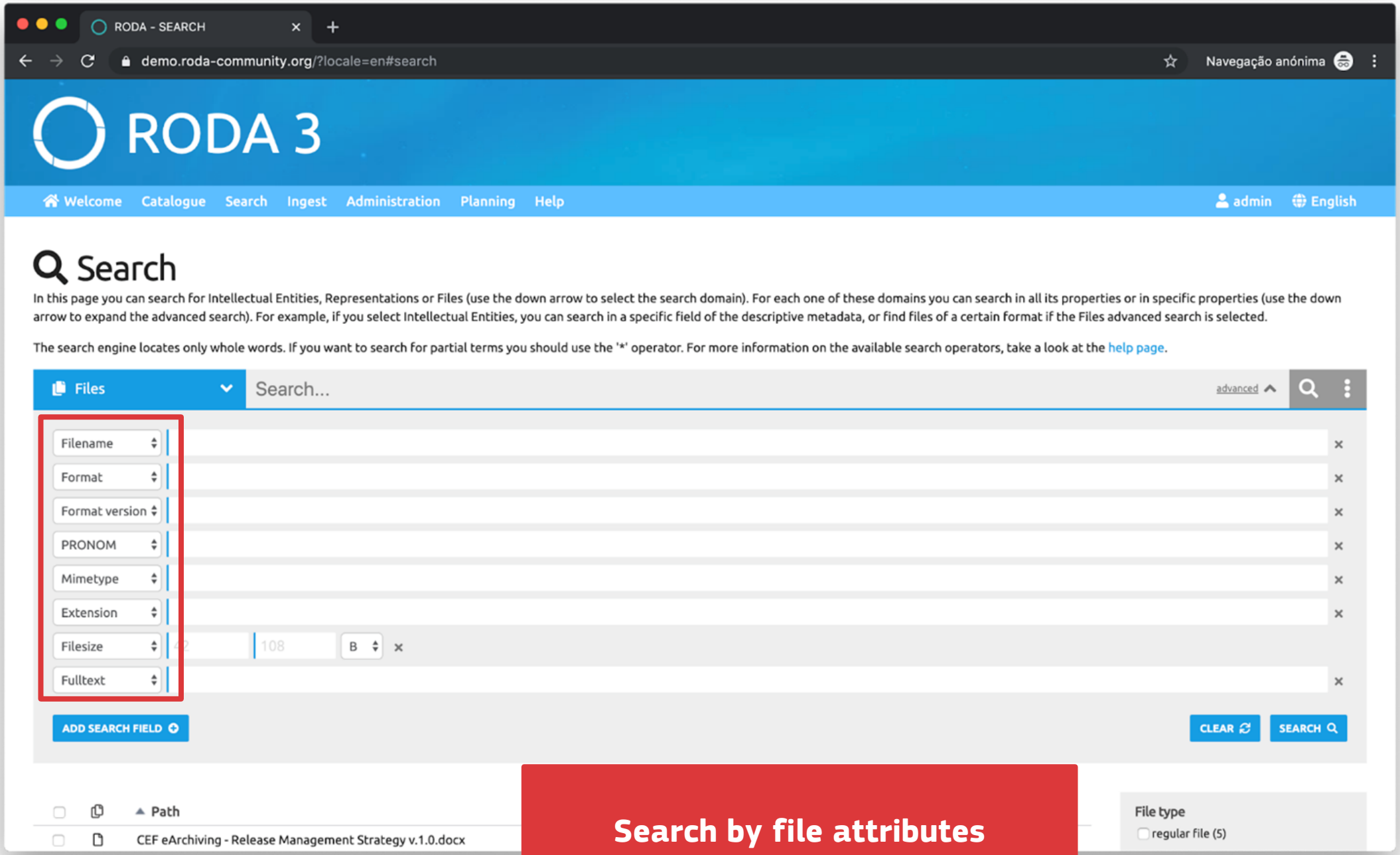


European Commission



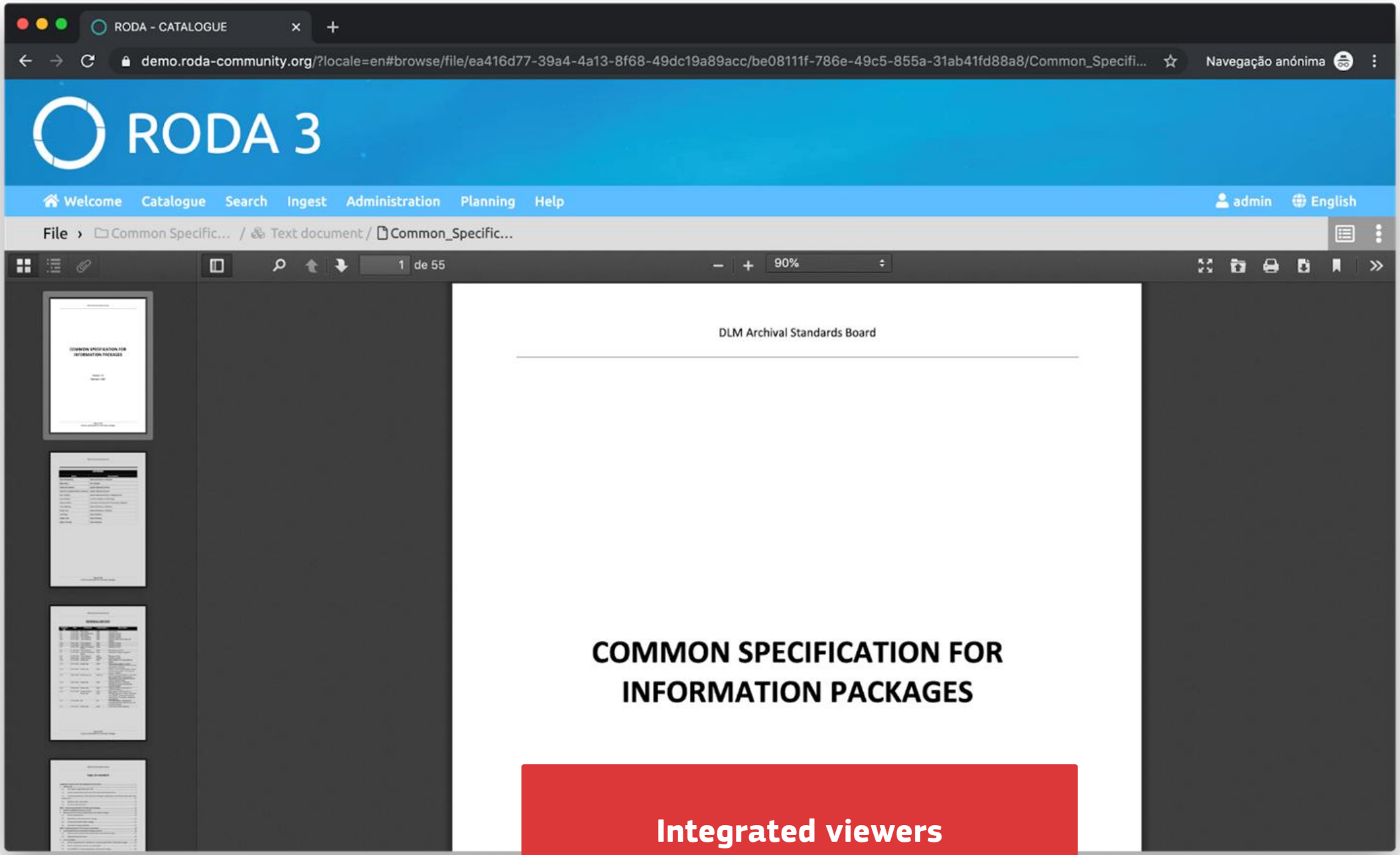
Search by metadata attributes

European Commission

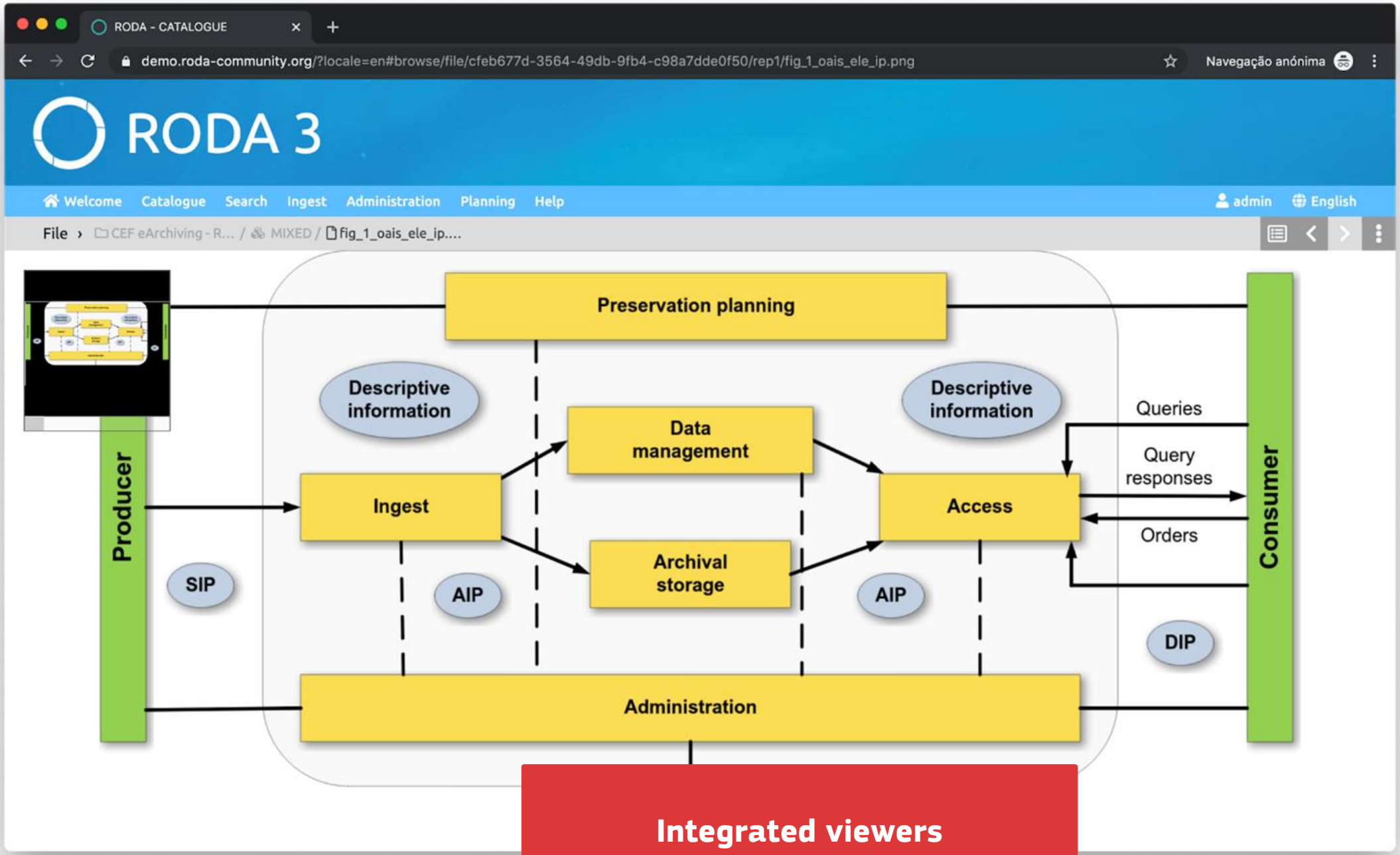


Search by file attributes

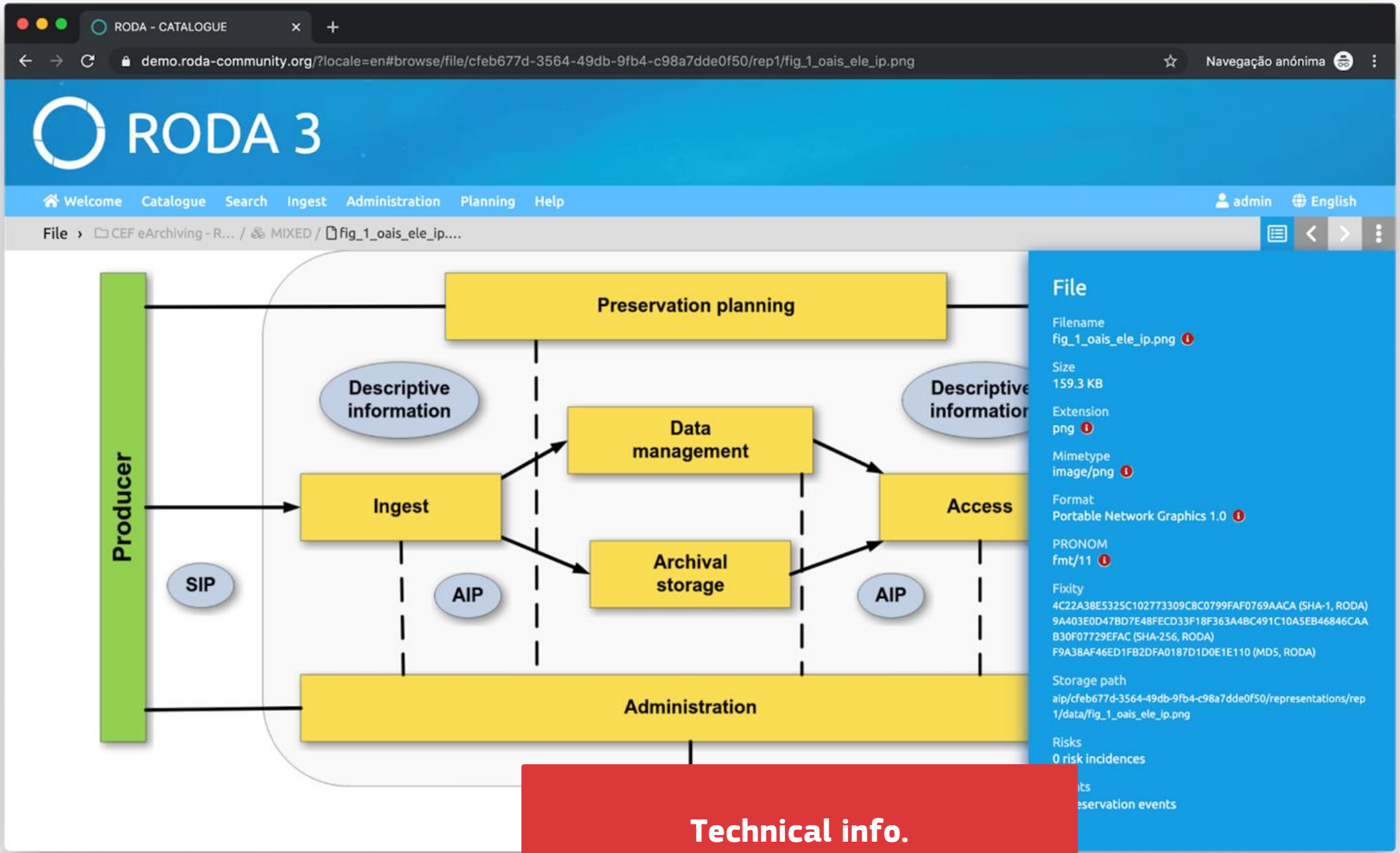
European Commission



Integrated viewers

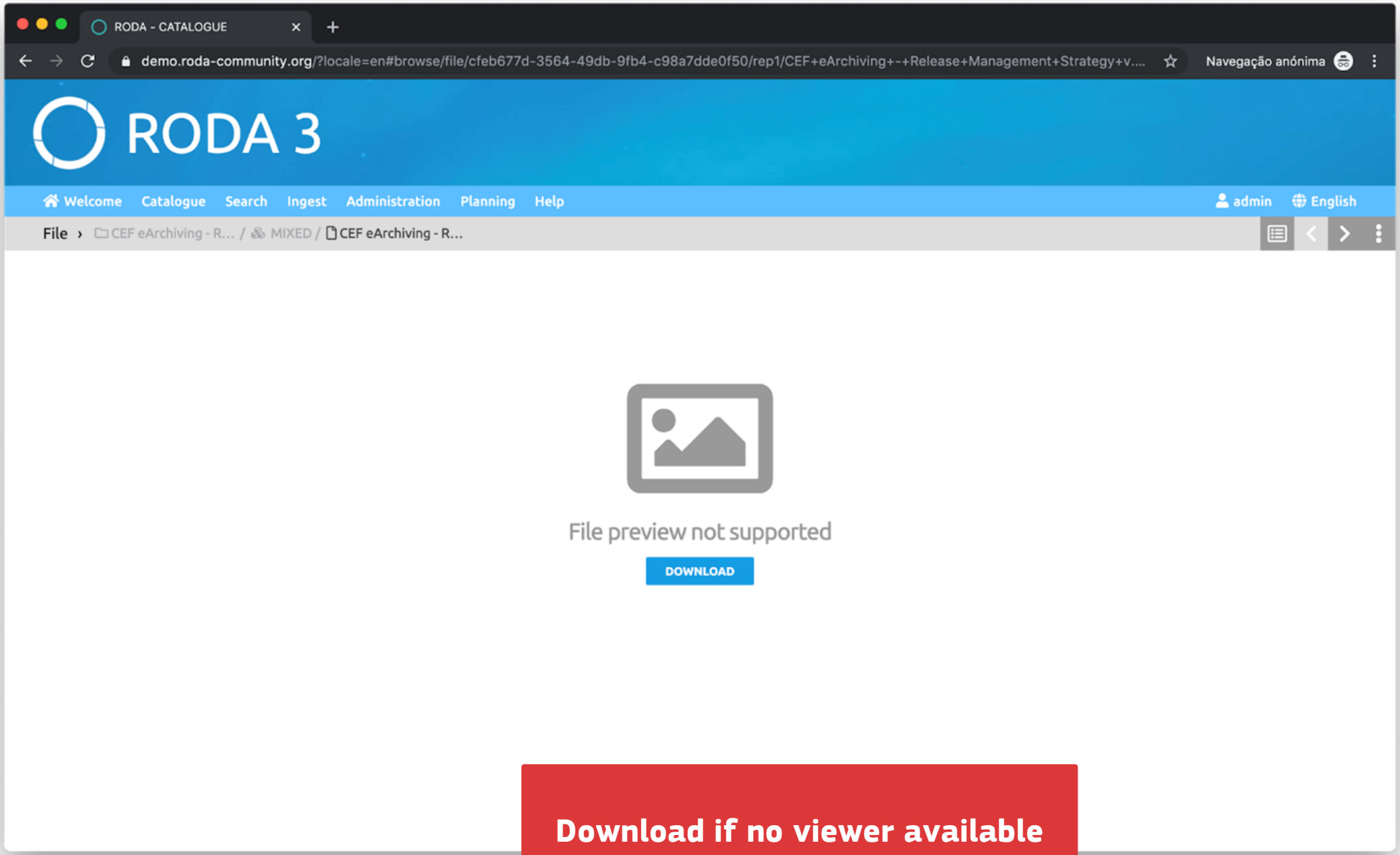


Integrated viewers



Technical info.

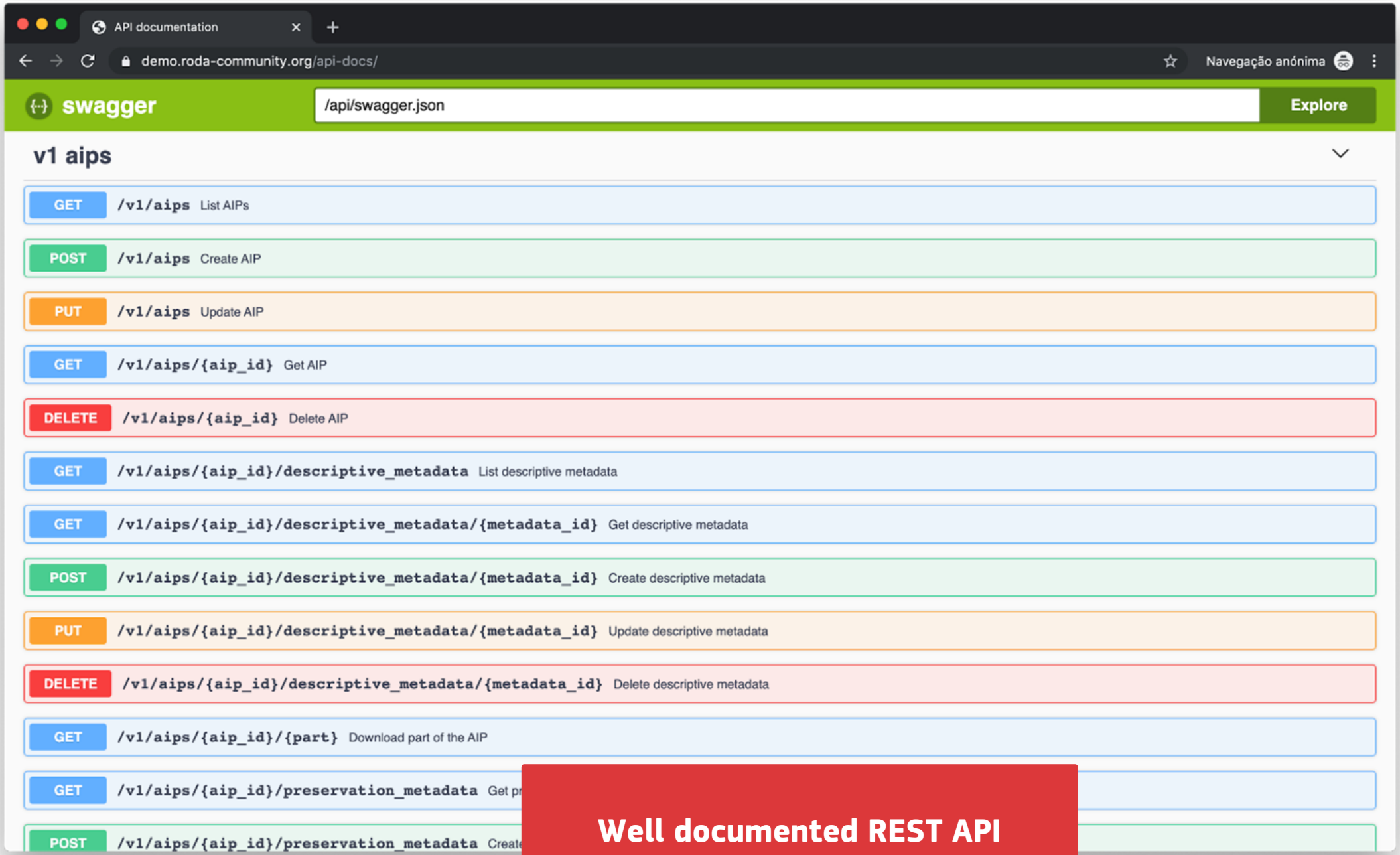
European Commission



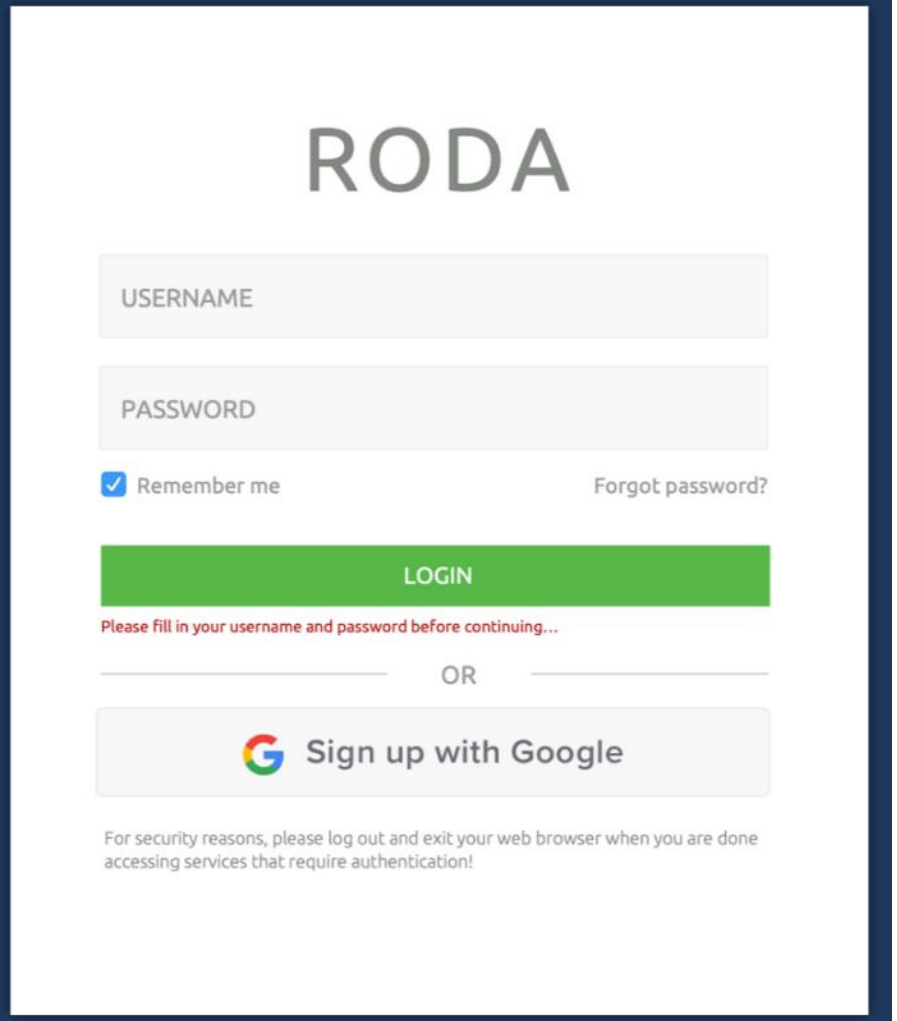
Download if no viewer available

European Commission

Worthwhile mentioning...



Well documented REST API



Central Authentication Service

Local user database

CAS v1, v2 and v3

SAML v1 and v2 Protocol

OAuth

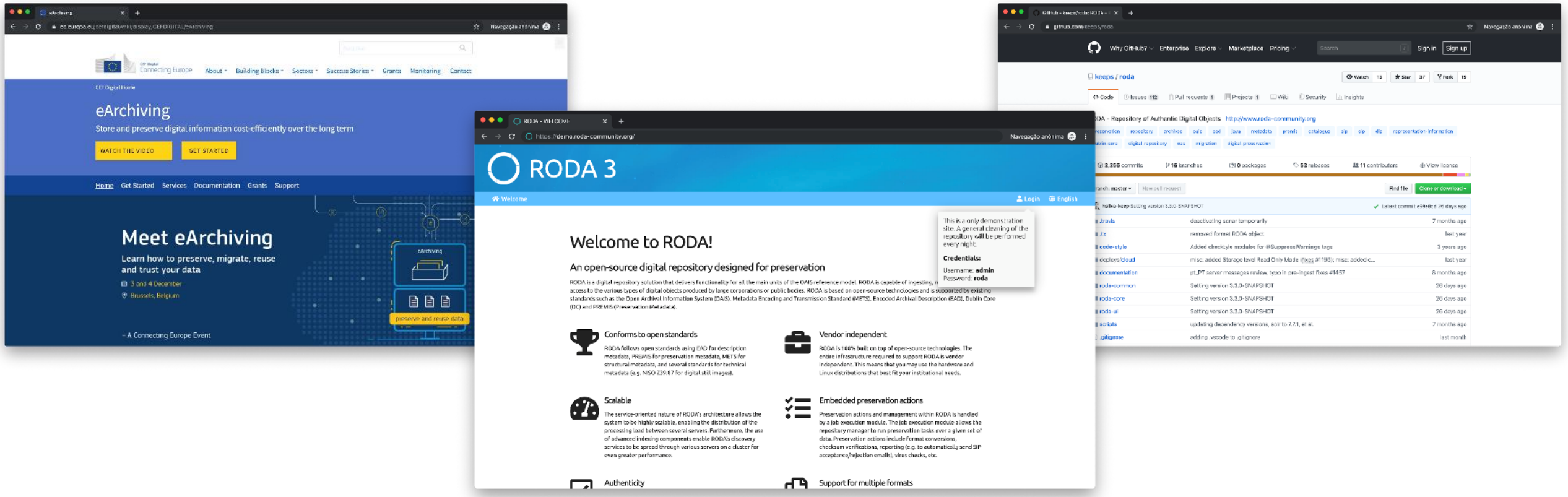
OpenID

WS-Federation Passive Requestor

Delegated authentication

... more

Where can I find it?



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

github.com/keeps/roda

www.roda-community.org

European Commission

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

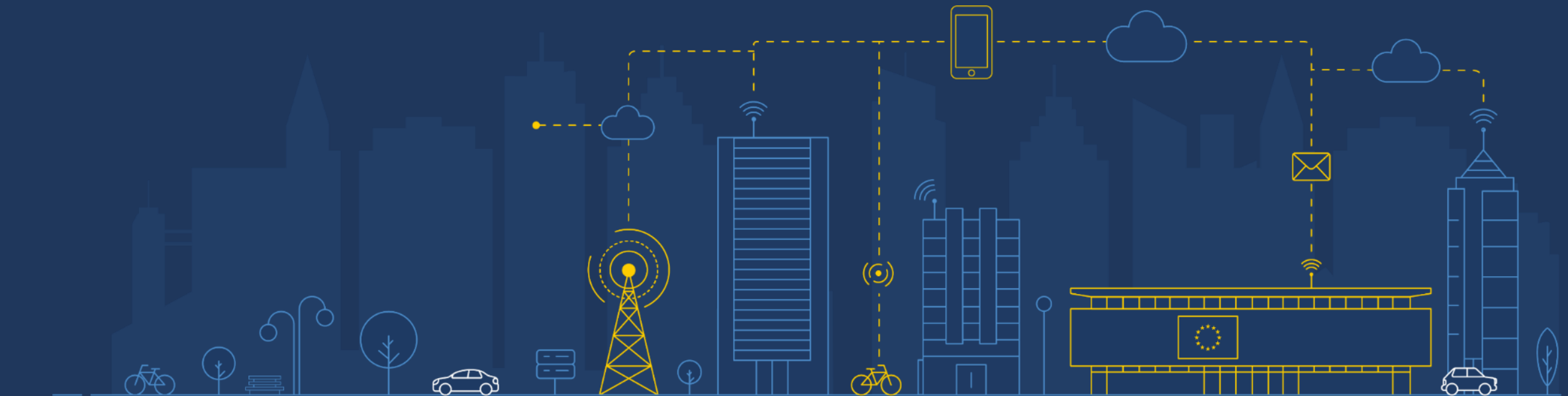
Contact us

info@keep.pt

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

Icons from www.flaticon.com

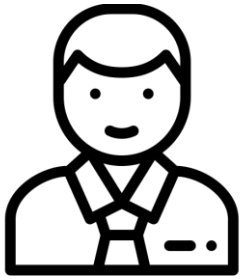
Database

Export tool

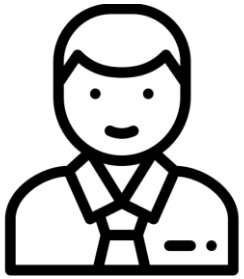
Preservation format

Icons from www.flaticon.com

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



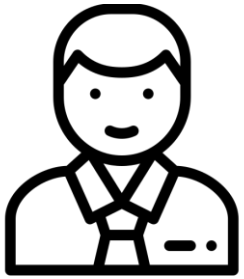
Icons from www.flaticon.com



Find

Repository

Icons from www.flaticon.com



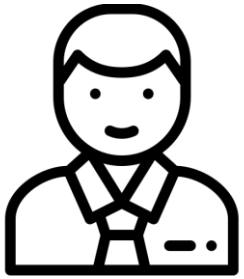
Find

Repository

Retrieve

Preservation format

Icons from www.flaticon.com



Find

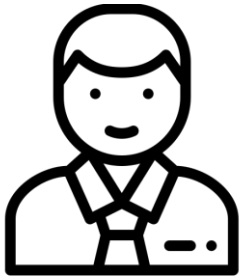
Repository

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

Icons from www.flaticon.com

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

database.toolkit
preservation

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

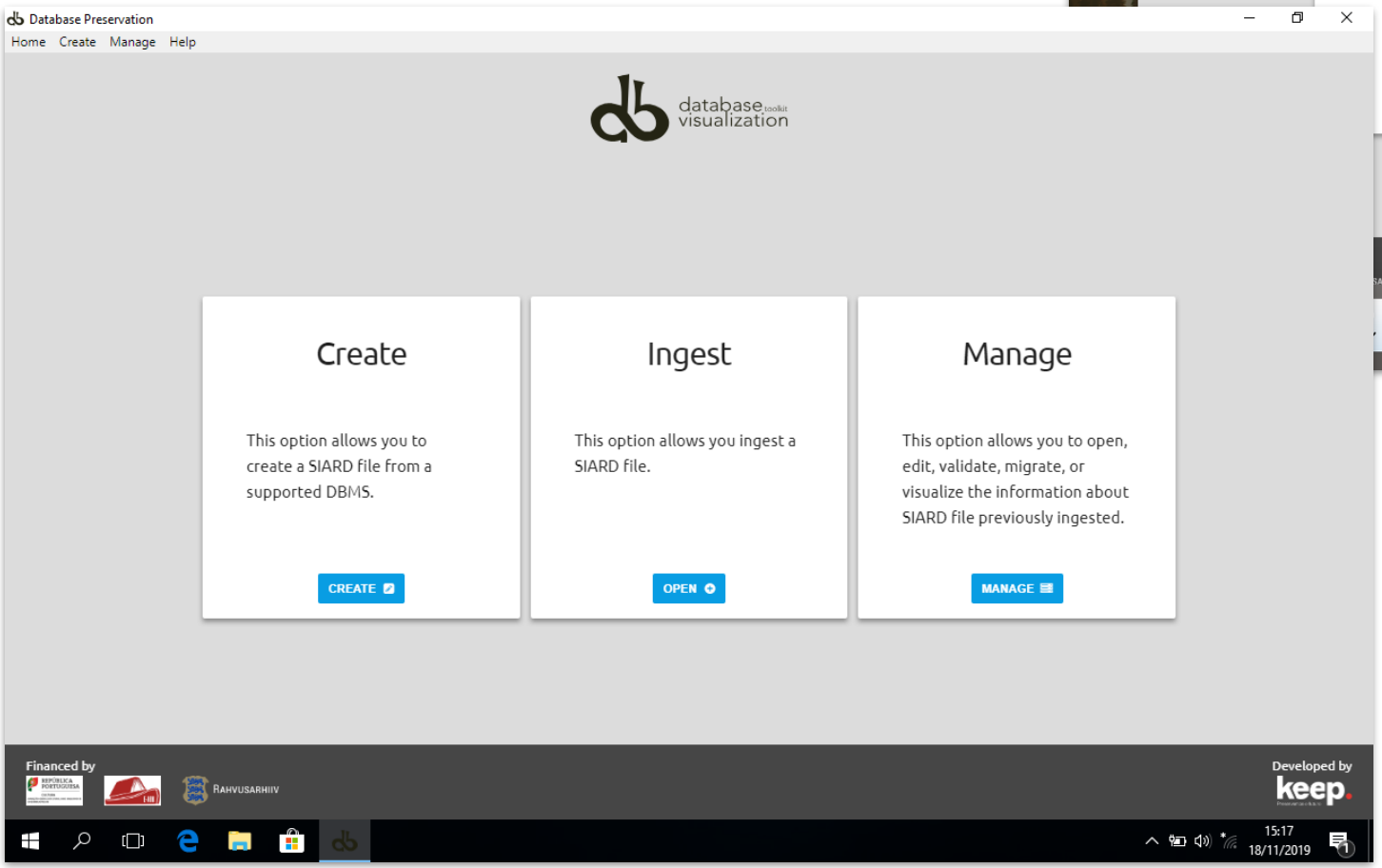
A command-line tool and development library for automation and system integration

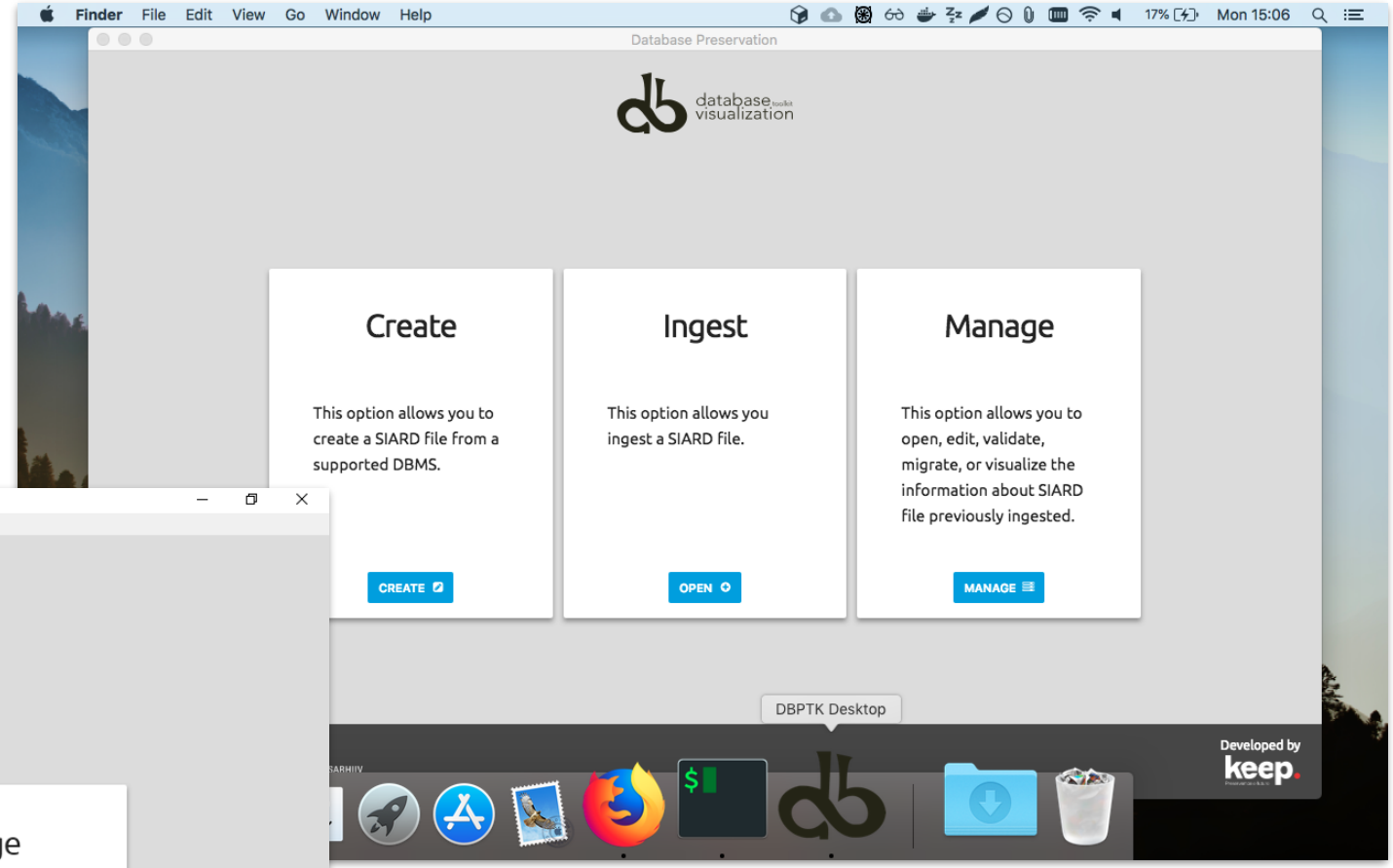
Icons from www.flaticon.com

DBPTK Desktop

Desktop Application

Windows





macOS

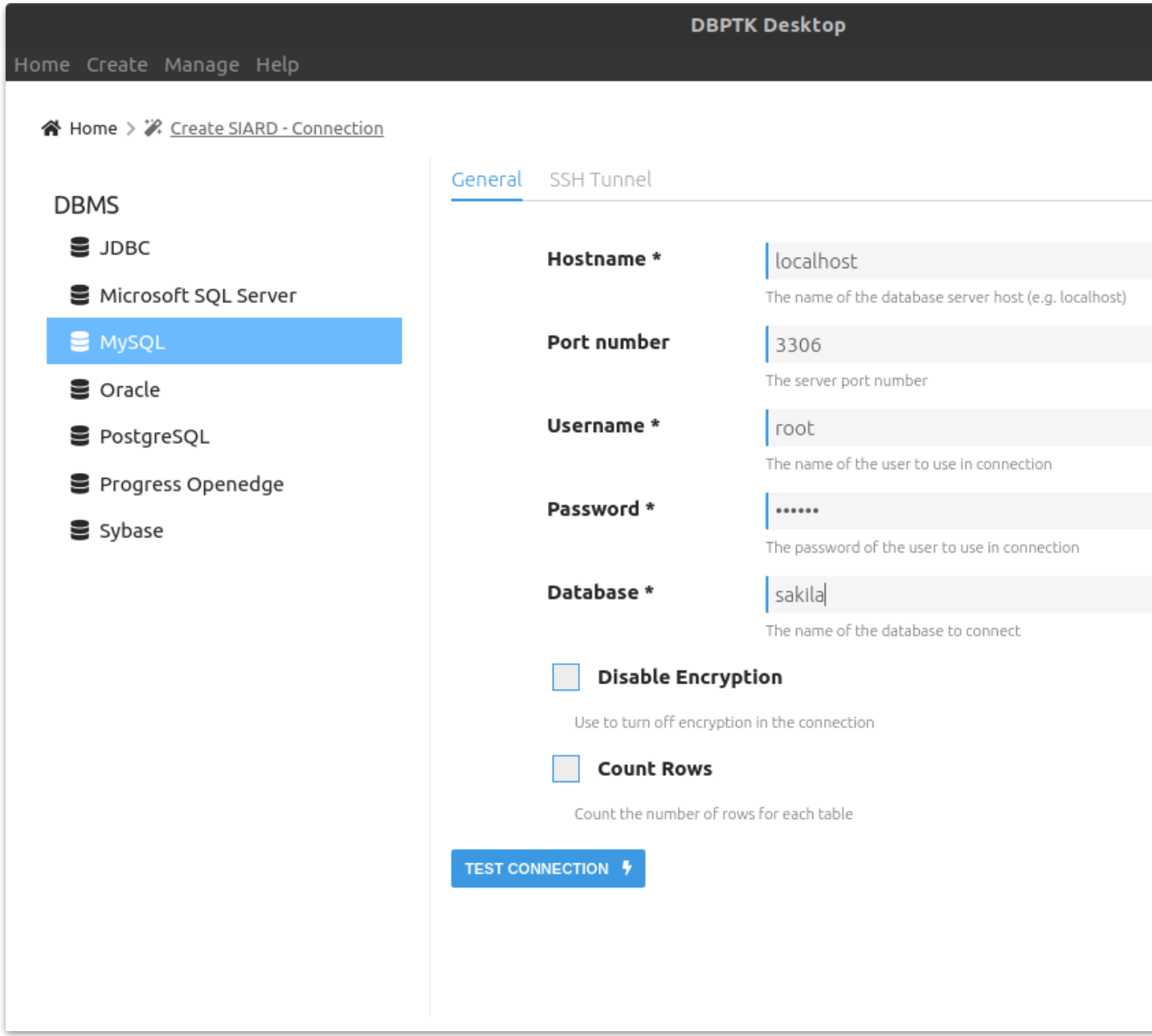
Also available on Linux

European Commission

SIARD creation

Export database to a preservation format

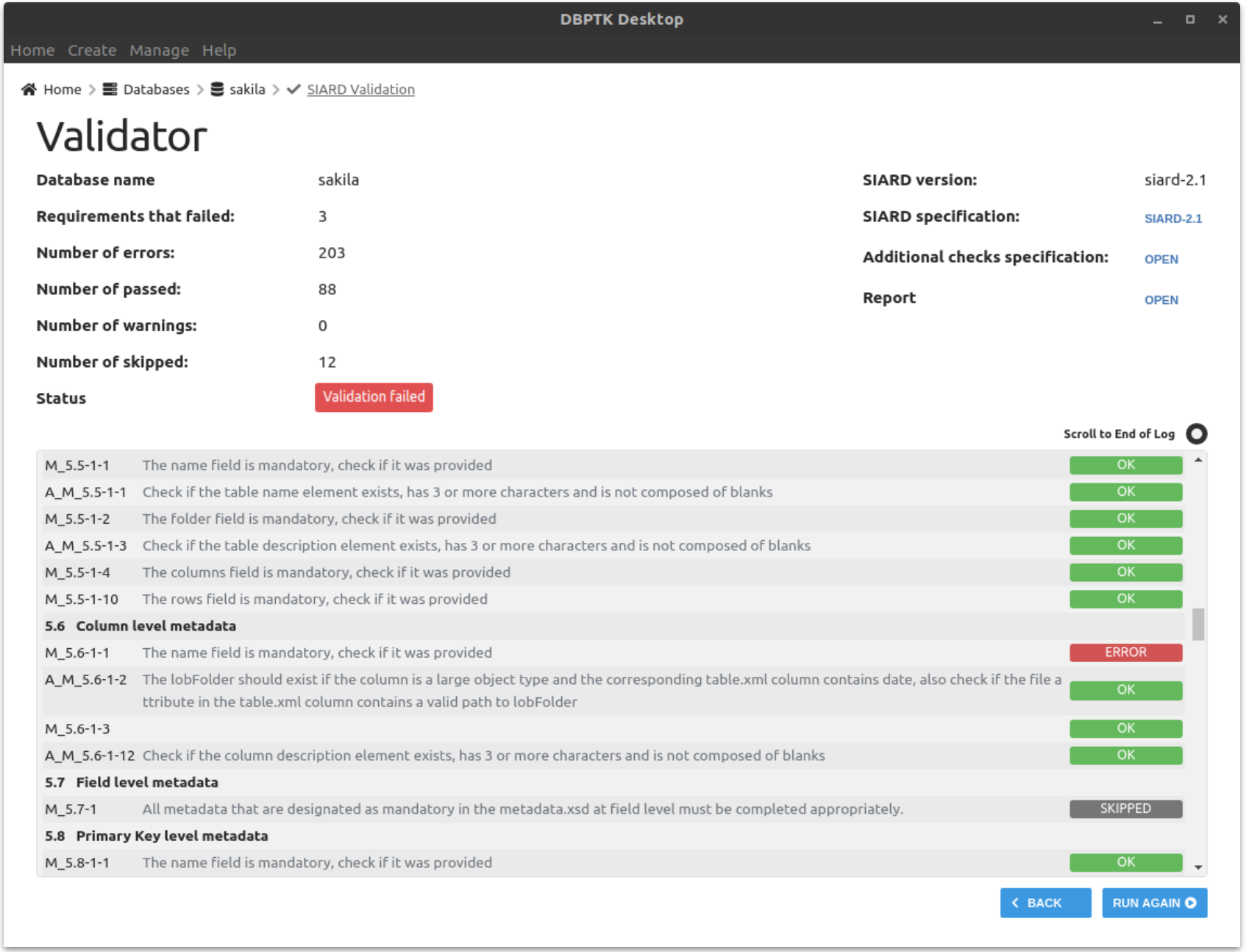
Connect to a local or remote database and save all content into a preservation format like SIARD



SIARD validation

Validate archived database

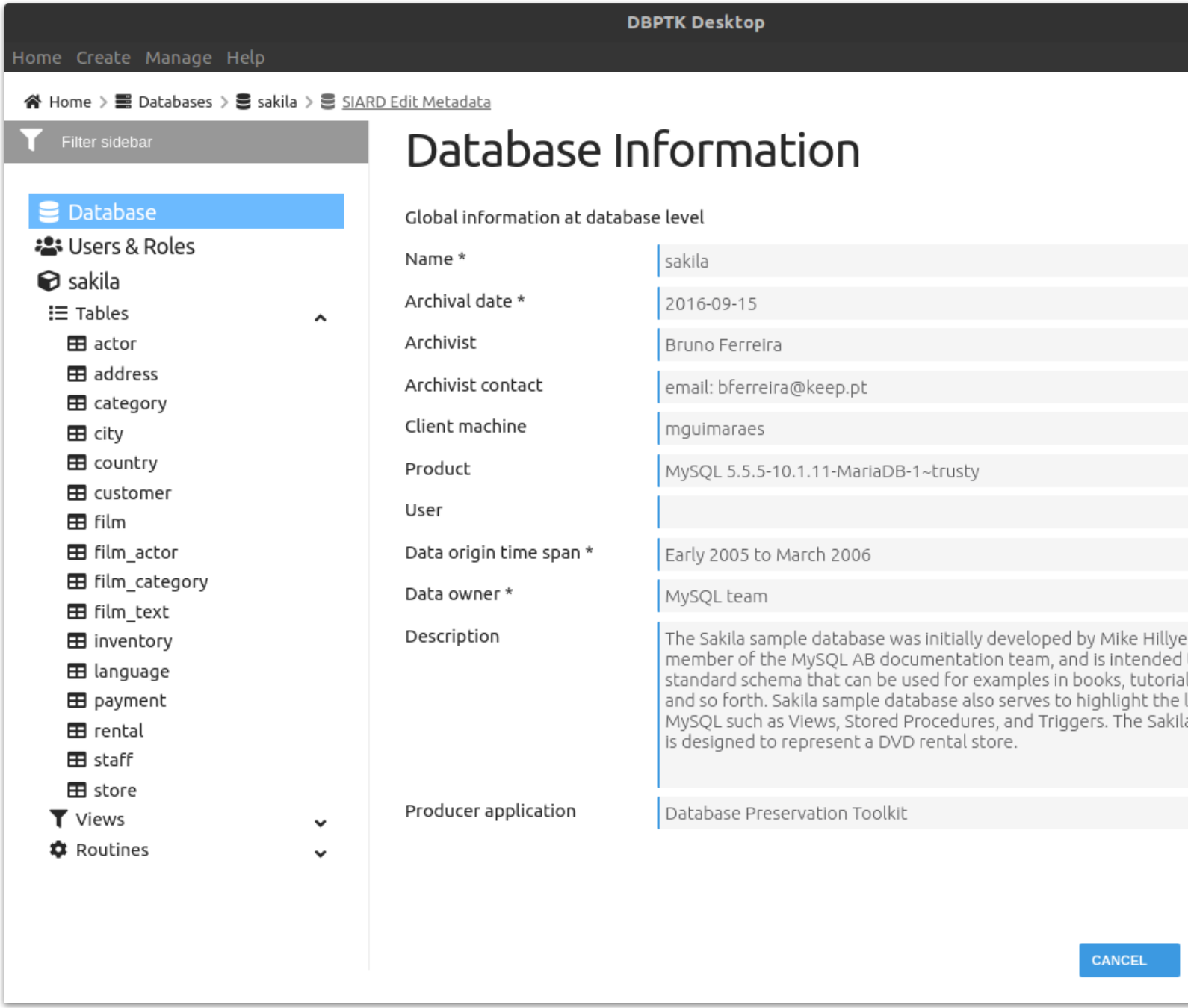
Validate SIARD against specification plus many additional checks for a thorough validation



Edit SIARD metadata

Enrich archived database with descriptions

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

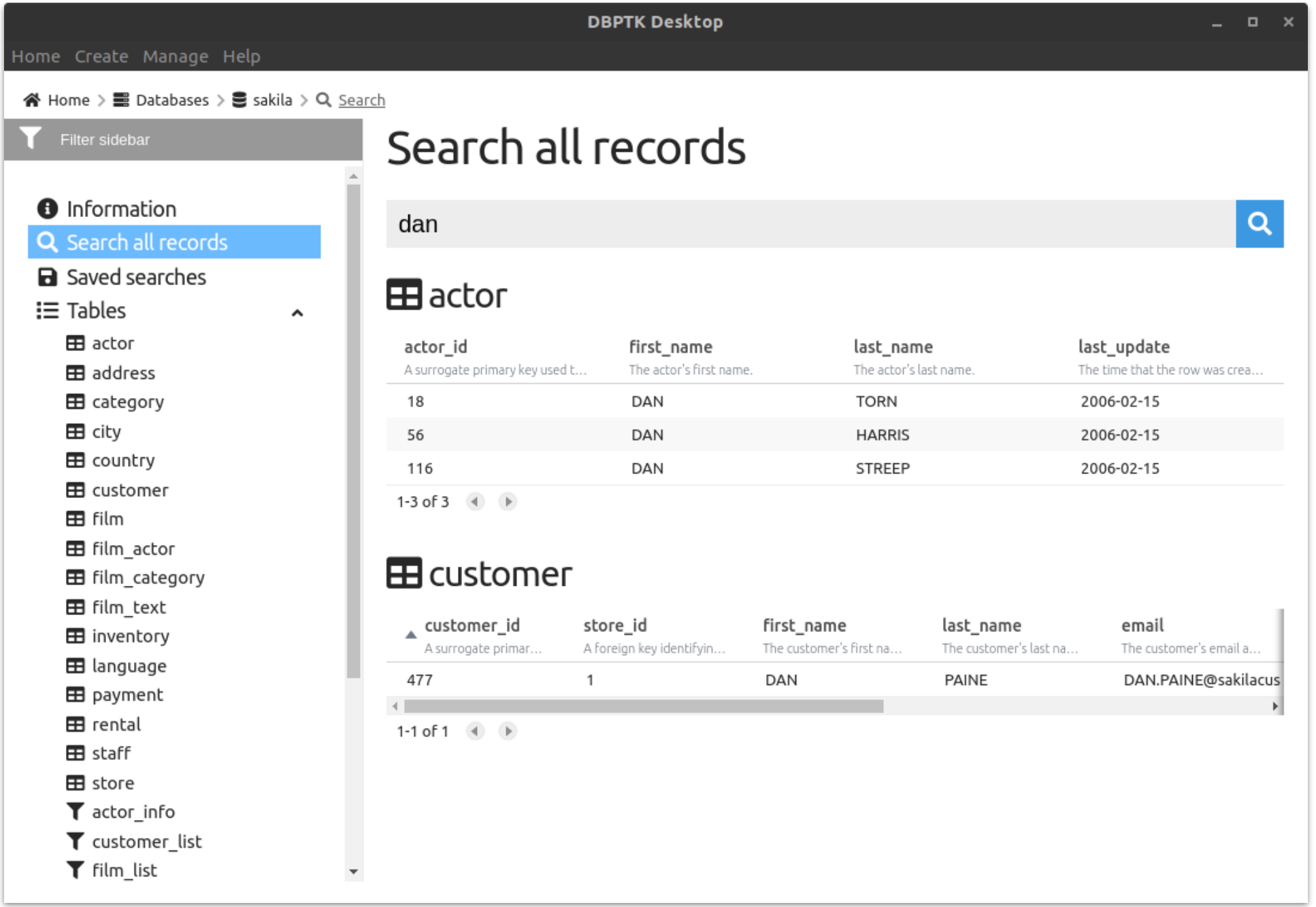


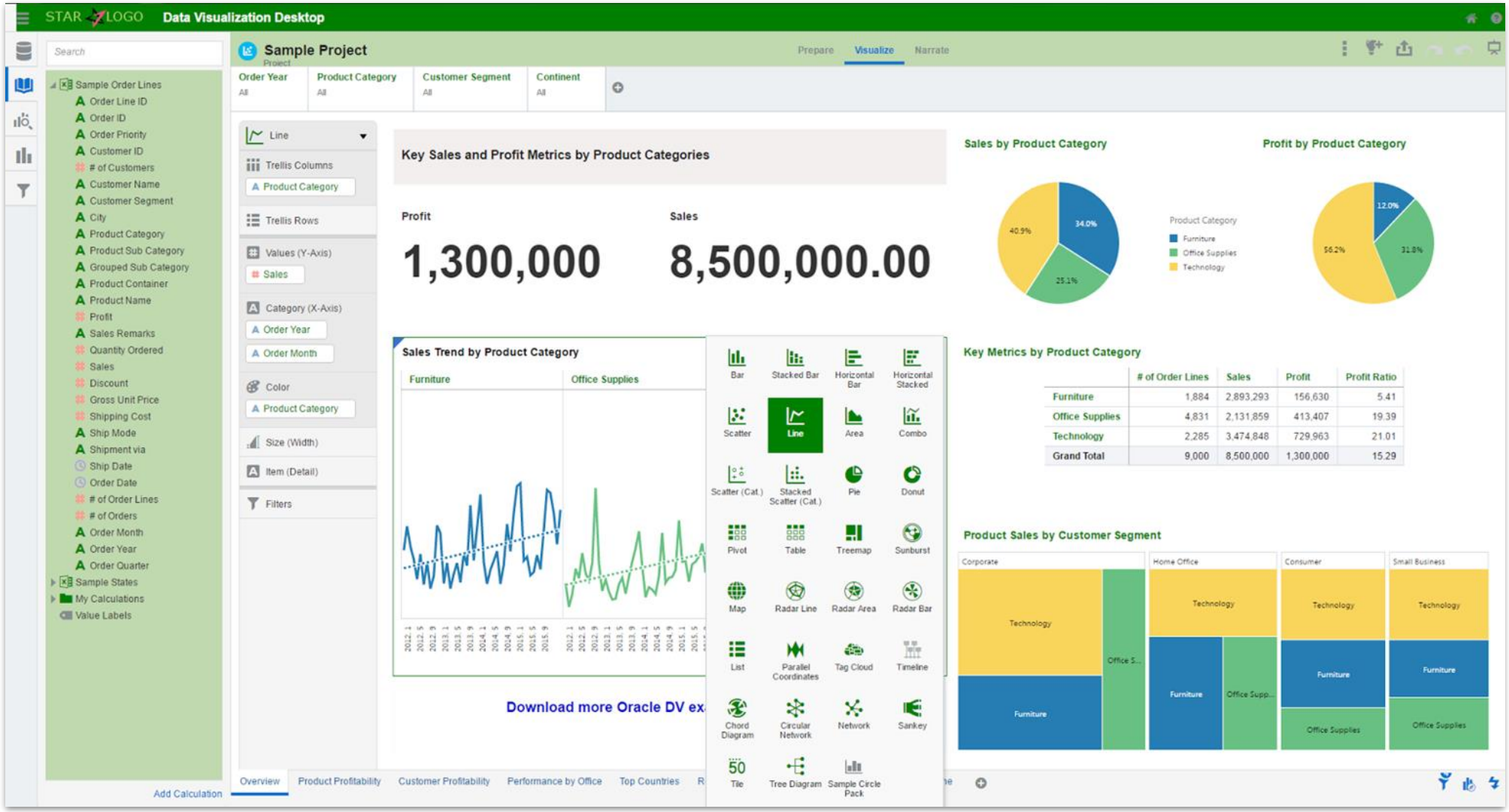
Search records

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.





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

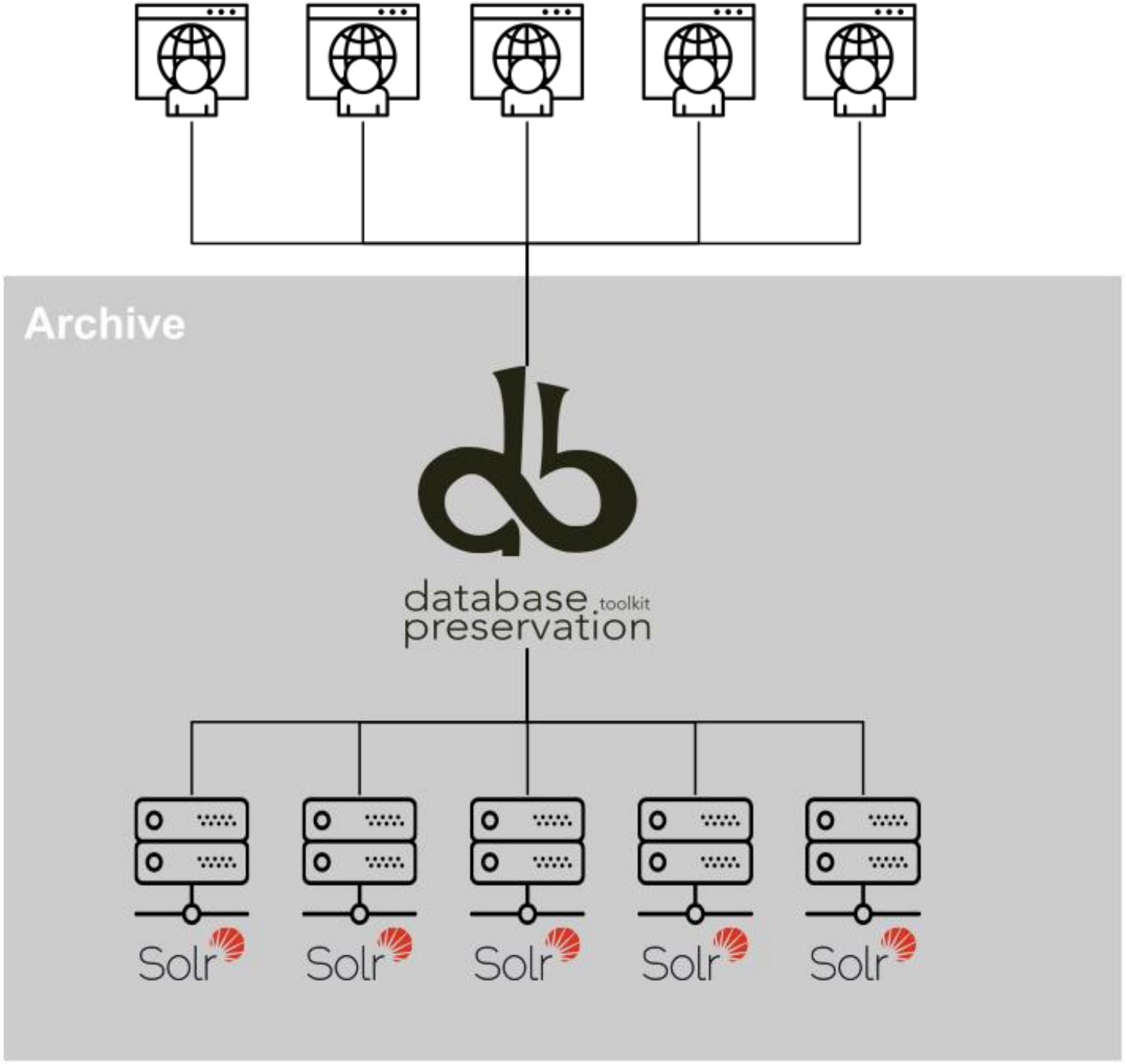
Image from Oracle Blog

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

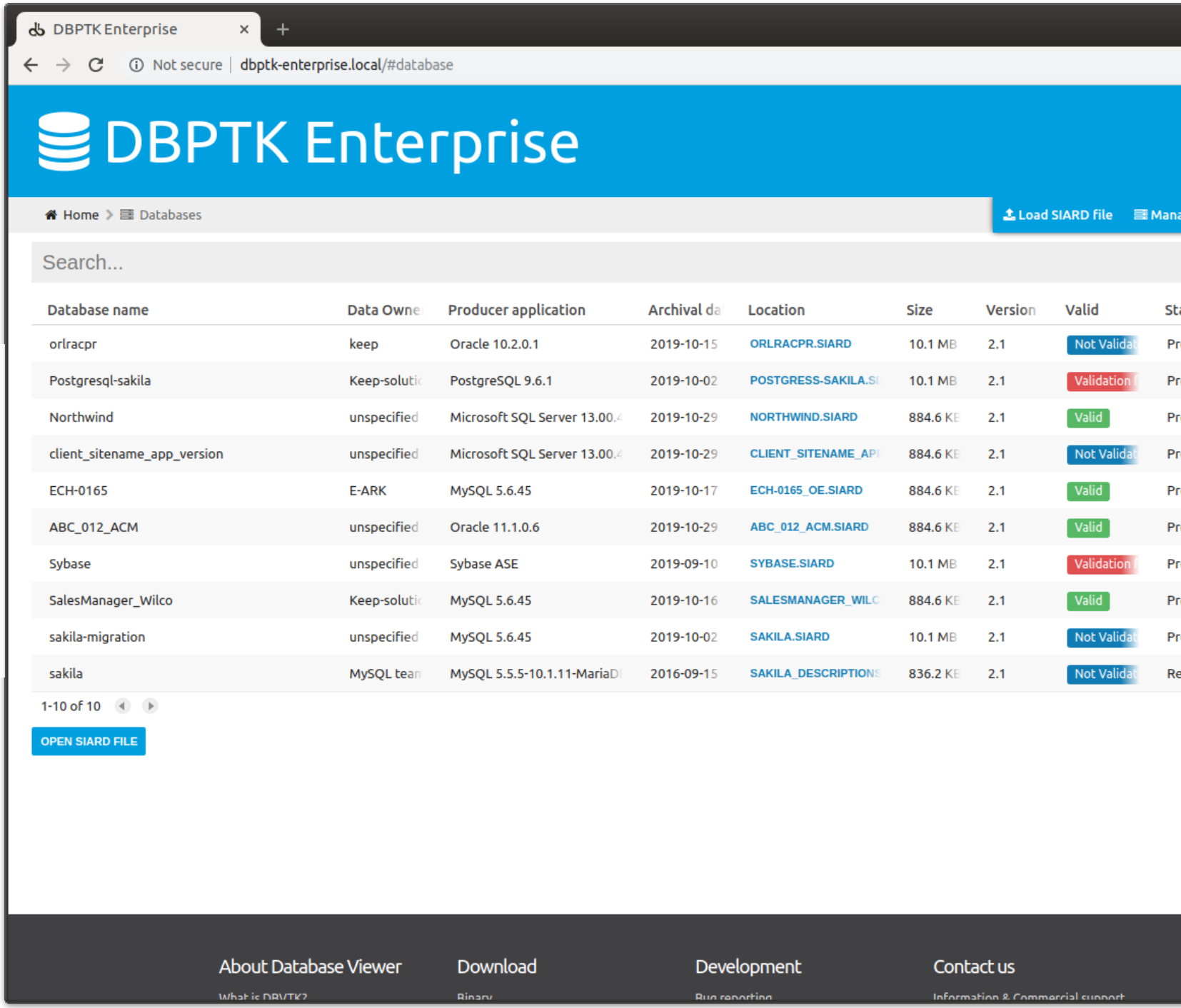


Icons from www.flaticon.com

Manage multiple databases

Single system, multiple databases

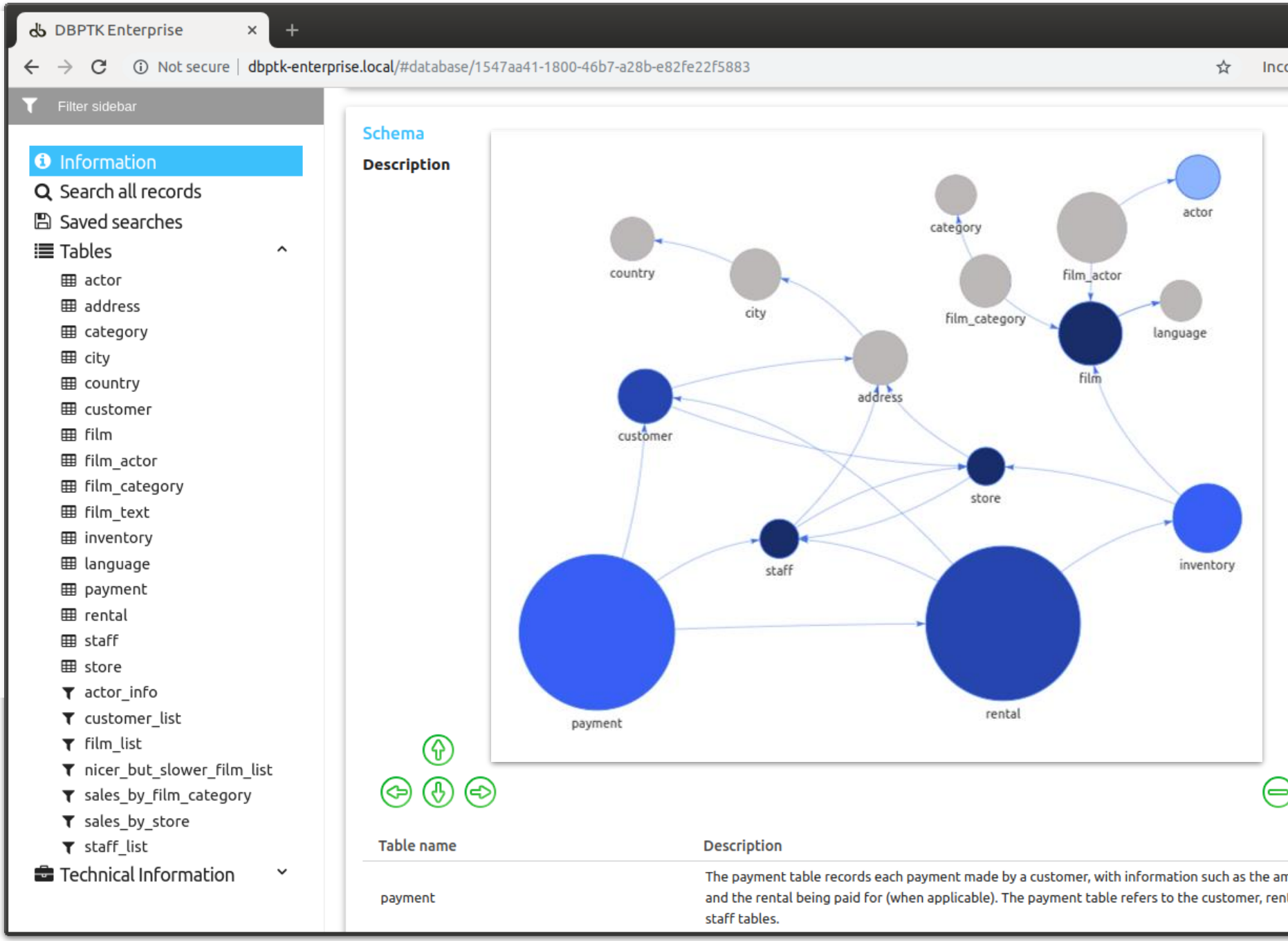
Search through the databases, manage their status, enrich their metadata, validate them, make them ready for users to search.



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

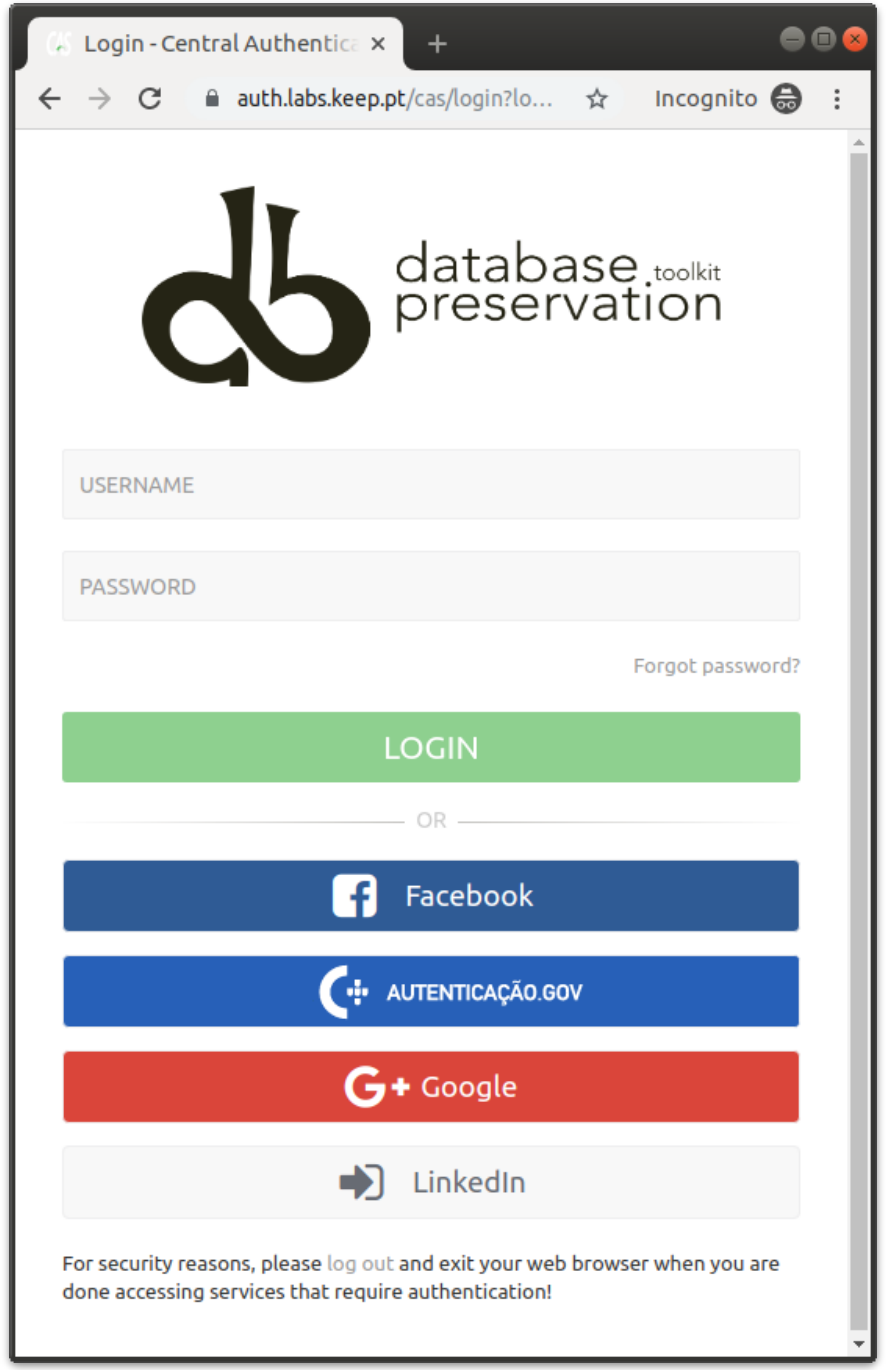


Single sign-on

Support for multiple protocols

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

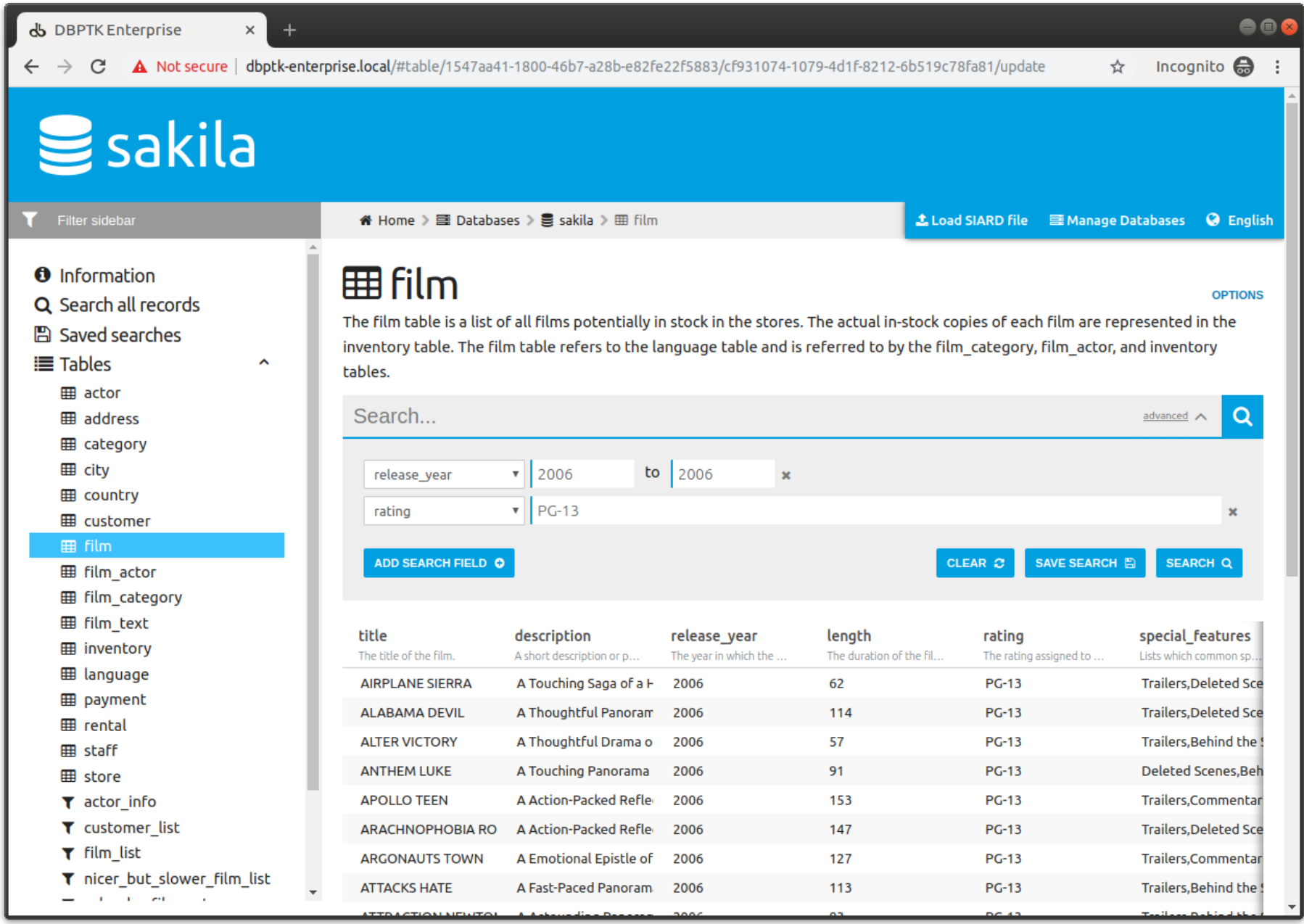
Supports internal authorization definition or configurable external authorization



Browse and search

Allow users to access database content on the Web

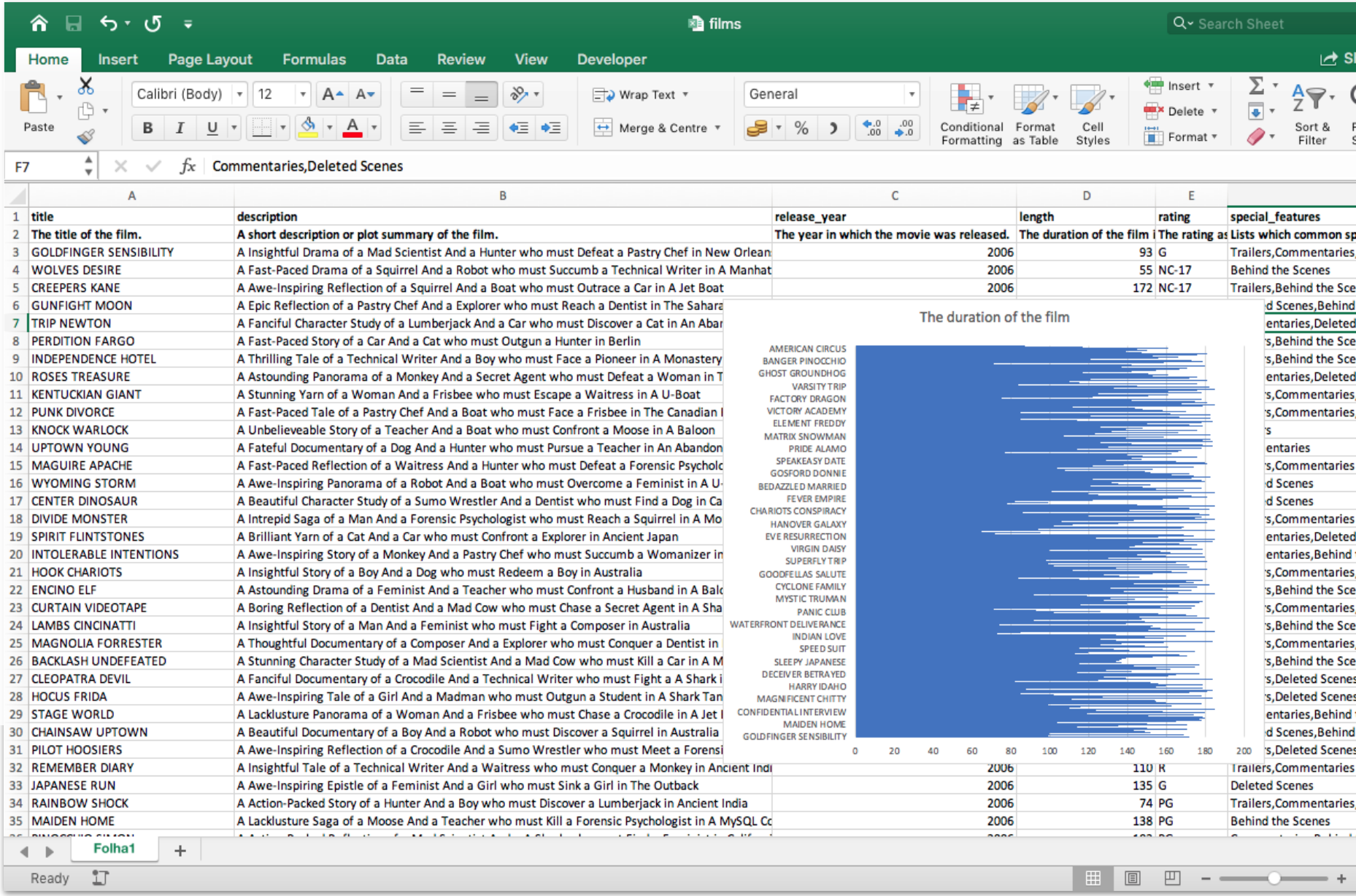
Allow them to search on a prepared, de-normalized and anonymized database content.



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



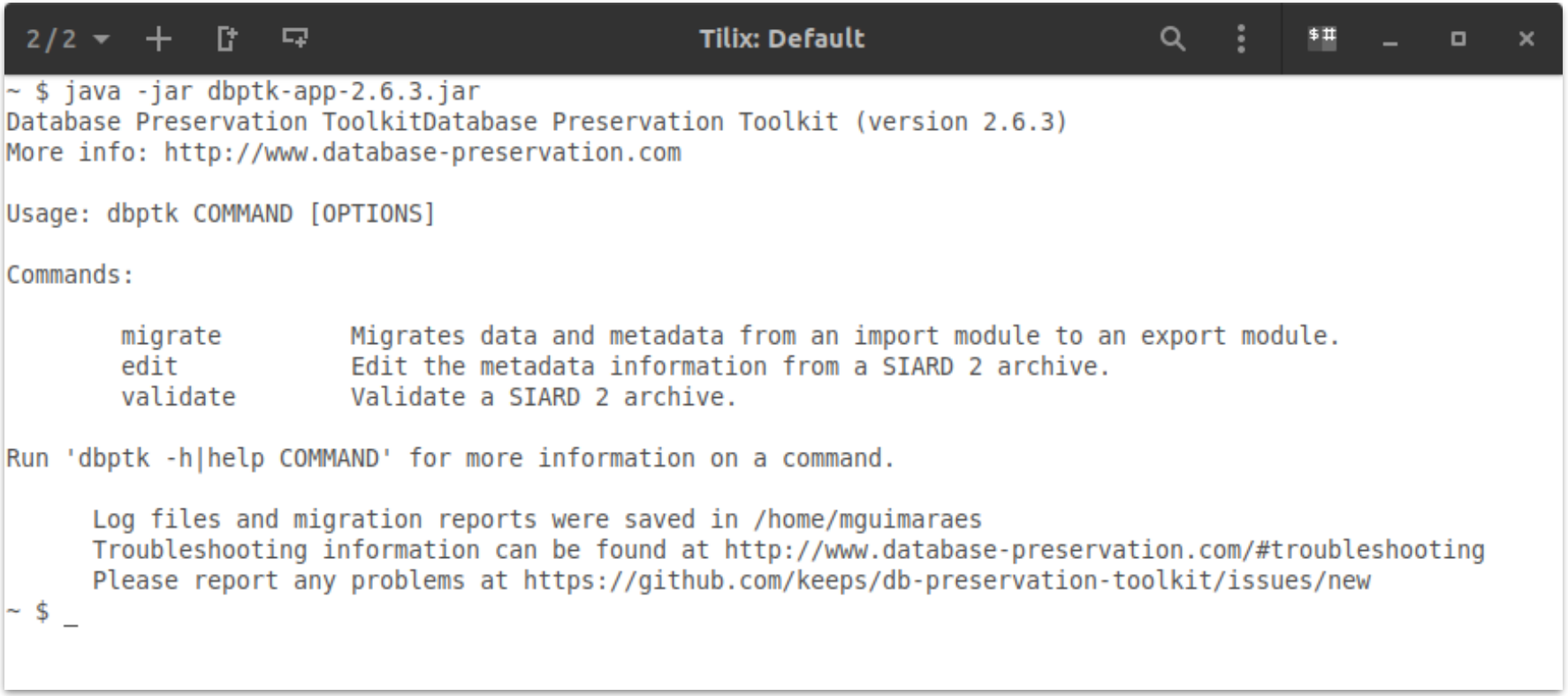
	A	B	C	D	E	
1	title	description	release_year	length	rating	special_features
2	The title of the film.	A short description or plot summary of the film.	The year in which the movie was released.	The duration of the film i	The rating as	Lists which common sp
3	GOLDFINGER SENSIBILITY	A Insightful Drama of a Mad Scientist And a Hunter who must Defeat a Pastry Chef in New Orlean	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				s,Behind the Sce
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				s,Commentaries
13	KNOCK WARLOCK	A Unbelieveable 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 Psycholo				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 CINCINATTI	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 Lacklusture Panorama of a Woman And a Frisbee who must Chase a Crocodile in A Jet				entaries,Behind
30	CHAINSAW UPTOWN	A Beautiful Documentary of a Boy And a Robot 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 Lacklusture 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

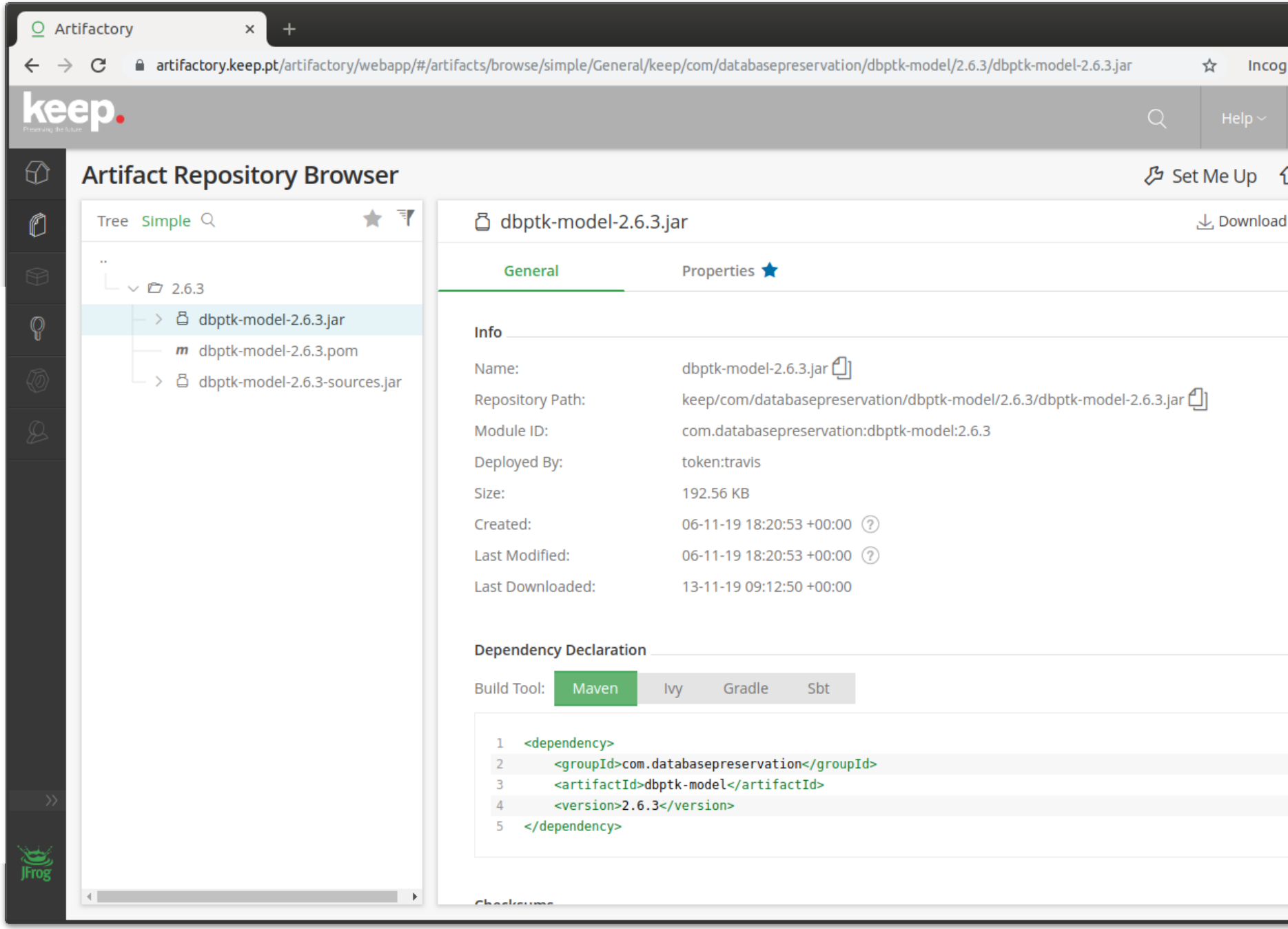
Command line interface allows easy automation of periodic tasks like saving database to preservation format, validating, and editing metadata.



Systems integration

Java library

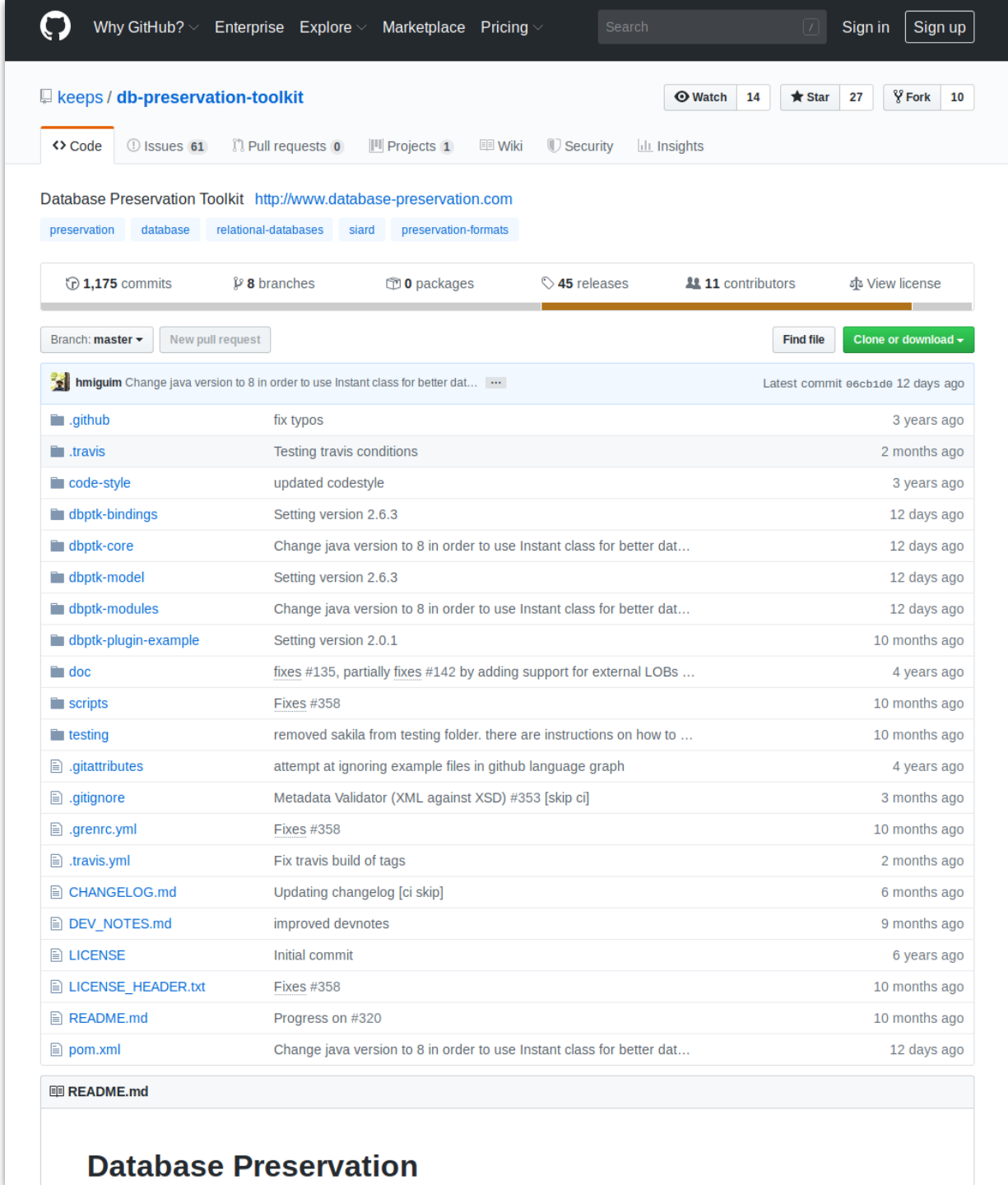
Library to allow integration of production systems to directly use database preservation features.



Open source

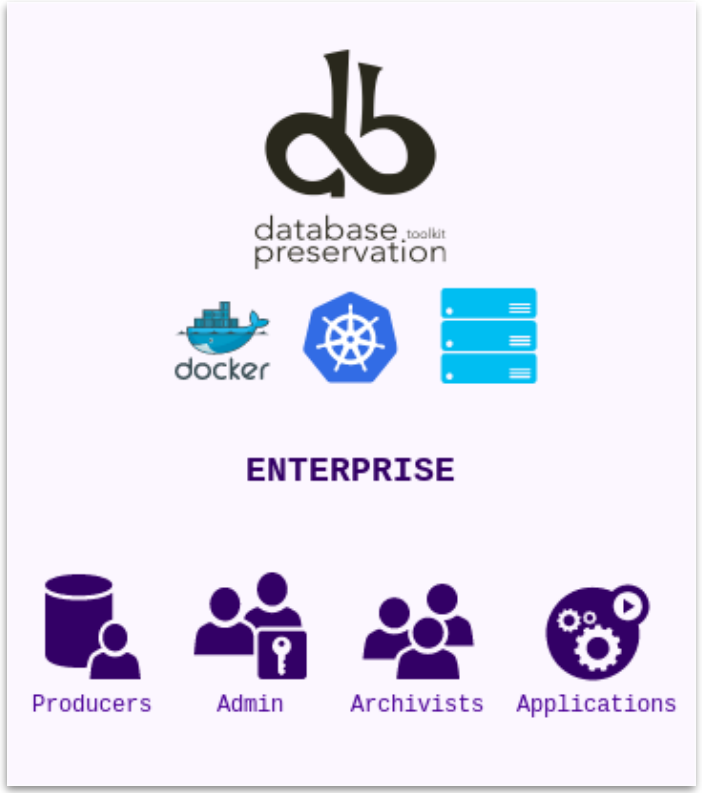
For custom development

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



How to choose?


database preservation toolkit
DESKTOP
Producers
Archivists


database preservation toolkit
docker
ENTERPRISE
Producers
Admin
Archivists
Applications

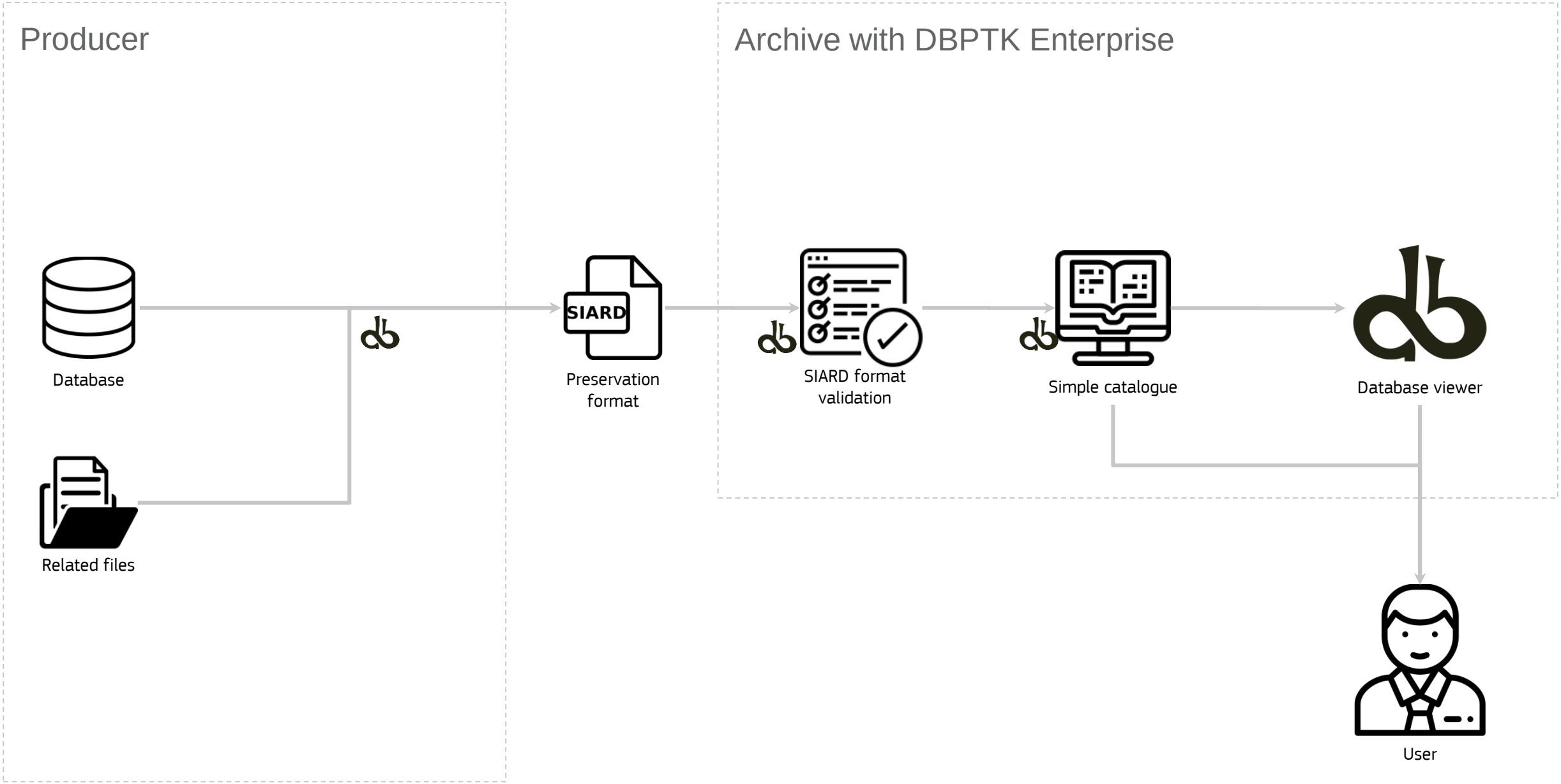

database preservation toolkit
DEVELOPER
Developers
Applications
Producers

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

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

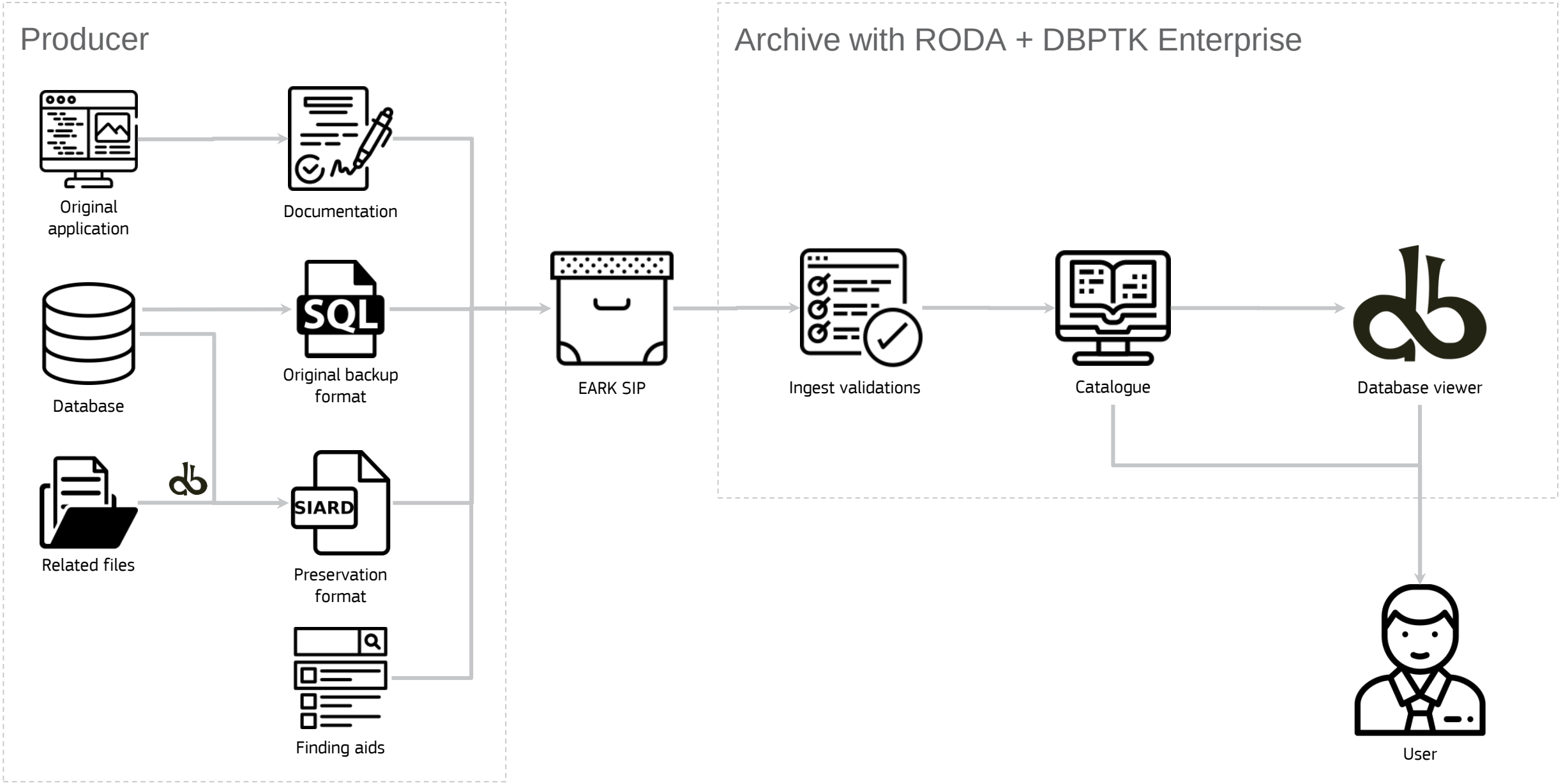
** Limited functionally, may need very high requirements for a Desktop

Simple database archival flow



Icons from www.flaticon.com

Full database archival ecosystem flow



Icons from www.flaticon.com

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

Coffee break

We will resume at 15:15