European Commission

Welcome to this live webinar on Distributed Digital Preservation in practice

Start 10:00

18 May 2023

Audience notes for the Live Webinar

Your **cameras have been turned off** and **microphones muted**.

If you have any technical issues during the event, please use the chat function.

Please **use the Q&A for questions to speakers**. These will be addressed at the end of the event.

Please note that this webinar is recorded. No attendee personal information will be captured in these recordings. You will receive a recording link directly from WebEx following the event

Agenda

10:00 – 10:10

eArchiving Initiaitve welcome

Jaime Kaminski – eArchiving Initiative training activity lead

10:10 – 10:55

Distributed Digital Preservation in practice

Luís Faria – KEEP SOLUTIONS

10:55 – 11:00

Short Q&A / break

11:00 – 10:50

Demo

Miguel Guimarães– KEEP SOLUTIONS

10:55 – 11:00

Q&A / close

Distributed Digital Preservation in practice

Luís Faria and Miguel Guimarães
KEEP SOLUTIONS

eArchiving Initiative training webinar

Agenda

Distributed digital preservation Problem, mission, approach and vision Architecture Shallow E-ARK IPs (why, what and how) RODA agent and Synchronization Remote actions Available actions and external plugins	9:05 to 9:50
Break (10 minutes)	9:55 to 10:00
Live demonstration Setup your own network Synchronize with central and inspect content Request preservation actions and get back the result See how other information is provided from central to agents	10:00 to 10:50
Discussion and Q&A	10:50 to 11:00

Digital Preservation

The sum of **activities** (procedures, standards, best practices and technologies) necessary to ensure the **long-term access and reusability** of digital information.

MIME

EAD

Migration

WARC

JPEG

TIFF

OPF

METS

OAIS

RODA

AIP

DC

PREMIS

SIP

JHove

ISO

Metadata

PAIMAS

Authentication

NDSA

Refreshing

Digitisation

PDF/A

DCC

MPEG

DPC

DDI

DRAMBORA

Authenticity

Formats

SGML

PRONOM

DIP

XML

CCSDS

Checksum

DOI

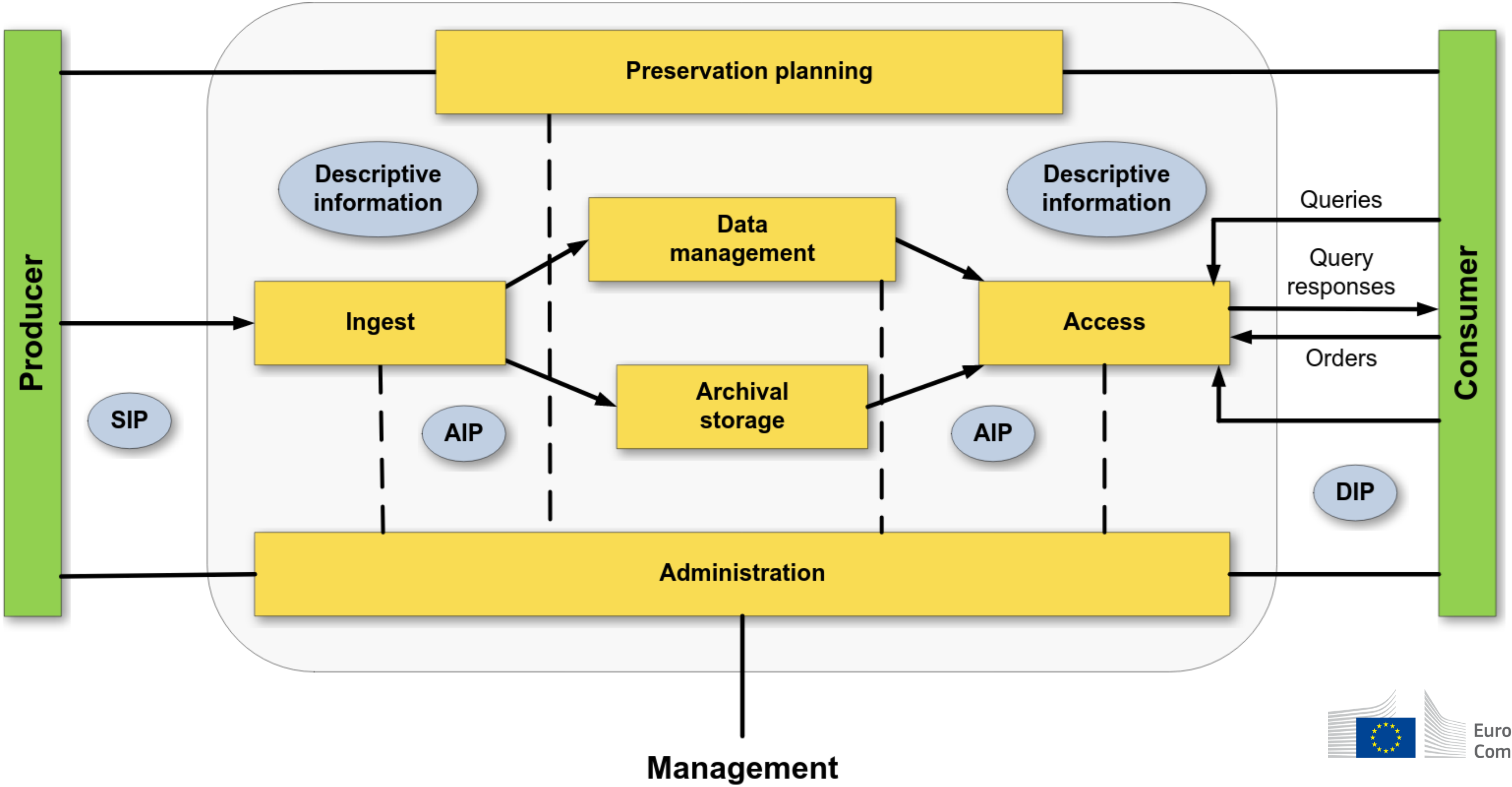
DROID

Emulation

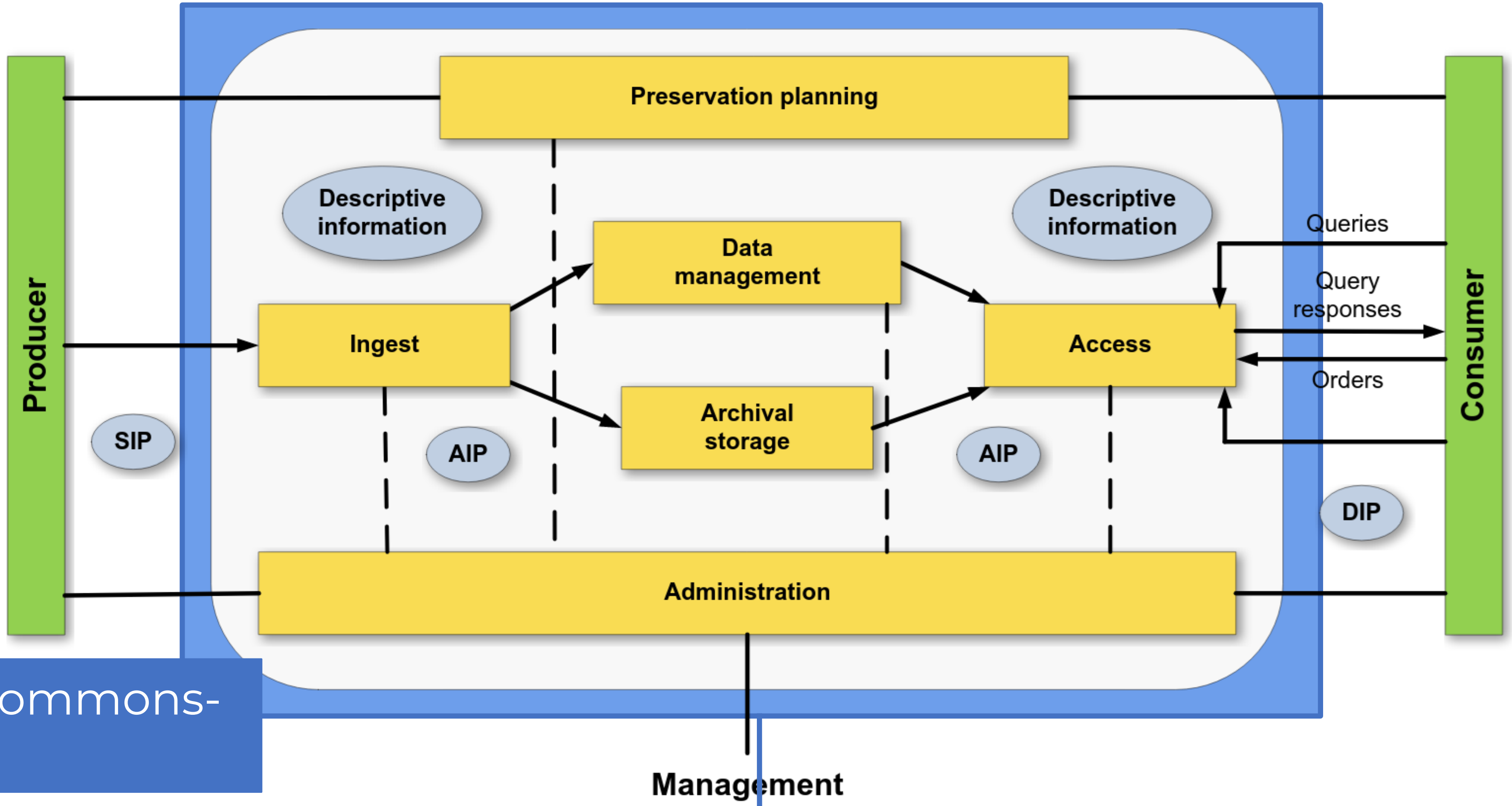
TRAC

HTML

ISO 14721:2012 (OAIS): Functional model



European Commission



E-ARK commons-ip

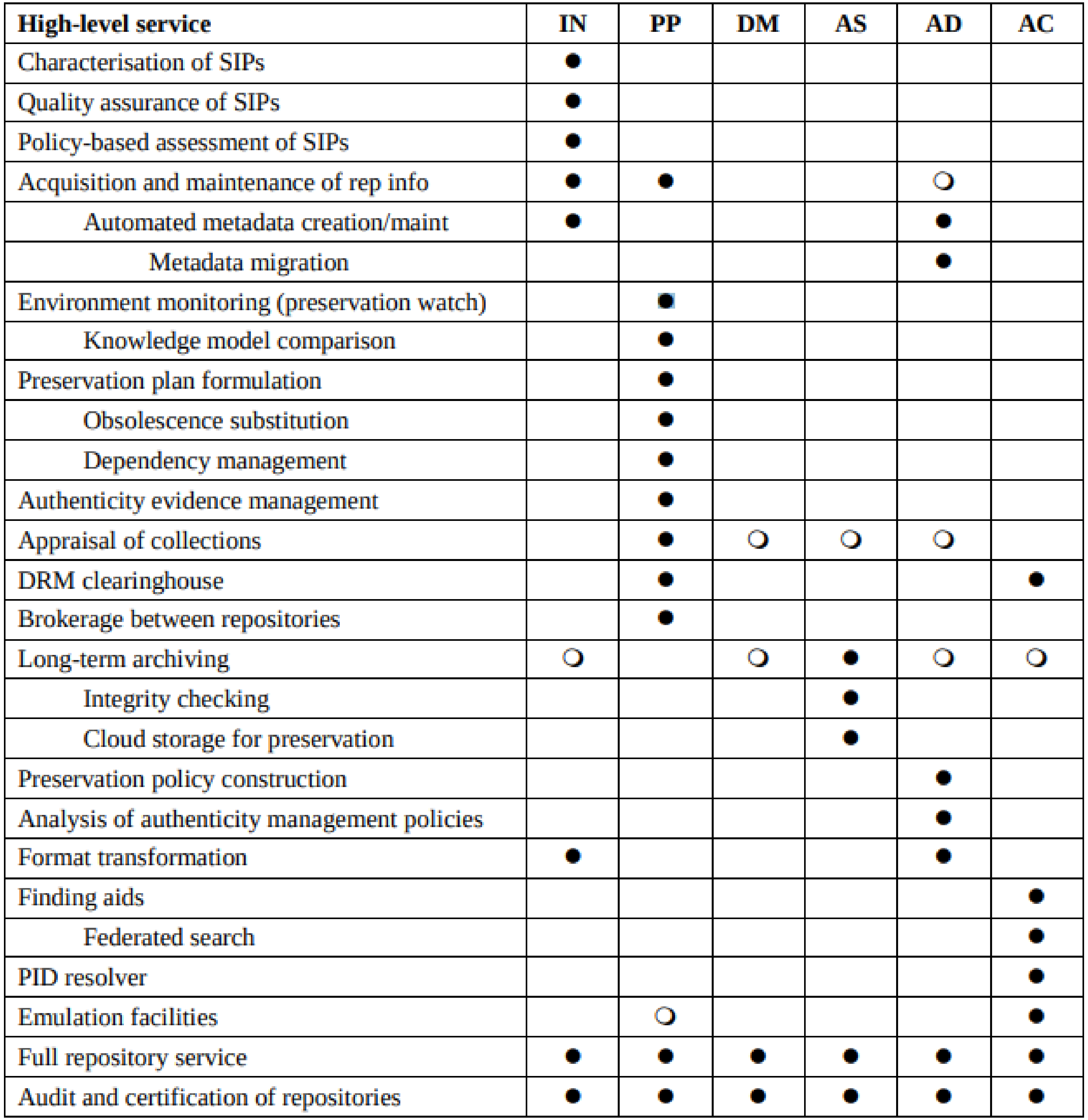
RODA
SIP creation tool

Preservation Toolkit

RODA

Preservation Toolkit

European Commission



High-level service	IN	PP	DM	AS	AD	AC
Characterisation of SIPs	●					
Quality assurance of SIPs	●					
Policy-based assessment of SIPs	●					
Acquisition and maintenance of rep info	●	●			○	
Automated metadata creation/maint	●				●	
Metadata migration					●	
Environment monitoring (preservation watch)		●				
Knowledge model comparison		●				
Preservation plan formulation		●				
Obsolescence substitution		●				
Dependency management		●				
Authenticity evidence management		●				
Appraisal of collections		●	○	○	○	
DRM clearinghouse		●				●
Brokerage between repositories		●				
Long-term archiving	○		○	●	○	○
Integrity checking				●		
Cloud storage for preservation				●		
Preservation policy construction					●	
Analysis of authenticity management policies					●	
Format transformation	●				●	
Finding aids						●
Federated search						●
PID resolver						●
Emulation facilities		○				●
Full repository service	●	●	●	●	●	●
Audit and certification of repositories	●	●	●	●	●	●

APARSEN

D21.1 Overview of Preservation Services

The table represents the structure of preservation services developed according to the above principles.

Black circles indicate where a service is a key contributor to the corresponding OAIS functional entity; while circles indicate possible or marginal relevance.

http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D21_1-01-2_1_incURN.pdf

Institution staff profiles

Organisation manager

Financial manager

Project Manager

Information manager and operators

Information Technology manager and operators: hardware and software

Digital Preservation Manager

Data Governance Manager / Information Security Officer / others.

Digital Preservation Manager

Preservation Policy

Preservation Planning

Representation Information

Risk management oriented to digital preservation (long-term access and reusability)

Preservation Actions: diagnose, identify risks, mitigate, improve value

Technology Watch

Designated Community Watch

Audit and certification

Authenticity, Appraisal, DRM and IPR

Authenticity

Capability to prove (or vouch) that the digital object is according to the original.

Preserve authenticy

The **credibility** of the digital object authenticity is endowed by the **trustworthiness** of the digital **repository** and the **institution** that supports it.

This **trustworthiness** is a consequence of the **institution honourability and credibility** and is further improved on the repository by having **transparency** on the **mission**, **policies** and **procedures** in place for **digital preservation**, being **rigorous** on their application and being able to **prove**, based on **evidence**, that the defined **policies and procedures are correctly followed**.

Do smaller institutions (public or otherwise) have the necessary resources to properly plan and execute digital preservation?

Why not just transfer content to the National Archives?

Information Security

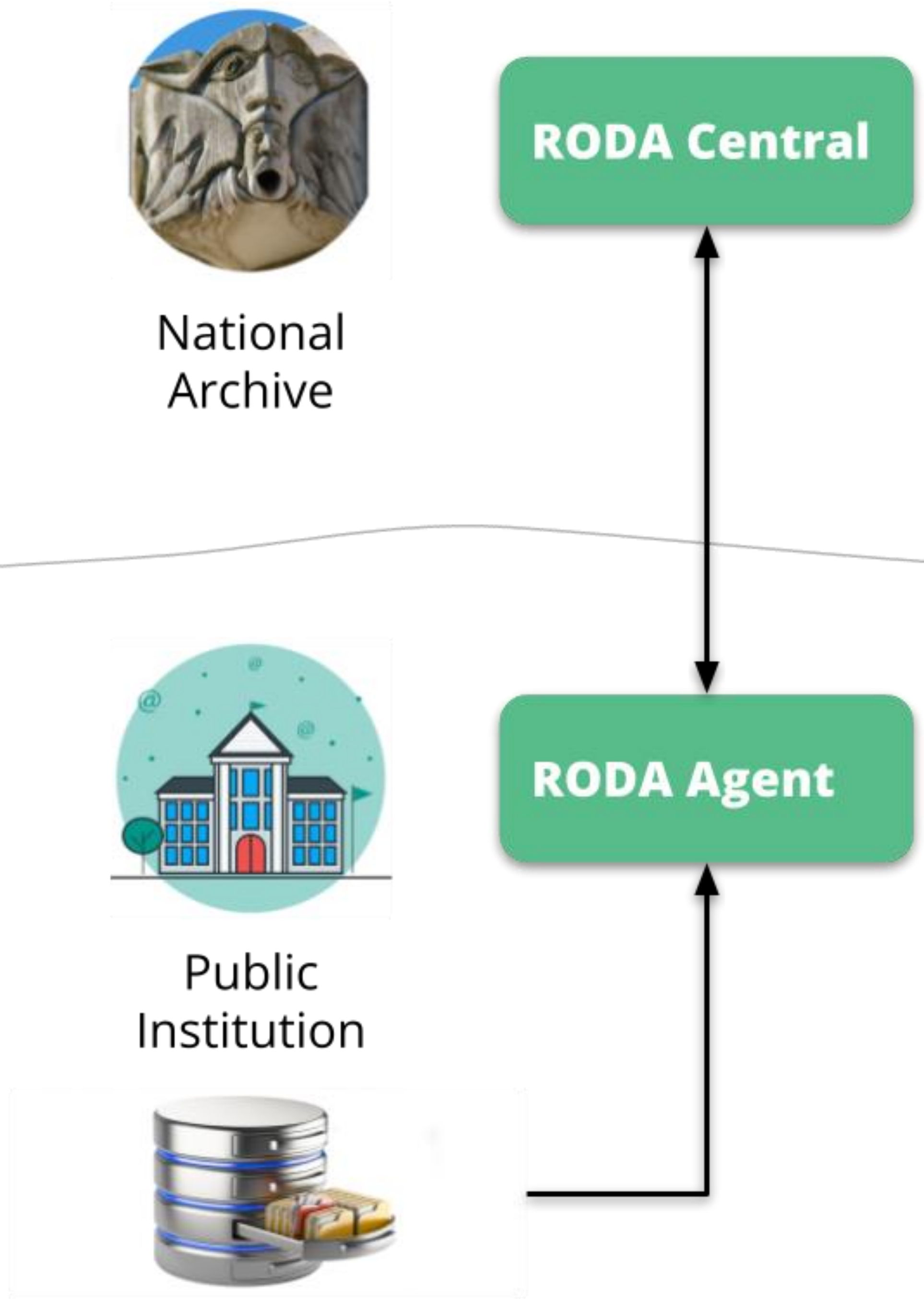
Local Access

Local Control

Continuous Production

Digital Preservation is too hard for smaller institutions

Can we keep the information, but delegate activities?



National Archive

RODA Central

- ✓ Preservation planning
- ✓ Risk management
- ✓ Technology watch
- ✓ Representation Network
- ✓ Diagnostic
- ✓ Preservation action recommendations
- ✓ Certification and audit

Public Institution

RODA Agent

- ✓ Ingest and data management
- ✓ Preservation action recommendation approval
- ✓ Automatic execution of preservation actions
- ✓ Continuous local access
- ✓ Complete data control

Distributed Digital Preservation

Set of functionalities that allow **delegating digital preservation functions** to a **central instance** of RODA, in order to create a digital preservation network where **preservation functions are defined centrally and distributed** among preservation agents installed in the local infrastructure of each participating institution.

–

Implementation of a **network** of institutions that **delegate the capacities for planning and executing digital preservation** to a central and authoritative entity.

Central institution

Institution with authority and capabilities to carry out digital preservation planning and operation functions, both within the institution itself and for other institutions that delegate these capabilities to the former.

RODA Central

RODA service managed by a central and authoritative institution with the capacity to carry out digital preservation planning and operation functions, both within the institution itself and for other institutions that delegate these capacities to the former.

Member institution

Institution with digital information that adheres to the distributed digital preservation service in order to delegate the capacities of planning and execution of digital preservation to the Central Institution.

RODA Agent

RODA service managed by a Member Institution that subscribes to the digital preservation service to a RODA Central, delegating the capacities of planning and execution of digital preservation to the Central Institution.

FUNCTION	LOCATION	NOTES
Ingest	Locally	The ingest must be done in the place where the data resides.
Data management	Metadata management should be carried out locally at each institution, but there is representation information that can be managed centrally.	The management of discovery services (supported by descriptive metadata) must be carried out locally in each participating institution (possibly using existing catalogs in the institutions). However, a very relevant part of this functional unit, such as the elaboration of a representation information database, this database can be carried out centrally.
Archival Storage	Locally	The storage and performance of integrity verification routines must be carried out where the data resides.
Access	Locally	The data exists to serve the institution, so the access component must be locally at the institution that holds the data.
Administration	Locally	Daily administration functions (e.g. user management) must be carried out locally on each member.
Preservation planning	Centrally	The activities inherent to preservation planning such as risk management, definition of preservation plans, technological surveillance and the like, definition of representation information, development and execution of preservation actions can be carried out centrally by a service provider with specialized knowledge in the area of preservation, such as the National Archive.

SERVICE	LOCATION
Characterization of SIPs	Locally
Quality assurance of SIPs	Locally
Policy-based assessment of SIP	Locally
Acquisition and maintenance of representation information	Centrally
Automated metadata creation/maintenance	Locally
Metadata migration	Locally
Environment monitoring (preservation watch)	Centrally
Knowledge model comparison	Centrally
Preservation plan formulation	Centrally
Authenticity evidence management	Locally
Appraisal of collections	Locally
DRM clearinghouse	Locally
Brokerage between repositories	Locally
Long-term archiving	Locally
Integrity checking	Locally
Cloud storage for preservation	Centrally*
Preservation policy construction	Centrally
Analysis of authenticity management policies	Locally
Format transformation	Locally (central decision)
Finding aids	Locally
Federated search	Centrally
PID resolver	Centrally**
Emulation facilities	Locally
Audit and certification of repositories	Centrally

* Central data replication option was considered unfavorable.

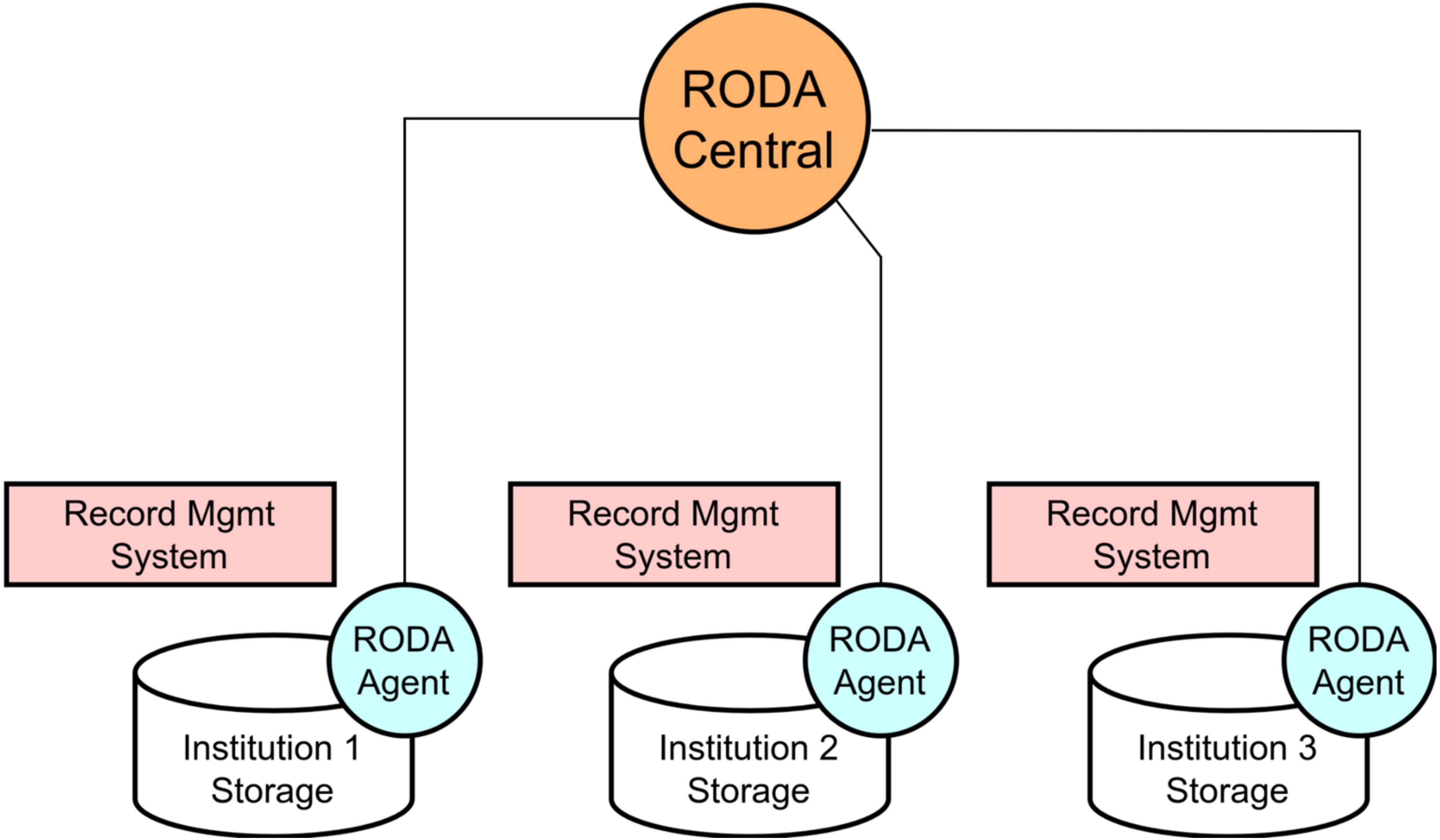
** The persistent identifier is decentralized (UUID), but the instance it belongs to (location) can be found centrally.

Architecture

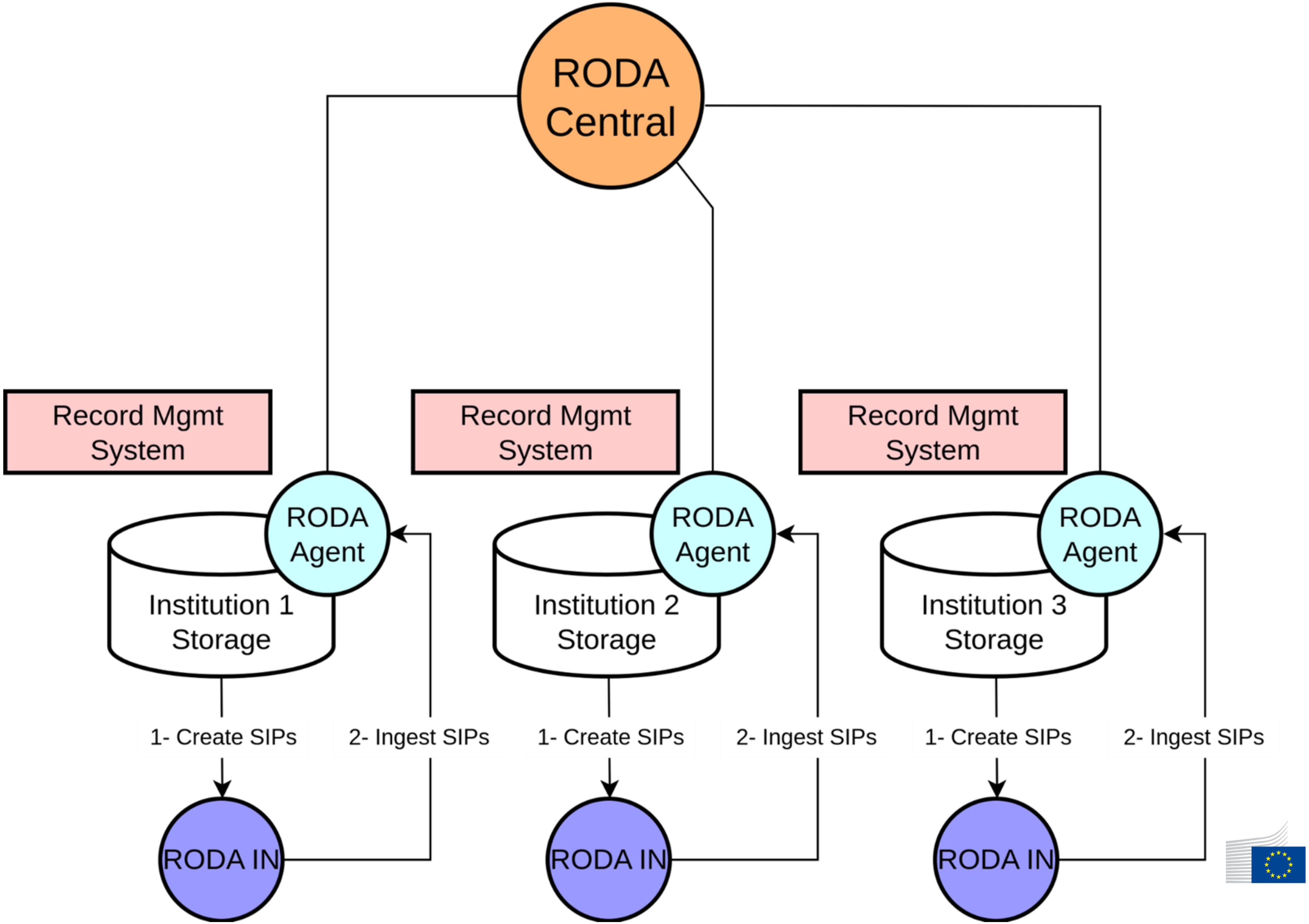
Components, formats and processes

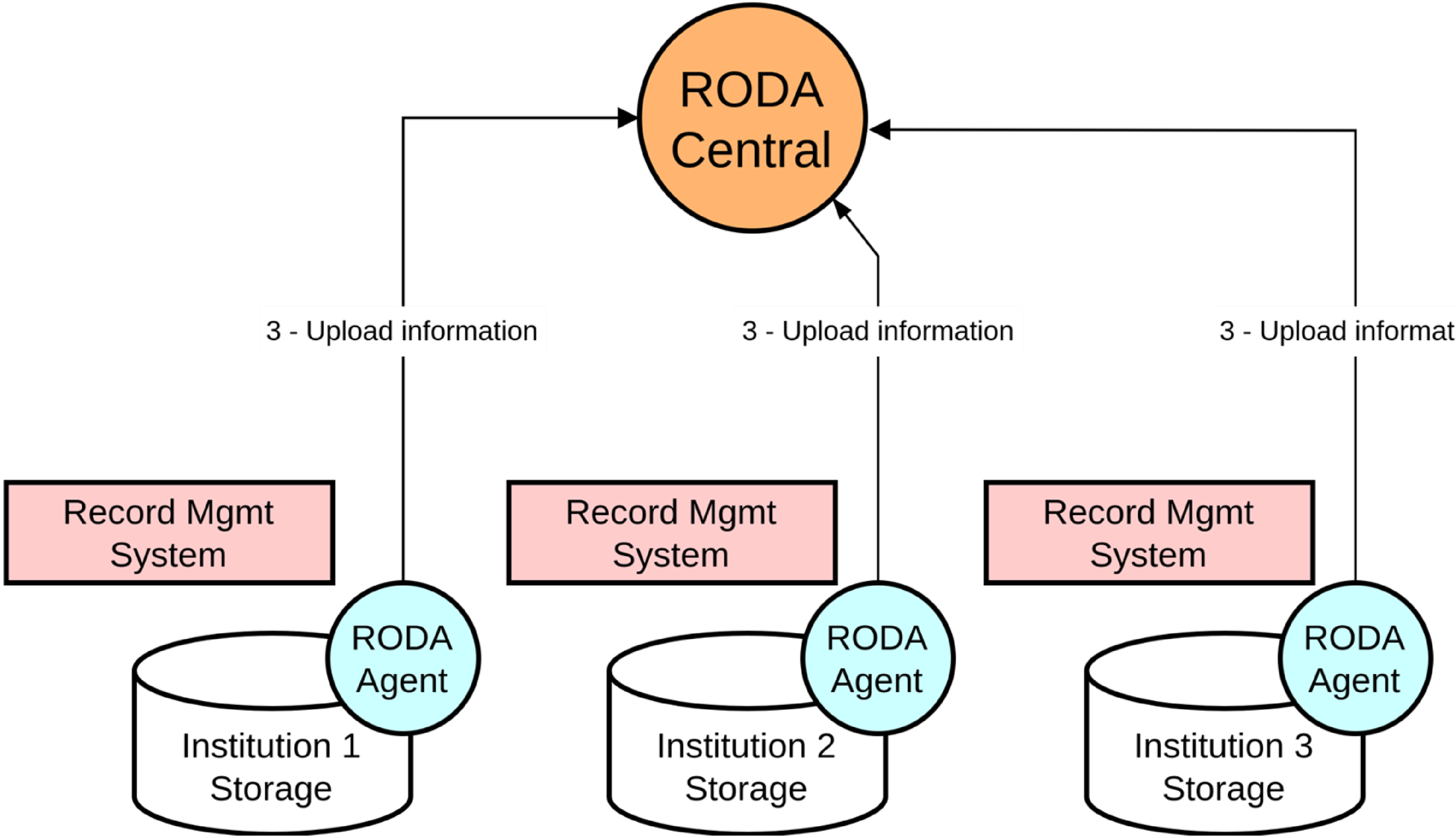
Workflow

1. Creation of SIPs in RODA agent
   Using RODA-in or custom integrations using commons-ip
2. Ingest of SIPs in the RODA agent
3. Upload information from RODA agent to RODA central
   Shallow AIPs, ingest and action processes, process reports, risk incidences, etc.
4. Preservation planning at the RODA central
5. RODA central requests execution of actions in RODA agent
   Diagnostic action or risk mitigation actions
6. Download information from the RODA central to the RODA local
   Action requests, risks, representation information
7. RODA agent execution of actions



European Commission

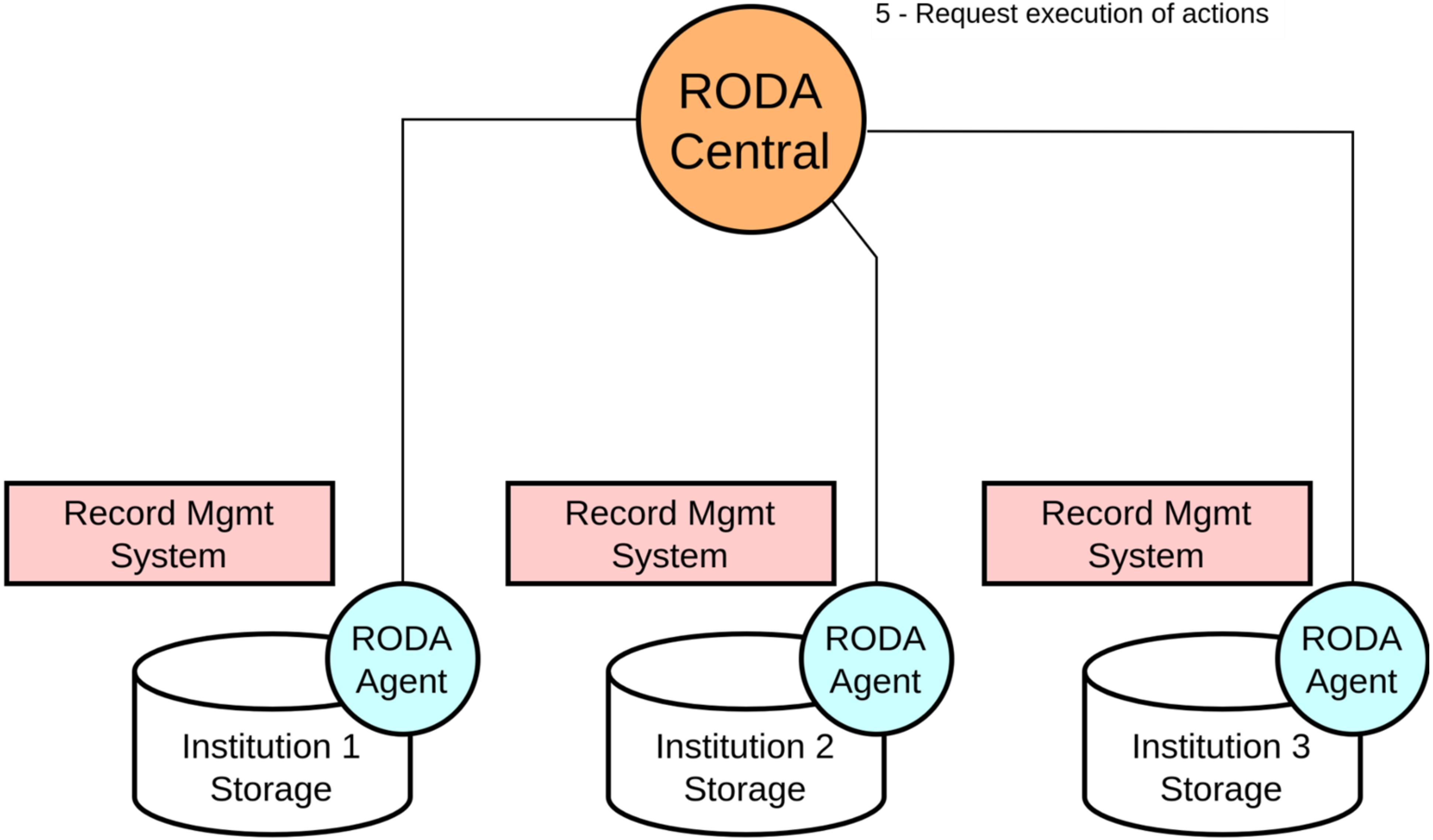

RODA Central
Record Mgmt System
Record Mgmt System
Record Mgmt System
RODA Agent
RODA Agent
RODA Agent
Institution 1 Storage
Institution 2 Storage
Institution 3 Storage
1- Create SIPs
2- Ingest SIPs
1- Create SIPs
2- Ingest SIPs
1- Create SIPs
2- Ingest SIPs
RODA IN
RODA IN
RODA IN

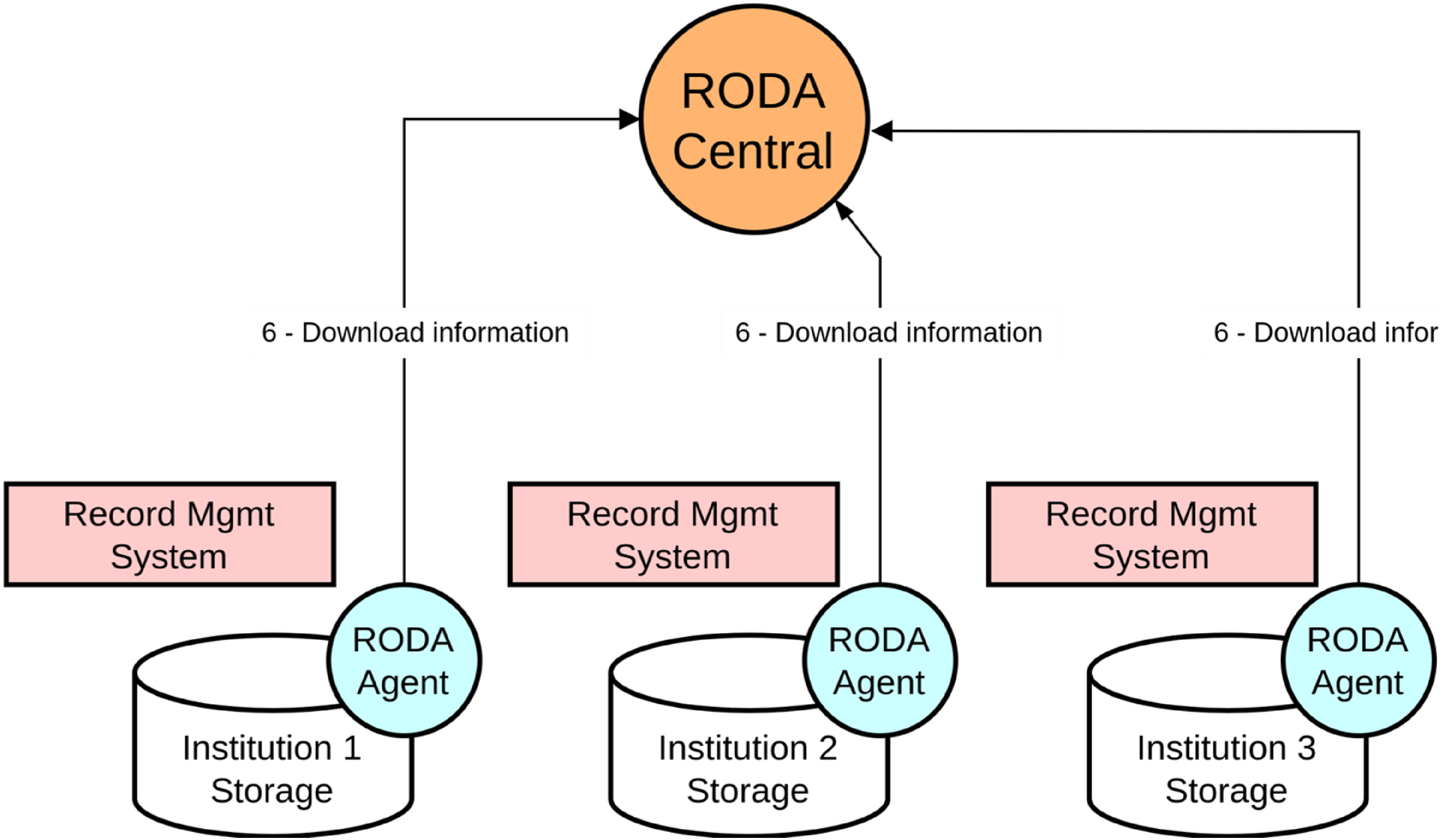


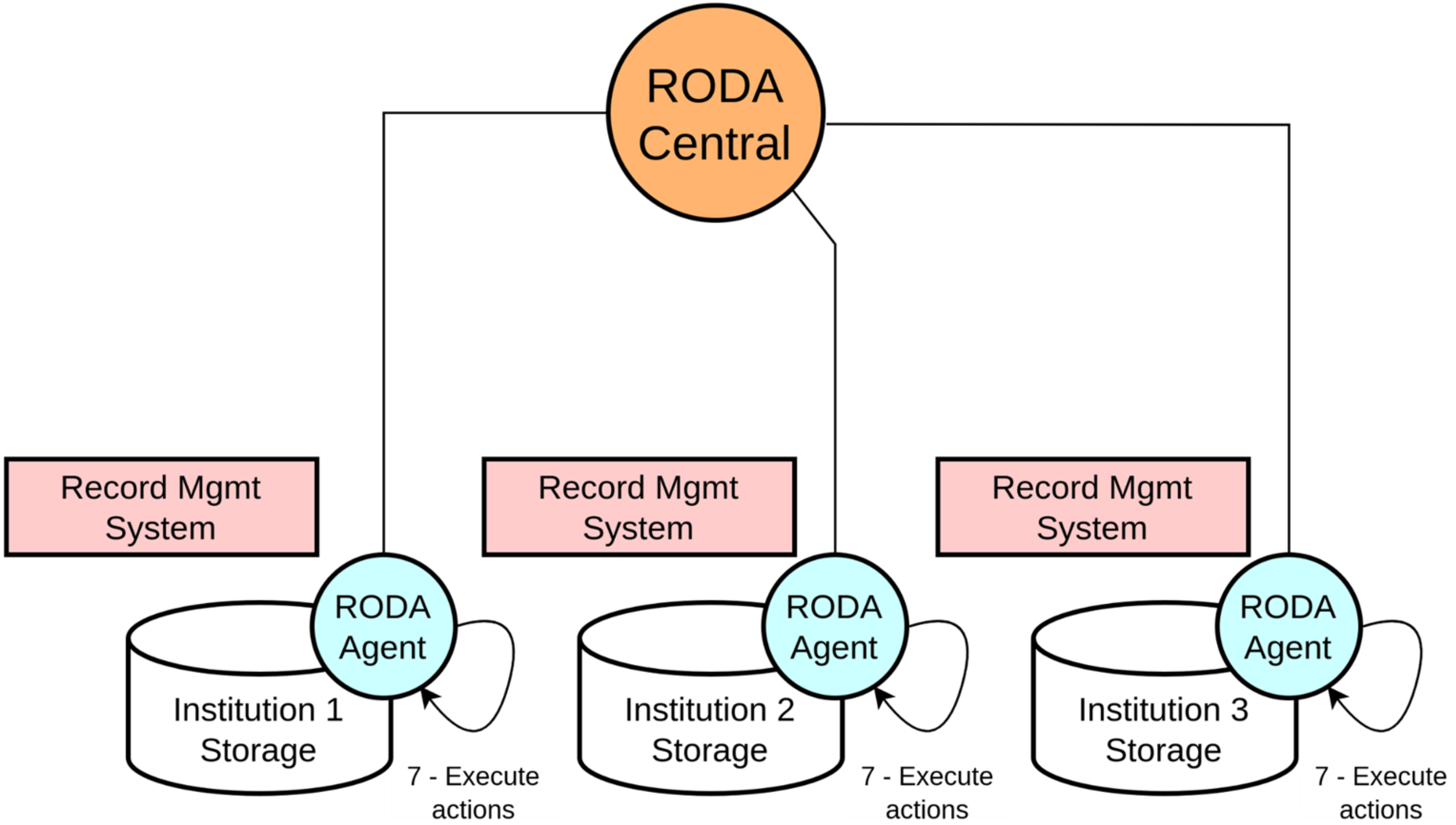
European Commission

4 - Preservation planning

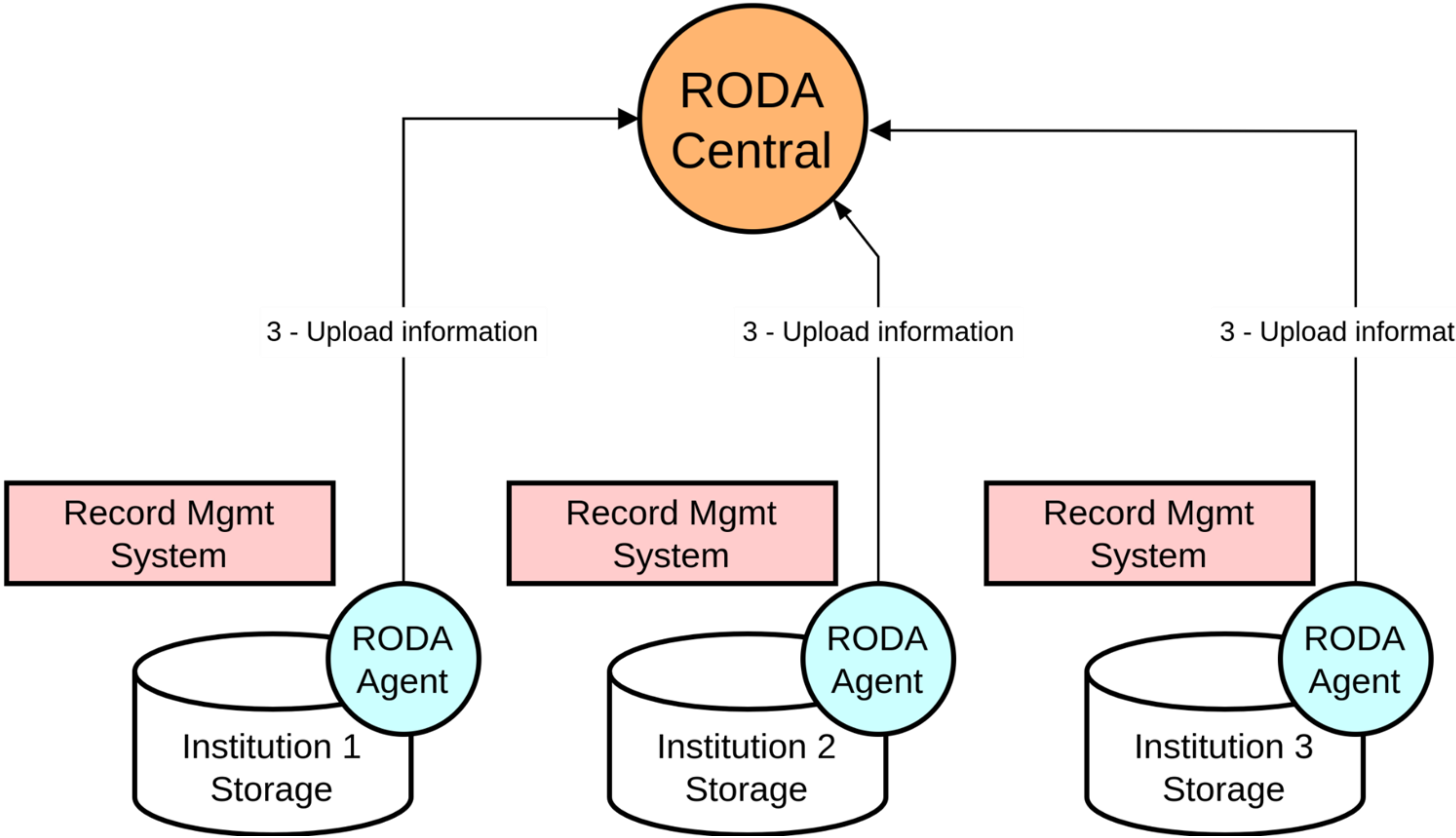
5 - Request execution of actions




RODA Central
6 - Download information
6 - Download information
6 - Download information
Record Mgmt System
Record Mgmt System
Record Mgmt System
RODA Agent
RODA Agent
RODA Agent
Institution 1 Storage
Institution 2 Storage
Institution 3 Storage



European Commission


RODA Central
3 - Upload information
3 - Upload information
3 - Upload information
Record Mgmt System
Record Mgmt System
Record Mgmt System
RODA Agent
RODA Agent
RODA Agent
Institution 1 Storage
Institution 2 Storage
Institution 3 Storage

Architectural requirements

Same RODA components and plugins in RODA central and agent

Bi-directional information passing with uni-directional contact

Always from RODA agent to RODA central

Periodic synchronisation process

Remote action requests requested by central and executed by agent

Shallow E-ARK IPs

Using files by reference in Information Packages

Why use shallow E-ARK IPs?

Lower the entry barrier for Institutions to Digital Preservation activities

Less storage infrastructure

Do NOT duplicate from the current business supporting systems.

Digital Preservation as an added value without drawbacks

Minimum additional infrastructure, minimum additional staff and staff training

Enabling of Distributed Digital Preservation strategy

RODA Central keeps information from remote RODA agents as shallow AIPs

E-ARK Information Packages

Specifications for SIP, AIP and DIP formats

Common base designated Common spec

Maintained by the DILCIS Board

Digital Information LifeCycle Interoperability Standards Board

Developed in the E-ARK project

Supported by the European Commission eArchiving Activity and supervised by the DLM Forum

Shallow E-ARK SIP 2

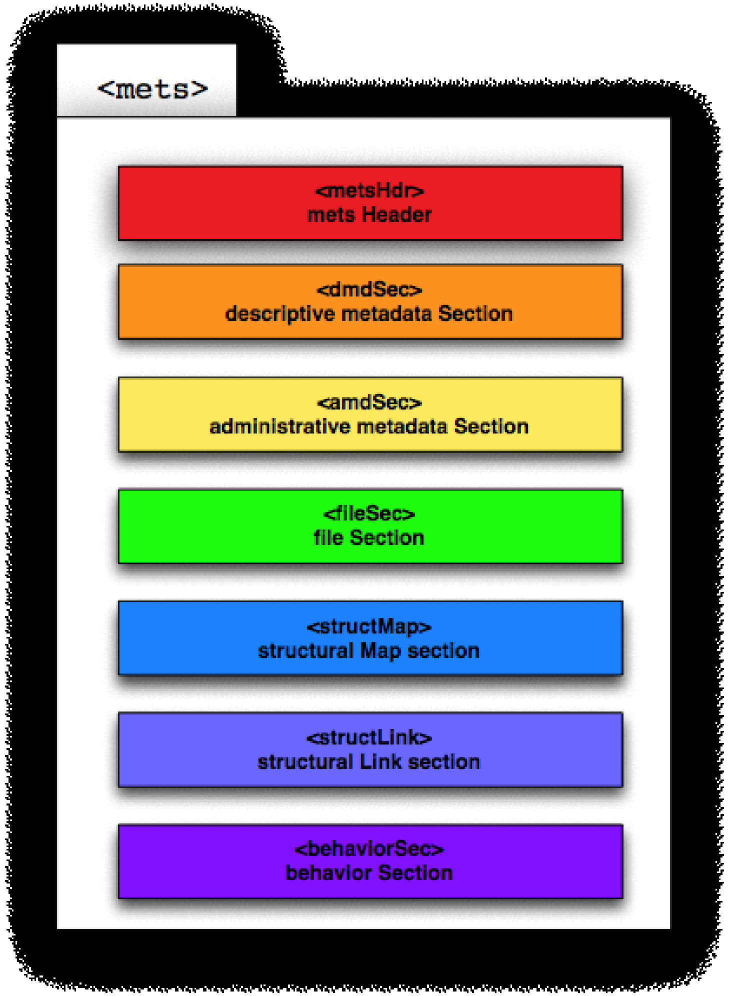
Follows the specification of the E-ARK SIP 2 format but with an extension in the data representations

It does NOT contain the data stored in the submission packages

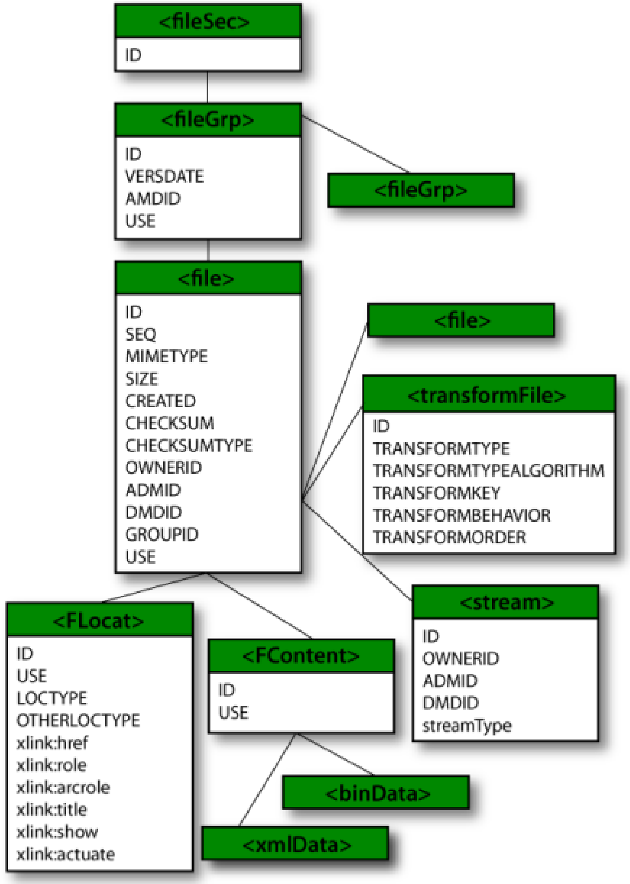
Instead, it has a reference to the files stored in an external location

Reference is made using the Uniform Resource Locator (URL) standard

METS Standard



<mets>
<metsHdr> mets Header
<dmdSec> descriptive metadata Section
<amdSec> administrative metadata Section
<fileSec> file Section
<structMap> structural Map section
<structLink> structural Link section
<behaviorSec> behavior Section



<FILE> FILE SECTION

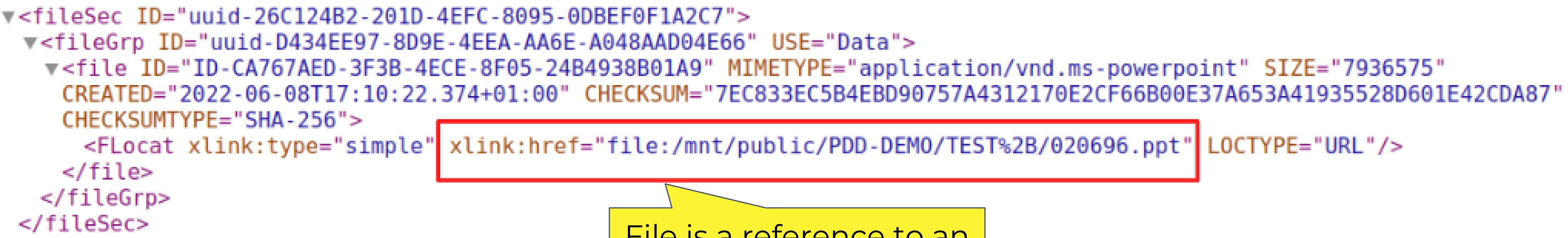
Source: https://www.loc.gov/standards/mets/METSPrimer.pdf

METS: File location

The file location element provides a pointer to the location of a content file. It uses the XLink reference syntax to provide linking information indicating the actual location of the content file, along with other attributes specifying additional linking information.

NOTE: is an empty element. The location of the resource pointed to MUST be stored in the xlink:href attribute.

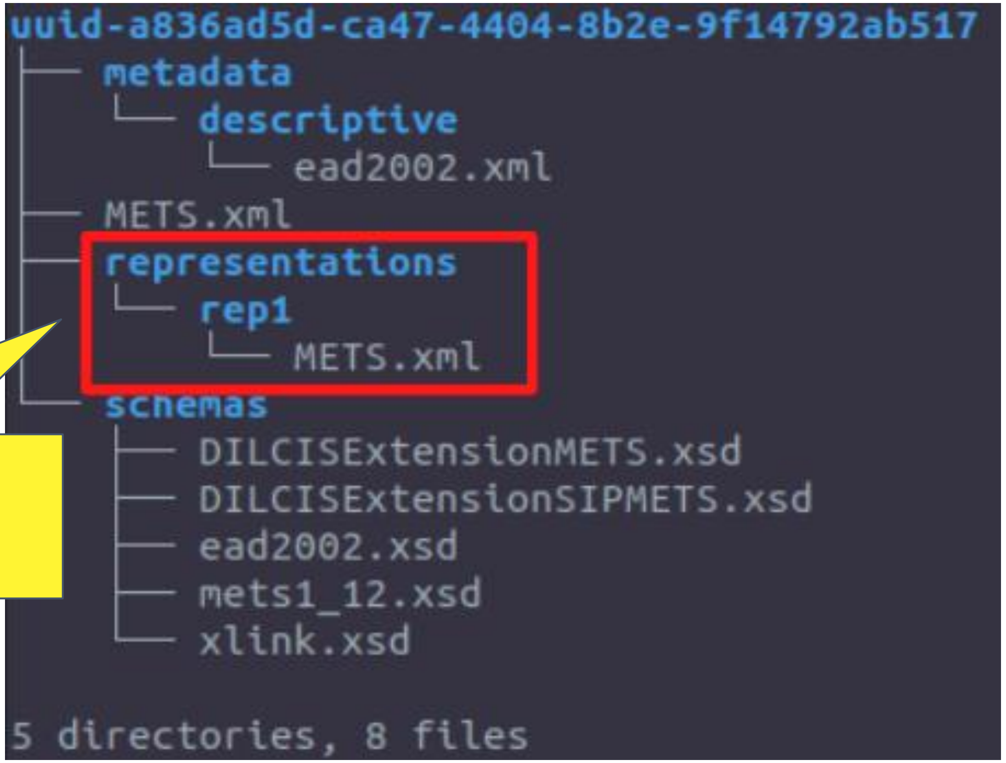
Shallow E-ARK SIP 2 (METS.xml)



```
▼<fileSec ID="uuid-26C124B2-201D-4EFC-8095-0DBEF0F1A2C7">
  ▼<fileGrp ID="uuid-D434EE97-8D9E-4EEA-AA6E-A048AAD04E66" USE="Data">
    ▼<file ID="ID-CA767AED-3F3B-4ECE-8F05-24B4938B01A9" MIMETYPE="application/vnd.ms-powerpoint" SIZE="7936575"
     CREATED="2022-06-08T17:10:22.374+01:00" CHECKSUM="7EC833EC5B4EBD90757A4312170E2CF66B00E37A653A41935528D601E42CDA87"
     CHECKSUMTYPE="SHA-256">
       <FLocat xlink:type="simple" xlink:href="file:/mnt/public/PDD-DEMO/TEST%2B/020696.ppt" LOCTYPE="URL"/>
     </file>
   </fileGrp>
 </fileSec>
```

File is a reference to an external location.

Shallow E-ARK SIP 2 (SIP.zip)



```
uuid-a836ad5d-ca47-4404-8b2e-9f14792ab517
├── metadata
│   └── descriptive
│       └── ead2002.xml
├── METS.xml
├── representations
│   └── rep1
│       └── METS.xml
└── schemas
    ├── DILCISExtensionMETS.xsd
    ├── DILCISExtensionSIPMETS.xsd
    ├── ead2002.xsd
    ├── mets1_12.xsd
    └── xlink.xsd

5 directories, 8 files
```

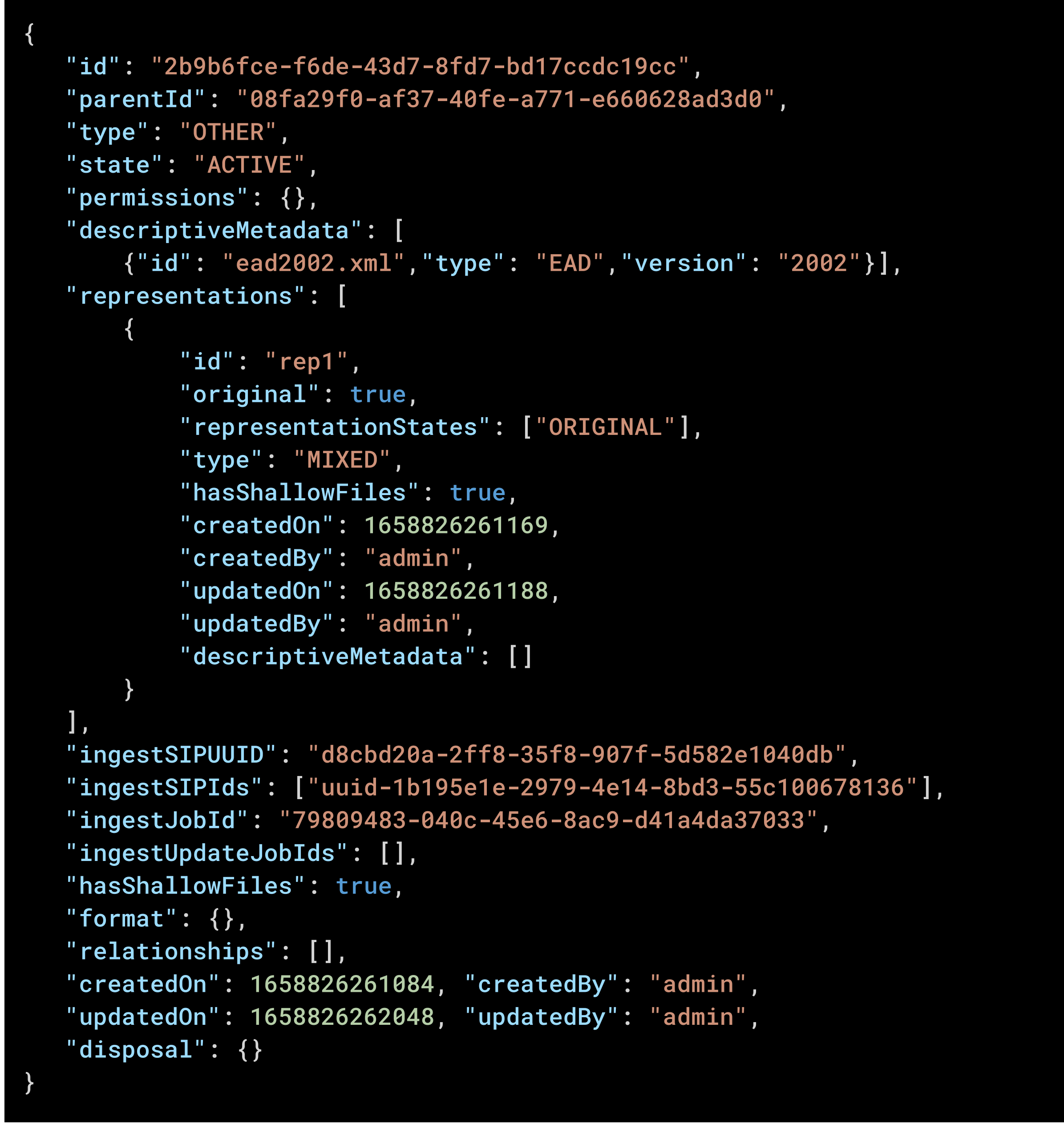
Representation contains only the METS file

E-ARK AIP & RODA AIP

E-ARK AIP has a METS.xml file containing a list of all files and their respective checksums

RODA AIP uses aip.json for performance and efficiency reasons

E-ARK AIP METS.xml file can be generated at any time using plugin: "E-ARK AIP Manifest Updater"



```
{
   "id": "2b9b6fce-f6de-43d7-8fd7-bd17ccdc19cc",
   "parentId": "08fa29f0-af37-40fe-a771-e660628ad3d0",
   "type": "OTHER",
   "state": "ACTIVE",
   "permissions": {},
   "descriptiveMetadata": [
       {"id": "ead2002.xml","type": "EAD","version": "2002"}],
   "representations": [
       {
           "id": "rep1",
           "original": true,
           "representationStates": ["ORIGINAL"],
           "type": "MIXED",
           "hasShallowFiles": true,
           "createdOn": 1658826261169,
           "createdBy": "admin",
           "updatedOn": 1658826261188,
           "updatedBy": "admin",
           "descriptiveMetadata": []
       }
   ],
   "ingestSIPUUID": "d8cbd20a-2ff8-35f8-907f-5d582e1040db",
   "ingestSIPIds": ["uuid-1b195e1e-2979-4e14-8bd3-55c100678136"],
   "ingestJobId": "79809483-040c-45e6-8ac9-d41a4da37033",
   "ingestUpdateJobIds": [],
   "hasShallowFiles": true,
   "format": {},
   "relationships": [],
   "createdOn": 1658826261084, "createdBy": "admin",
   "updatedOn": 1658826262048, "updatedBy": "admin",
   "disposal": {}
}
```

Shallow E-ARK AIP & RODA AIP

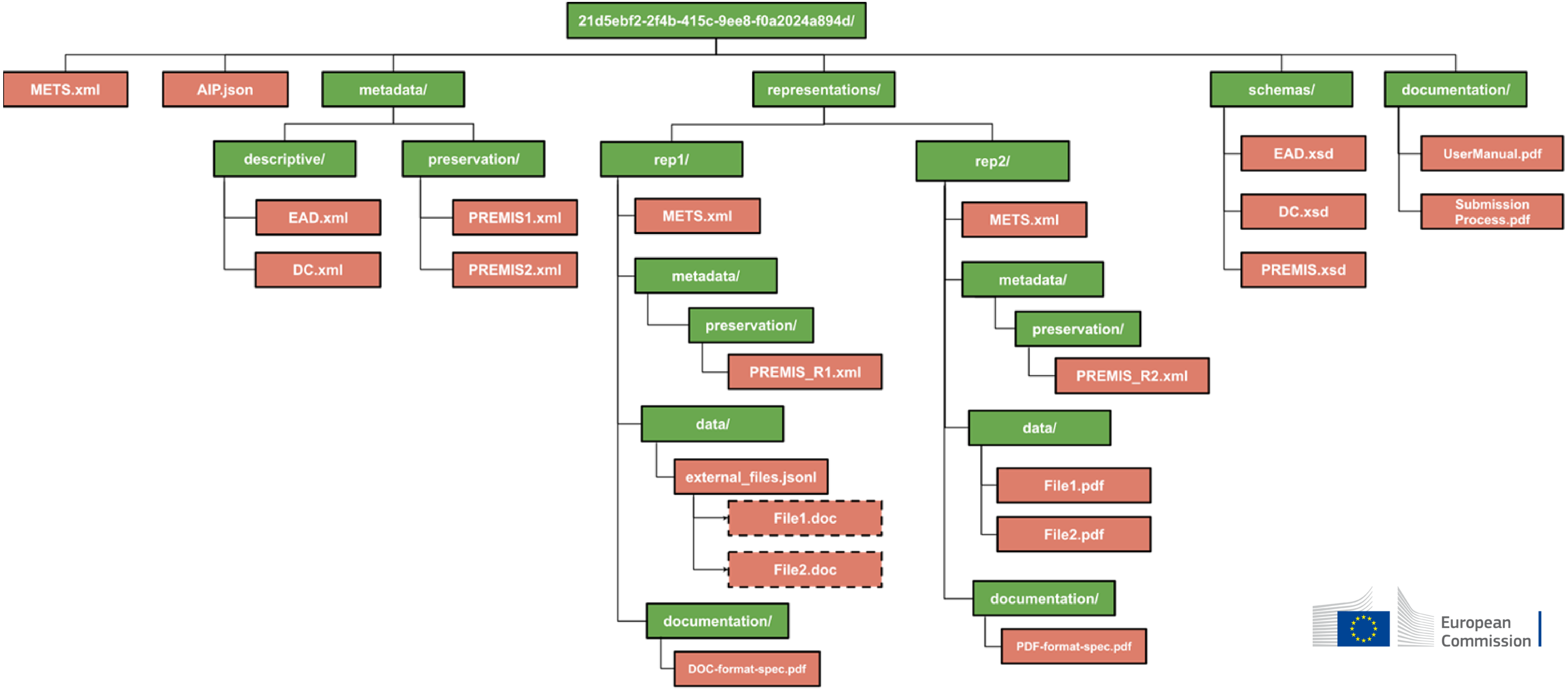
Shallow AIPs follows the E-ARK AIP 2.0.4 specification, except for the remote files in the representation data folder

There is an auxiliary file for each representation of an AIP containing the location of all external files existing in that representation

The file is in JSON Lines format and has the name: external_files.jsonl

https://jsonlines.org/

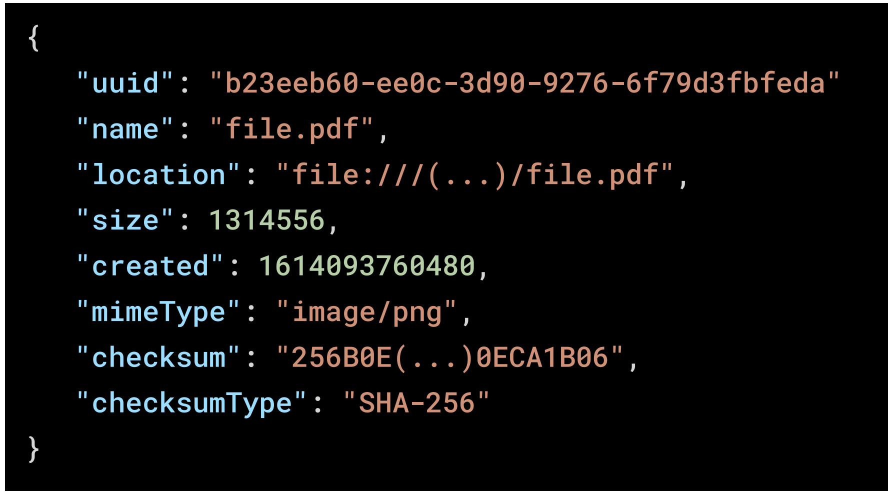
Shallow E-ARK AIP & RODA AIP



external_files.jsonl

Essential file information:

- Persistent Identifier (UUID)
- File name
- Location (URL)
- Size (in bytes)
- Creation date (seconds since epoch)
- File format (MIME Type)
- Checksum (value and algorithm)

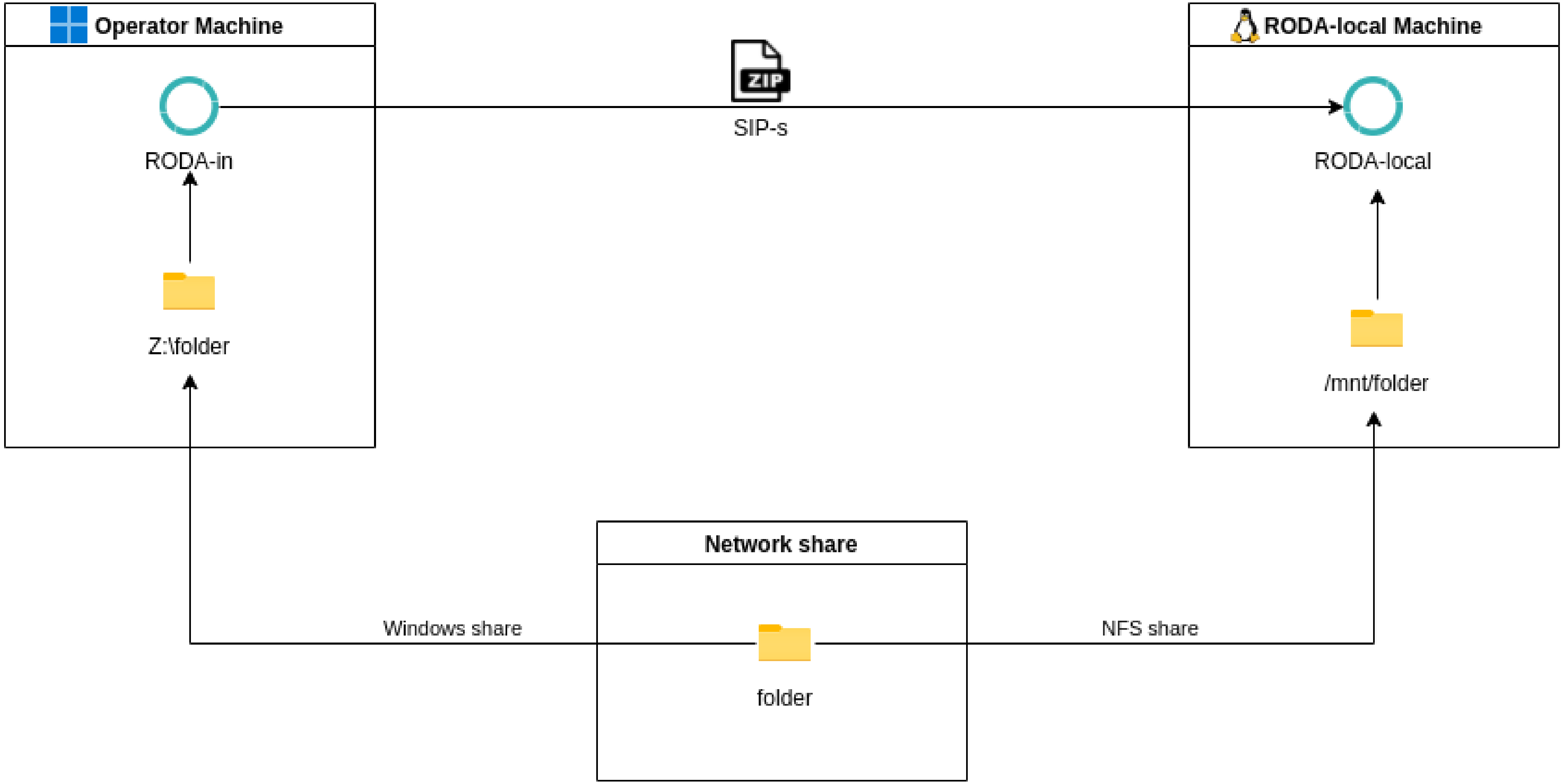


```
{
   "uuid": "b23eeb60-ee0c-3d90-9276-6f79d3fbfeda"
   "name": "file.pdf",
   "location": "file:///(...)/file.pdf",
   "size": 1314556,
   "created": 1614093760480,
   "mimeType": "image/png",
   "checksum": "256B0E(...)0ECA1B06",
   "checksumType": "SHA-256"
}
```

Creating a Shallow SIP

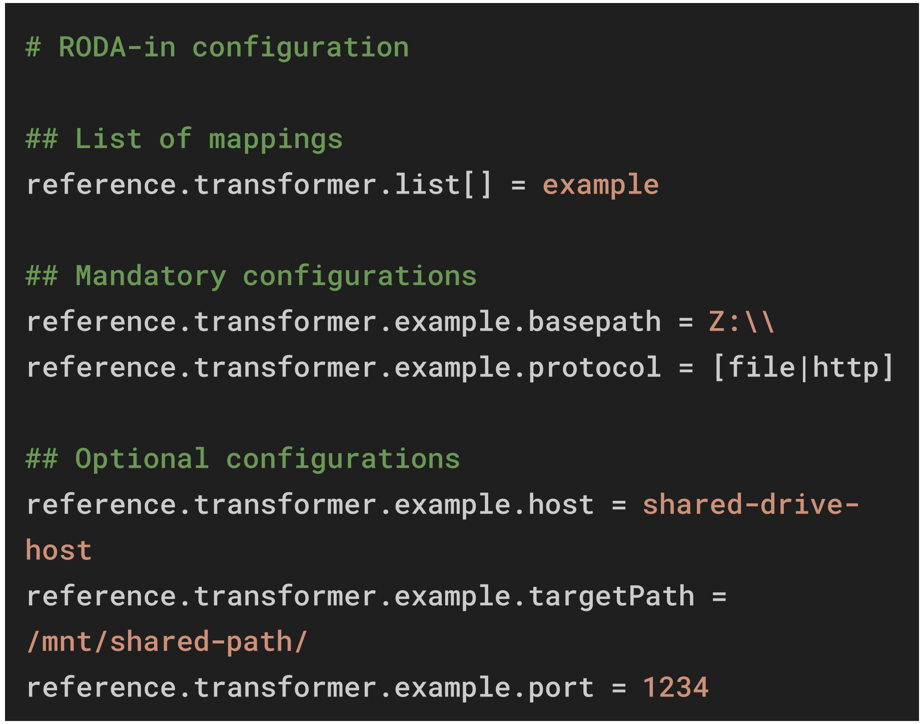
Using RODA-in

Use case


Operator Machine
RODA-in
Z:\folder
SIP-s
RODA-local Machine
RODA-local
/mnt/folder
Network share
folder
Windows share
NFS share

Workflow

1. Configure shared folder in RODA Agent
2. Configure shared folder on operator machine (Map network drive…)
3. Configure RODA-in on the operator's machine
4. Create SIP-S on operator machine
5. Transfer SIP-S to RODA-local and start ingestion



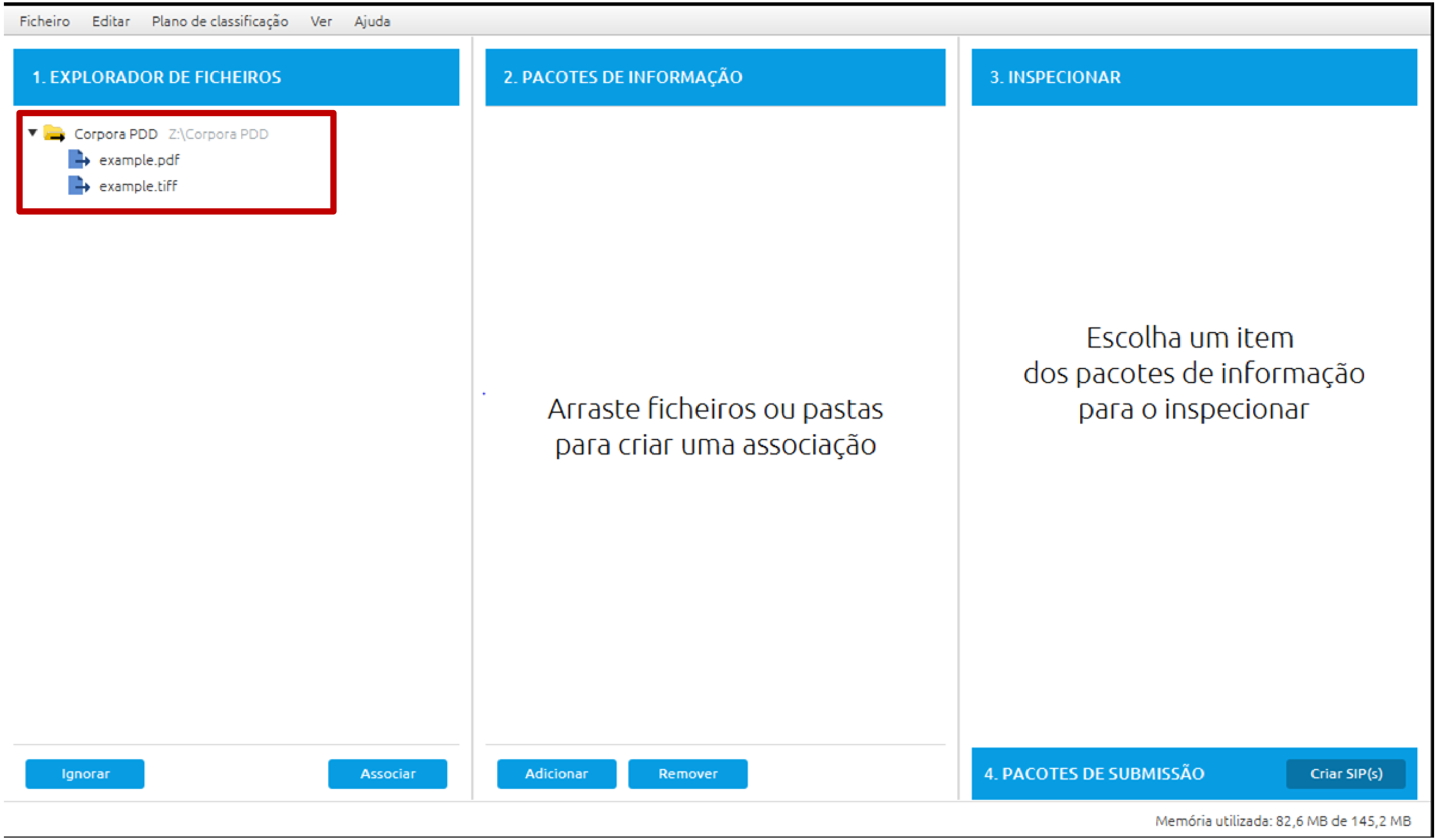
```
# RODA-in configuration

## List of mappings
reference.transformer.list[] = example

## Mandatory configurations
reference.transformer.example.basepath = Z:\\
reference.transformer.example.protocol = [file|http]

## Optional configurations
reference.transformer.example.host = shared-drive-host
reference.transformer.example.targetPath = /mnt/shared-path/
reference.transformer.example.port = 1234
```

RODA-in: icons show files by reference

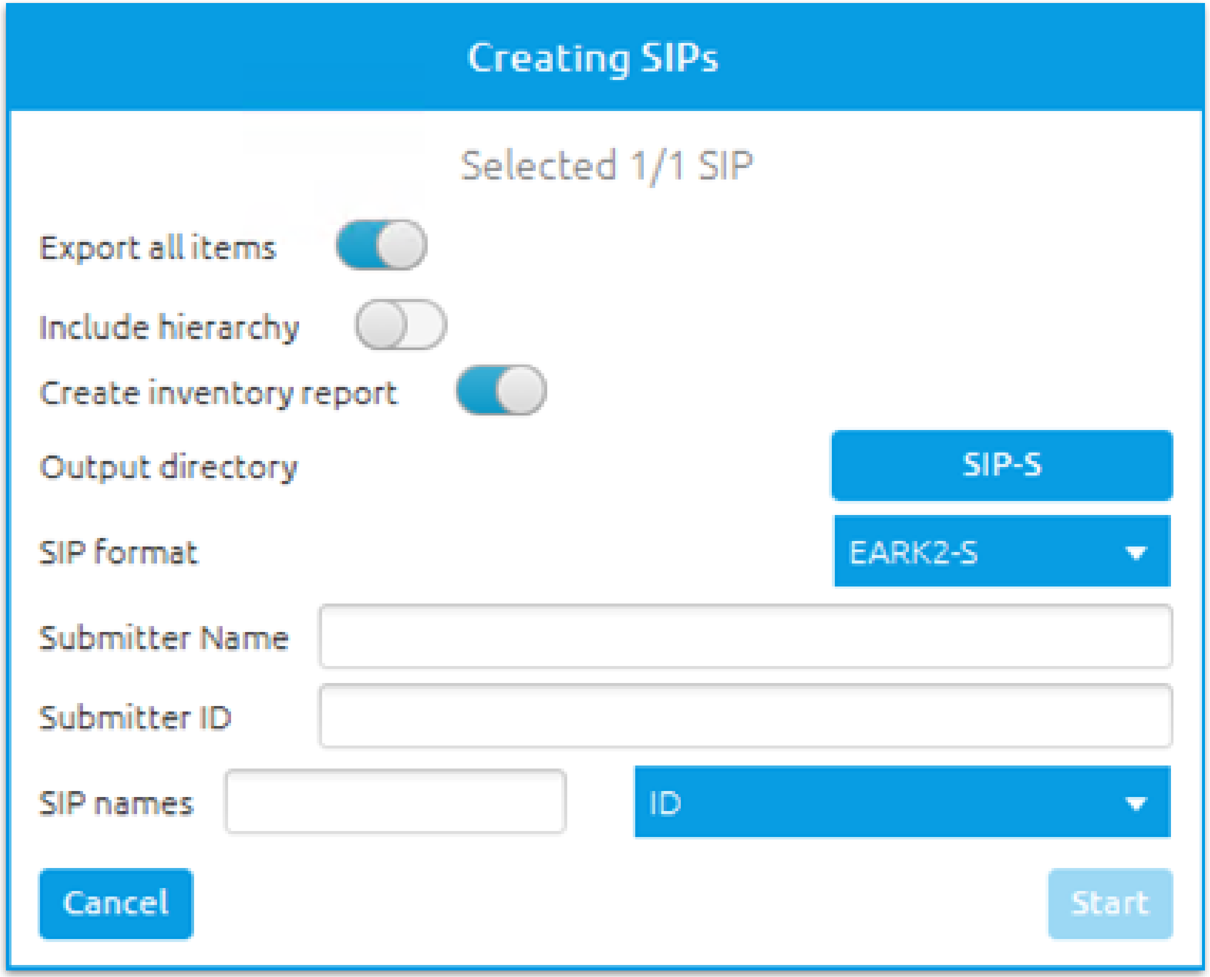


Arraste ficheiros ou pastas para criar uma associação

Escolha um item dos pacotes de informação para o inspecionar

Memória utilizada: 82,6 MB de 145,2 MB

RODA-in: SIP Format E-ARK2-S (Shallow)



Creating SIPs

Selected 1/1 SIP

Export all items: on

Include hierarchy: off

Create inventory report: on

Output directory: SIP-S

SIP format: EARK2-S

Submitter Name

Submitter ID

SIP names: ID

Other use cases

File servers
↳Manual procedure using RODA-in

Document Management Systems
↳Integration built upon commons-ip

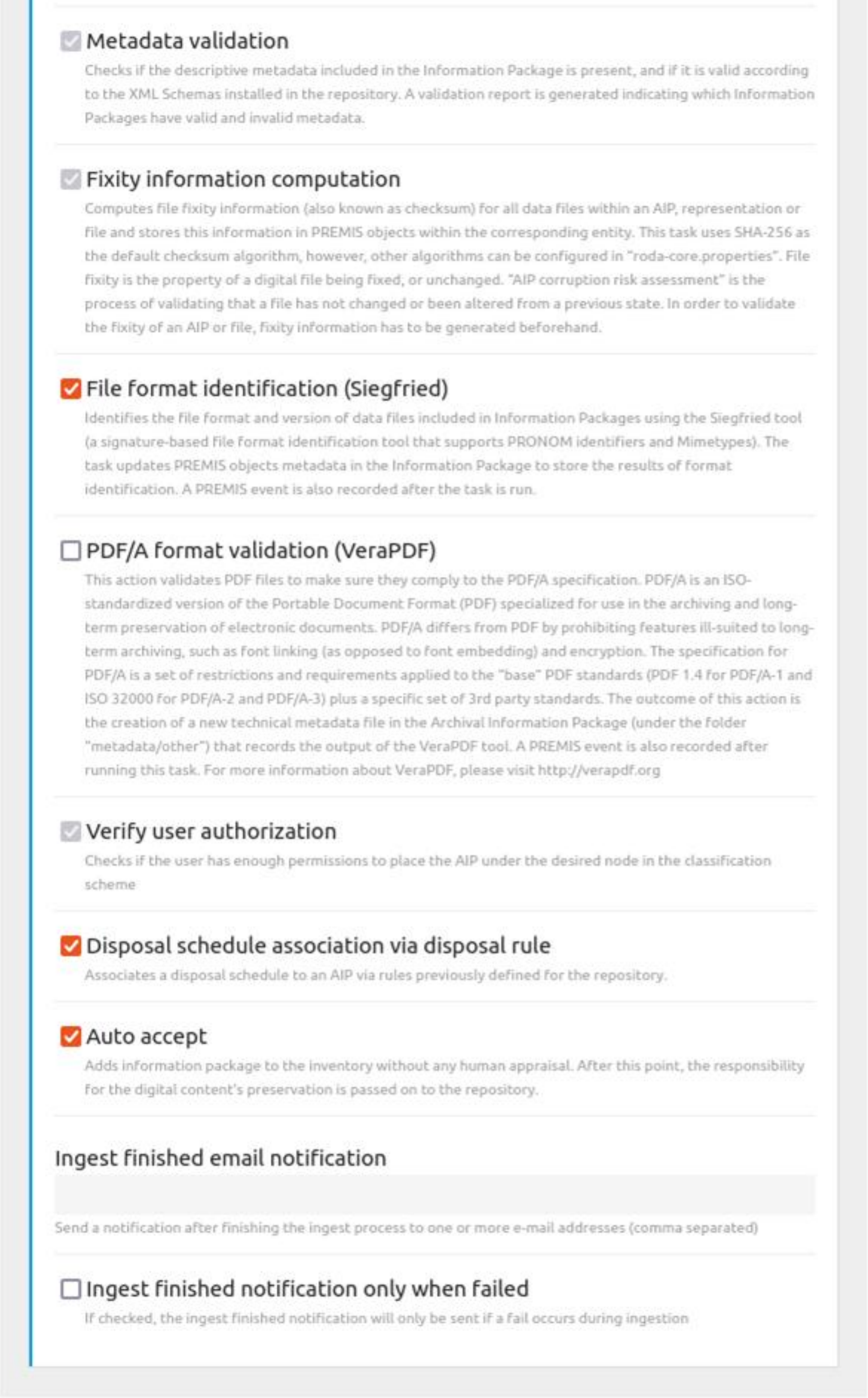
Relational databases
↳Database preservation toolkit (DBPTK) and then one of the previous

European Commission

RODA agent

Ingest and actions with shallow IPs

Default ingest workflow



☑ 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

☑ Verify user authorization

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

☑ Disposal schedule association via disposal rule

Associates a disposal schedule to an AIP via rules previously defined for the repository.

☑ 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

1. Validate Shallow SIP and convert to Shallow AIP
2. Override parent node
3. Validate descriptive metadata
4. Generate preservation metadata (fixity information computation)
5. Identify file formats
6. Other plugins: Validate PDF/A, Fulltext extraction etc.
7. Verify user authorisation
8. Disposal schedule association
9. Auto-accept (skip manual validation)
10. Notifications: emails, webhooks, file reports

File reference as an URL

```
URL = protocol://service/path
```

Protocol
Defines how access is provided and identifies the protocol manager who knows how to operate the service to access the files. Protocol examples are HTTPS, FTP, NFS, and FILE.

Service
Defines the location on the network (domain or IP, optionally port) where the service providing access is operating under the protocol identified above.

Path
Identifies the file in the context of the service providing it.

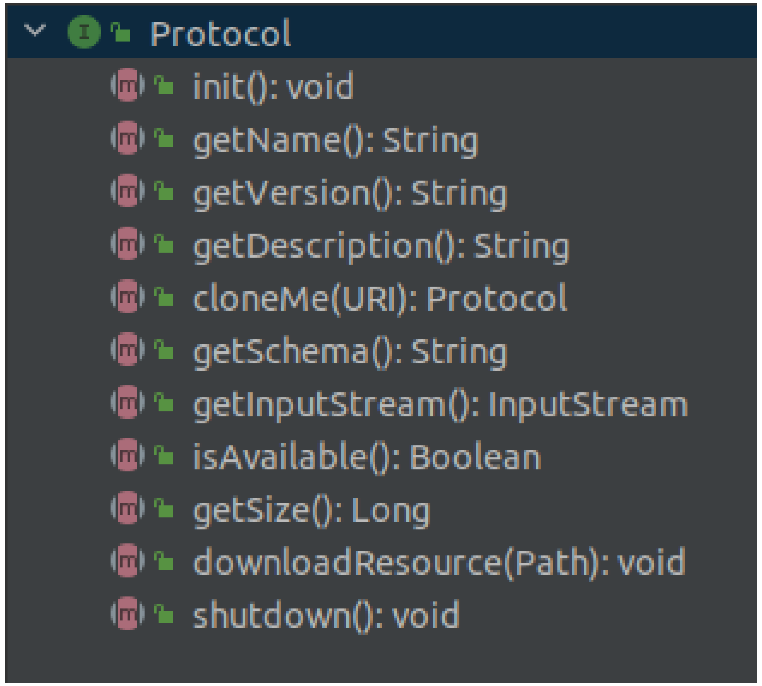
Protocol manager

READ-ONLY access to a file reference.

Configuration of authentication method.

Provides access to:

- ➔ File content (streaming, partial access)
- ➔ File technical metadata (size, checksum, format)



```
Protocol
    init(): void
    getName(): String
    getVersion(): String
    getDescription(): String
    cloneMe(URI): Protocol
    getSchema(): String
    getInputStream(): InputStream
    isAvailable(): Boolean
    getSize(): Long
    downloadResource(Path): void
    shutdown(): void
```

Available Protocol managers:

- ➔ **FILE**: for file system access for read-only mounts (e.g. NFS)
- ➔ **HTTP(S)**: for HTTP or REST-API resources
- ➔ **RODA**: for distributed digital preservation use case
  (i.e. how central sees the agent's files)

RODA working with files by reference (1 of 2)

1. Files created with the RODA API are always local
   External files can only be created from ingesting Shallow SIP
2. External files updated by the RODA API are made local
3. External files can be removed by the RODA API

The **result of preservation actions** always creates local files

Such as the creation of new representations or disseminations from format migration actions

RODA working with files by reference (2 of 2)

Avoid filename conflicts between local and external ones

But if there is a conflict, it should prefer the local file.

Metadata is always local

Including descriptive, preservation, or other.

Access to the content of external files in a transparent way.

Random access to parts of the file must also be allowed

Given the protocol manager supports it

Backwards compatibility is a must

All internal actions, preservation actions and plugins remain functional when executed on external files.

Synchronization

Formats and processes

Synchronization requirements

RODA agent must always be the source of connection

RODA agents inside a DMZ

Atomic batch synchronization

Create a synchronization package with the differential since last synchronisation.
Avoid network issues during synchronization.

RODA instance identification

All entities being synchronized will identify the instance they belong to.

RODA instance identification

Generated during RODA initial setup

Can be changed to be more representative of the institution

Is present in every entity in RODA

AIP, Representation Information, Risks, etc

In RODA Central is presented in the UI

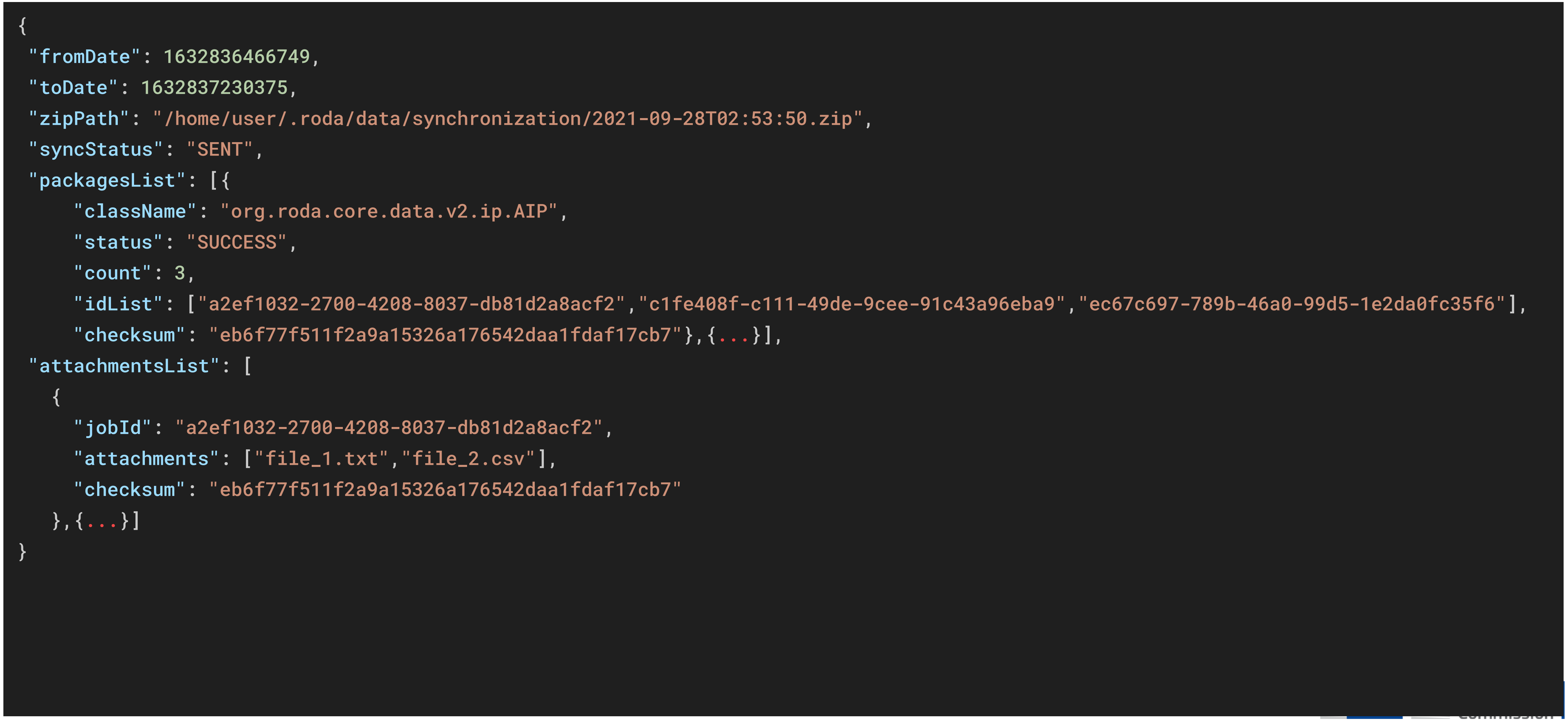
To identify the remote instance the content belongs to

RODA Central can have content from both local and remote instances (hybrid)

Batch synchronization format (ZIP)

```
2021-09-28T02:53:50.zip
├── state.json
├── storage
│              ├── aip
│              ├── job
│              ├── job-report
│              └── preservation
└── attachments
        └── job_id
                              └── attachment_id
```

Batch synchronization format: Manifest (state.json)



```
{
 "fromDate": 1632836466749,
 "toDate": 1632837230375,
 "zipPath": "/home/user/.roda/data/synchronization/2021-09-28T02:53:50.zip",
 "syncStatus": "SENT",
 "packagesList": [{
     "className": "org.roda.core.data.v2.ip.AIP",
     "status": "SUCCESS",
     "count": 3,
     "idList": ["a2ef1032-2700-4208-8037-db81d2a8acf2","c1fe408f-c111-49de-9cee-91c43a96eba9","ec67c697-789b-46a0-99d5-1e2da0fc35f6"],
     "checksum": "eb6f77f511f2a9a15326a176542daa1fdaf17cb7"},{...}],
 "attachmentsList": [
   {
     "jobId": "a2ef1032-2700-4208-8037-db81d2a8acf2",
     "attachments": ["file_1.txt","file_2.csv"],
     "checksum": "eb6f77f511f2a9a15326a176542daa1fdaf17cb7"
   },{...}]
}
```

Batch synchronization format: Manifest (state.json)

ATTRIBUTE	DESCRIPTION
fromDate and toDate	Time range that is used to filter and include the entities to be added to the sync batch
zipPath	Location of the packaged batch in zip format on the RODA agent
syncStatus	Status of the synchronization plug-in between agent and central RODA
packageList	List of packages generated by each RODA entity
className	Bundle entity class, used for re-indexing process
status	Status of package creation in the RODA agent
count	Number of packaged objects of the entity
idList	List of packaged object identifiers
checksum	Merkle tree top hash of package contents
attachmentsList	List of processes with their attachments to be synchronized
jobId	Identifier of the process that generated the attachment file
attachments	List of attachments generated by the process

Synchronization process

1. RODA Agent creates sync package (differential from last sync)
2. RODA Agent send sync package to RODA Central
3. RODA Central validated and incorporates sync package
4. RODA Agent requests sync package from RODA Central
5. RODA Central creates sync package (differential from last sync for the requesting instance)
6. RODA Agent receives sync package from central, validates and incorporates.
7. RODA Agent executes action requests received from from sync package
8. RODA Agent send action reports, attachments and other updates in next sync

Information sent from agent to central

Processes about done actions and action reports:

- Ingest, Internal and (Preservation) Action processes

AIP:

- Basic structure of AIPs and representations
- Metadata (descriptive, preservation and other)
- References to files with basic metadata (size, checksum, formats)

DIP:

- Metadata
- References to files

Preservation metadata:

- Preservation events
- Preservation agents

Risk incidences (Specific events that relate risks with AIPs, Representation and Files)

Information sent from central to agent

(Preservation) Action requests

Risks

Representation Information

Including rules for connecting Representation Information with AIPs, Representations and Files.

Soon:

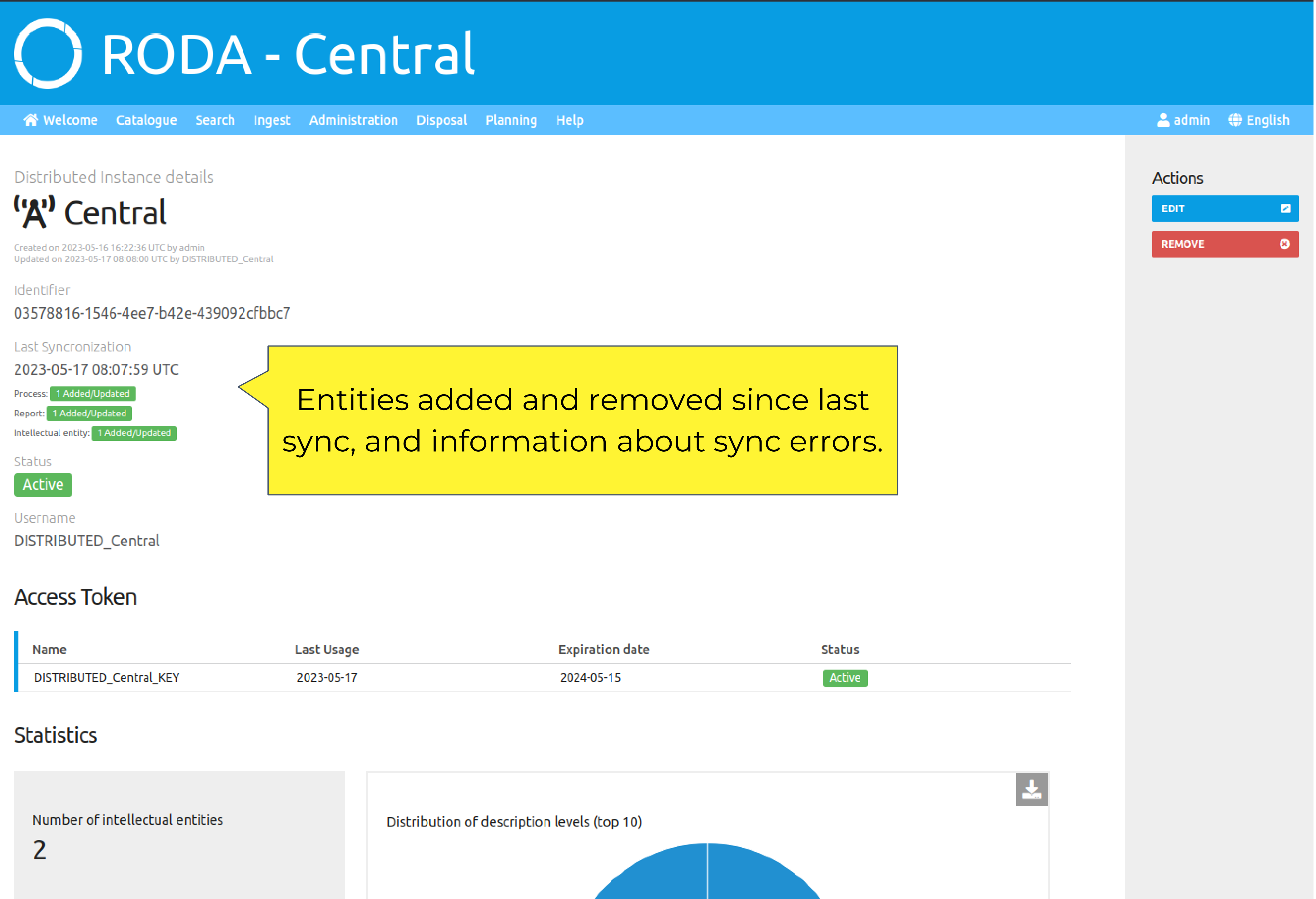
- Disposal schedules and rules

Deleted entities and completion validation

Local sends a complete list of: AIP IDs, DIP IDs, Risk Incidence IDs.

Central send a complete list of: Representation Information IDs, Risk IDs.

Sanity check report presented in sync status.



RODA - Central

Distributed Instance details

Central

Created on 2023-05-16 16:22:36 UTC by admin
Updated on 2023-05-17 08:08:00 UTC by DISTRIBUTED_Central

Identifier
03578816-1546-4ee7-b42e-439092cfbbc7

Last Syncronization
2023-05-17 08:07:59 UTC

Process: 1 Added/Updated
Report: 1 Added/Updated
Intellectual entity: 1 Added/Updated

Status
Active

Username
DISTRIBUTED_Central

Access Token

Name	Last Usage	Expiration date	Status
DISTRIBUTED_Central_KEY	2023-05-17	2024-05-15	Active

Statistics

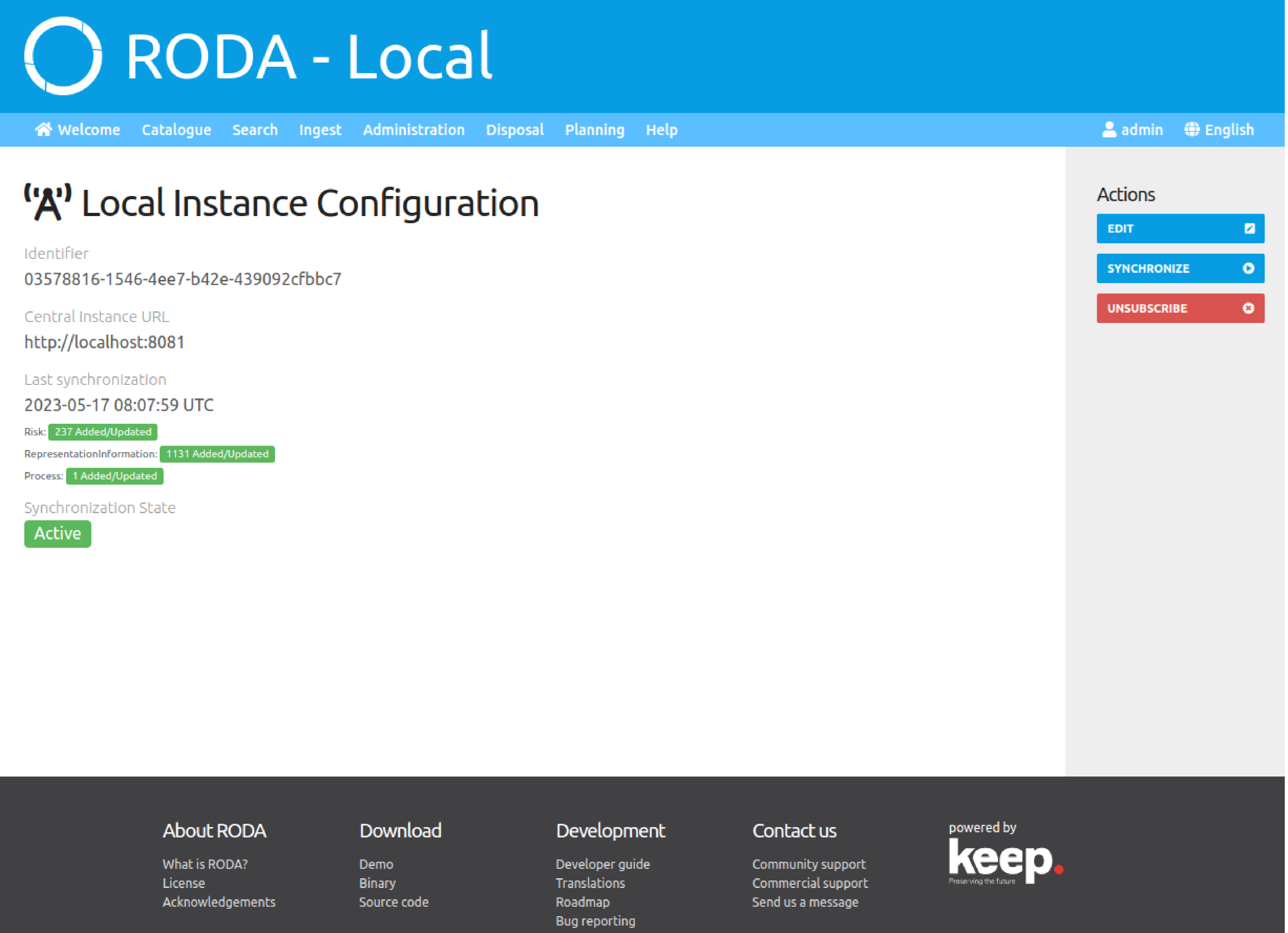
Number of intellectual entities
2

Distribution of description levels (top 10)

Actions

Entities added and removed since last sync, and information about sync errors.

European Commission



Local Instance Configuration

Identifier
03578816-1546-4ee7-b42e-439092cfbbc7

Central Instance URL
http://localhost:8081

Last synchronization
2023-05-17 08:07:59 UTC

Risk: 237 Added/Updated
RepresentationInformation: 1131 Added/Updated
Process: 1 Added/Updated

Synchronization State
Active

European Commission

Remote actions

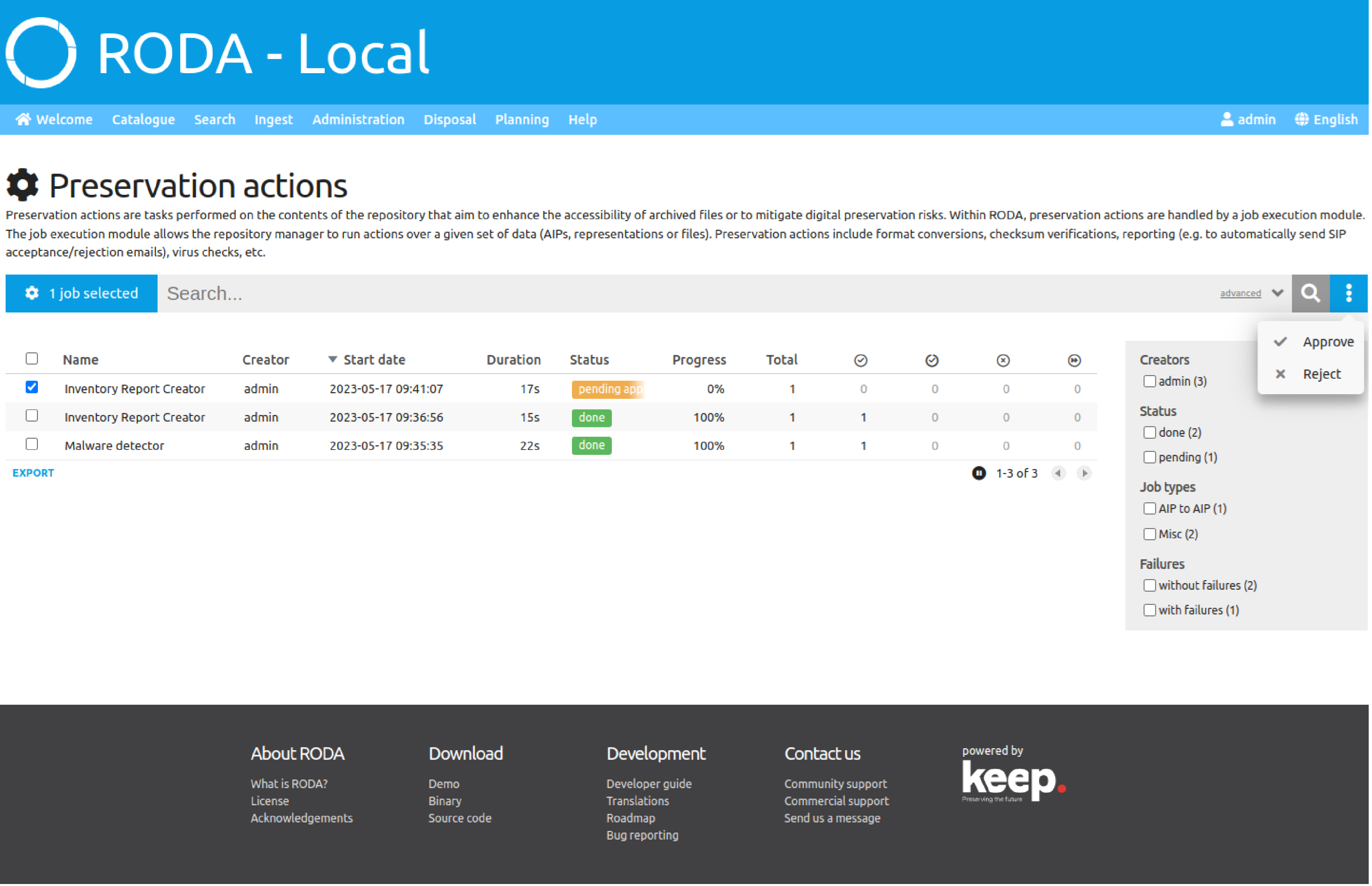
Requests and result feedback

Action request execution modes

Approval
RODA agent administrators must accept or reject the preservation action requested by RODA central.

Scheduled
RODA agent executes the requests automatically but on a predefined time window.

Immediate execution
RODA agent executes the requested action as soon as the instances are synchronized.



Preservation actions

Preservation actions are tasks performed on the contents of the repository that aim to enhance the accessibility of archived files or to mitigate digital preservation risks. Within RODA, preservation actions are handled by a job execution module. The job execution module allows the repository manager to run actions over a given set of data (AIPs, representations or files). Preservation actions include format conversions, checksum verifications, reporting (e.g. to automatically send SIP acceptance/rejection emails), virus checks, etc.

☐	Name	Creator	▼ Start date	Duration	Status	Progress	Total	✓	✓	⊗	⏩
☑	Inventory Report Creator	admin	2023-05-17 09:41:07	17s	pending app	0%	1	0	0	0	0
☐	Inventory Report Creator	admin	2023-05-17 09:36:56	15s	done	100%	1	1	0	0	0
☐	Malware detector	admin	2023-05-17 09:35:35	22s	done	100%	1	1	0	0	0

Creators
- ☐ admin (3)

Status
- ☐ done (2)
- ☐ pending (1)

Job types
- ☐ AIP to AIP (1)
- ☐ Misc (2)

Failures
- ☐ without failures (2)
- ☐ with failures (1)

About RODA
What is RODA?
License
Acknowledgements

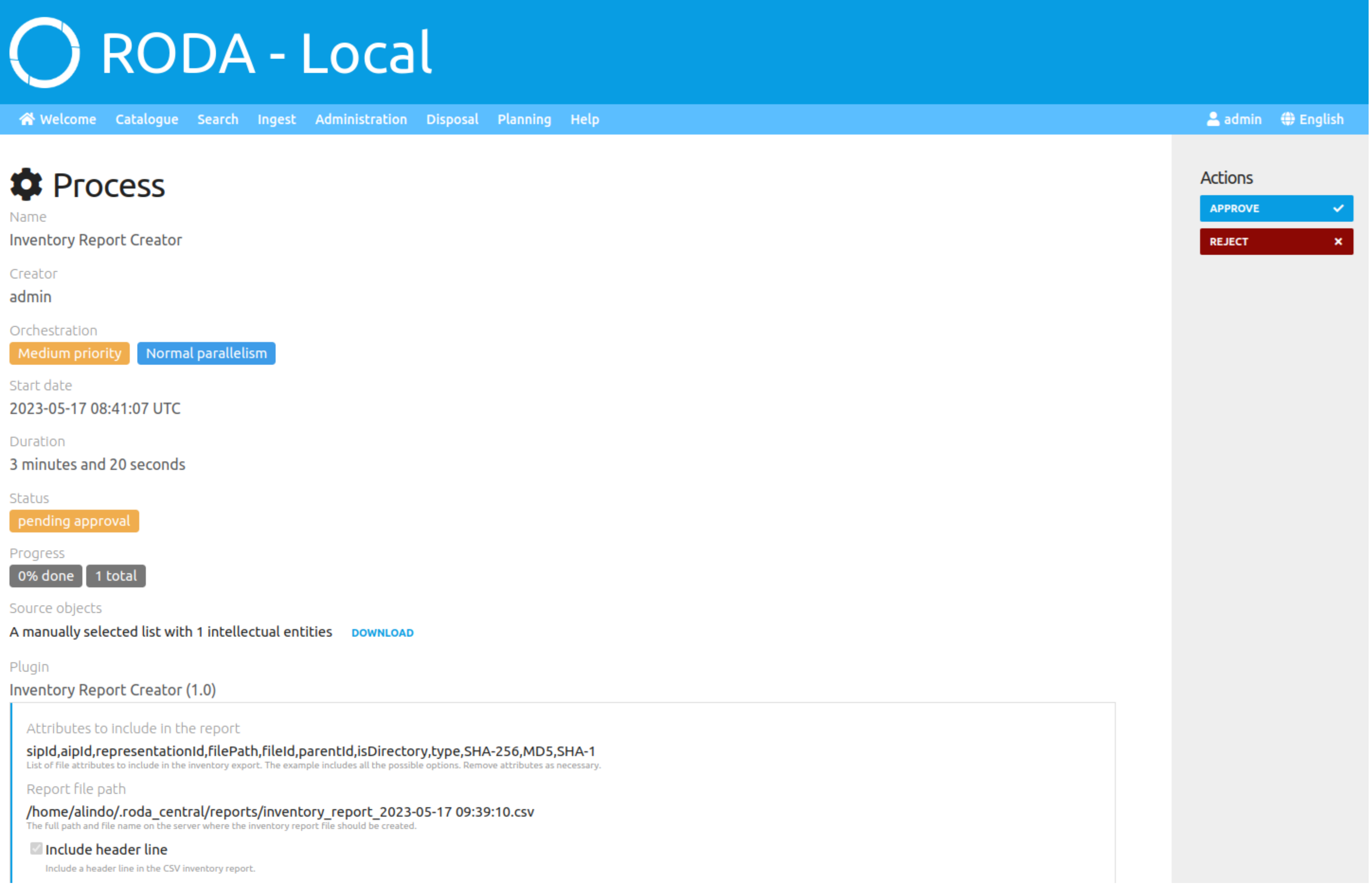
Download
Demo
Binary
Source code

Development
Developer guide
Translations
Roadmap
Bug reporting

Contact us
Community support
Commercial support
Send us a message

powered by keep. Preserving the future

European Commission



Process

Name
Inventory Report Creator

Creator
admin

Orchestration
Medium priority Normal parallelism

Start date
2023-05-17 08:41:07 UTC

Duration
3 minutes and 20 seconds

Status
pending approval

Progress
0% done 1 total

Source objects
A manually selected list with 1 intellectual entities DOWNLOAD

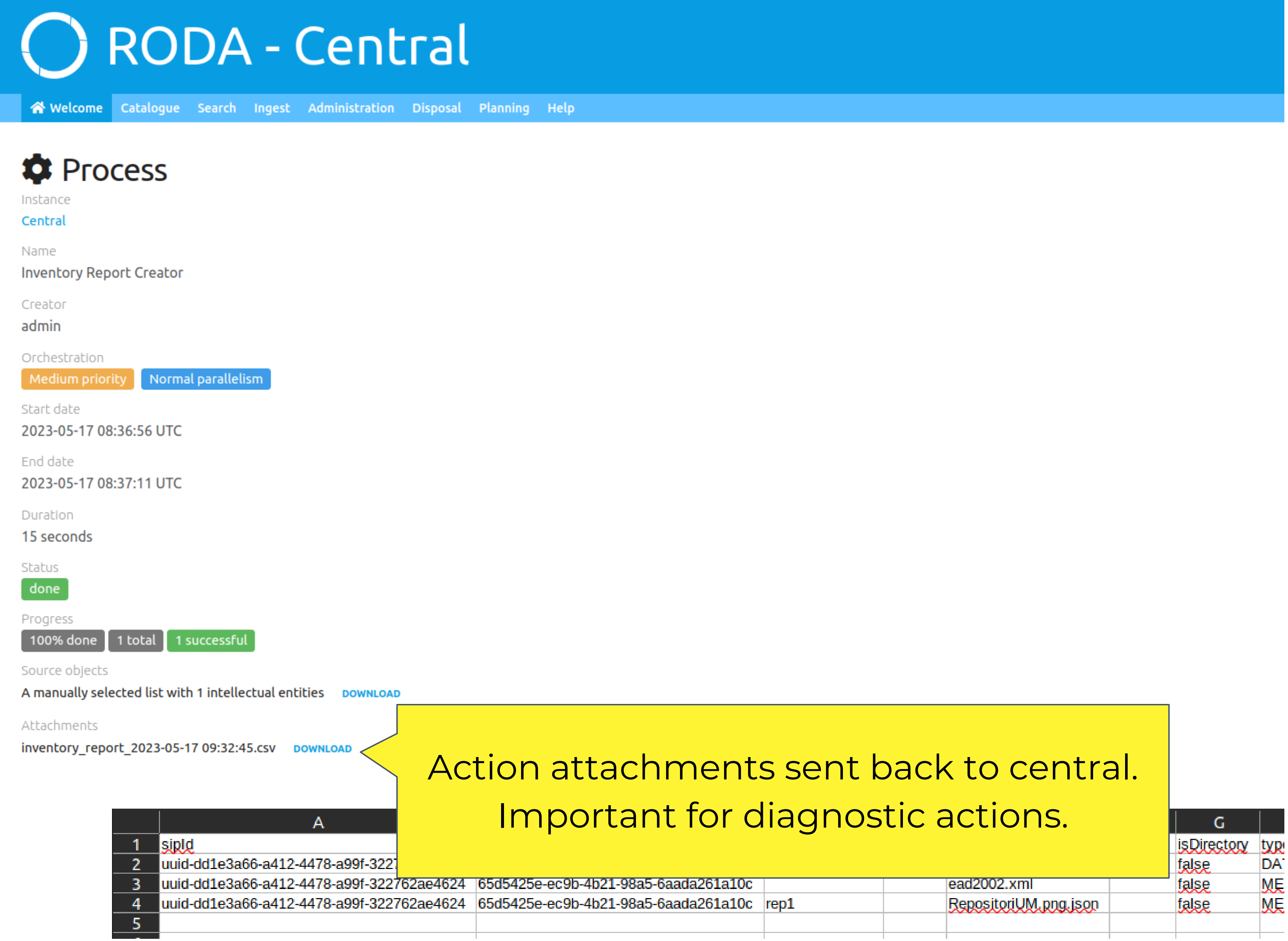
Plugin
Inventory Report Creator (1.0)

Attributes to include in the report
sipId,aipId,representationId,filePath,fileId,parentId,isDirectory,type,SHA-256,MD5,SHA-1
List of file attributes to include in the inventory export. The example includes all the possible options. Remove attributes as necessary.

Report file path
/home/alindo/.roda_central/reports/inventory_report_2023-05-17 09:39:10.csv
The full path and file name on the server where the inventory report file should be created.

☑ Include header line
Include a header line in the CSV inventory report.

European Commission



Process

Instance
Central

Name
Inventory Report Creator

Creator
admin

Orchestration
Medium priority
Normal parallelism

Start date
2023-05-17 08:36:56 UTC

End date
2023-05-17 08:37:11 UTC

Duration
15 seconds

Status
done

Progress
100% done
1 total
1 successful

Source objects
A manually selected list with 1 intellectual entities DOWNLOAD

Attachments
inventory_report_2023-05-17 09:32:45.csv DOWNLOAD

	A						G	
1	sipId						isDirectory	typ
2	uuid-dd1e3a66-a412-4478-a99f-3227						false	DA
3	uuid-dd1e3a66-a412-4478-a99f-322762ae4624	65d5425e-ec9b-4b21-98a5-6aada261a10c			ead2002.xml		false	ME
4	uuid-dd1e3a66-a412-4478-a99f-322762ae4624	65d5425e-ec9b-4b21-98a5-6aada261a10c	rep1		RepositoriUM.png.json		false	ME
5								

Action attachments sent back to central.
Important for diagnostic actions.

Available actions

Diagnose, identify risks, define mitigation strategy, request mitigation actions, evaluate and assure quality.

Characterization plugins

File Format Detector

The File Format Detector plugin is an essential tool for identifying and analysing various file formats. It provides comprehensive information about each file, including its name, designation, version, MIME type, and PRONOM identifier.

Office Documents Text Extractor

The Office Documents Text Extractor extracts the textual content from a vast array of document formats, including but not limited to Microsoft Office (Word, Excel, PowerPoint, etc.), PDF, RTF, ODT, HTML, XML, etc. The extracted textual content is then available for search, so you can find documents by searching words in their content.

File Feature Extractor

The File Feature Extractor is a powerful plugin that allows users to extract technical metadata from a wide range of file formats, making it an essential tool for digital curators.

Optical Character Recognition Extractor

The Optical Character Recognition Extractor is a powerful plugin designed to extract typed or printed text from digitalised images, making it an essential tool for professionals in various fields, including data analysis, document management, and research.

Validation plugins

Malware detector

This plugin provides robust security features by leveraging the ClamAV antivirus engine to scan files for potential threats, including trojans, viruses, malware, and other malicious content. ClamAV is a trusted, open-source (GPL) antivirus engine that is widely used in the industry for its exceptional accuracy and effectiveness in detecting threats.

Digitization profile validator for TIFF images

This plugin checks if the images produced through digitization processes meet the expectations defined in a digitization profile. The digitization profile typically outlines rules and guidelines for minimum DPI resolution, compression type, photometric interpretation, and other technical aspects of the image file format.

Format Validator for PDF/A

The Format Validator for PDF/A is a specialized tool designed to ensure compliance with the ISO-standardized Portable Document Format (PDF) specification for archival and long-term preservation of electronic documents. This plugin validates PDF files against the PDF/A specification, which imposes restrictions and requirements on the "base" PDF standards, including PDF 1.4 for PDF/A-1 and ISO 32000 for PDF/A-2 and PDF/A-3, as well as a set of additional third-party standards.

Conversion plugins

Image Converter

The Image Converter plugin harnesses the power of ImageMagick, a leading image manipulation tool, to effortlessly convert between over 200 different image formats including PNG, JPEG, JPEG-2000, GIF, TIFF, DPX, EXR, WebP, Postscript, PDF, and SVG.

Office Documents Converter

The Office Documents Converter is a versatile plugin that utilizes the "unoconv" (Universal Office Converter) technology to convert a wide range of office file formats. The supported formats include Open Document Format (odt), Microsoft Word (doc), Microsoft Office Open/Microsoft OOXML (ooxml), Portable Document Format (pdf), HTML (html), XHTML (xhtml), Rich Text Format (rtf), Docbook (docbook), and many others.

Video Converter

The Video Converter is a powerful plugin that leverages the capabilities of "avconv," a high-speed video and audio conversion tool. This converter can perform arbitrary sample rate conversions and resize video in real-time with a high-quality polyphase filter. The plugin allows for the conversion of files containing a variety of different stream types, including video, audio, subtitles, attachments, and data.

Audio Converter

The Audio Converter is a highly effective plugin that leverages the capabilities of "SoX" (Sound eXchange tool) a versatile cross-platform tool for audio file format conversion. With this plugin, users can convert audio files from one format to another and apply a variety of advanced effects such as volume adjustments, equalization, reverb, delay, chorus, flanging, tempo and pitch changes.

Digital signature plugins

Digital Signature Validator

The Digital Signature Validator performs a comprehensive evaluation of embedded digital signatures within files to ascertain their validity.

DIP Digital Signature Creator

The DIP Digital Signature Creator plugin is a powerful tool that enables users to generate a new Dissemination Information Package (DIP) for a specified Archival Information Package (AIP). The DIP contains all the files from the AIP, digitally signed with the repository's digital certificate.

Digital Signature Expiry Date Extractor

The Digital Signature Expiry Date Extractor plugin obtains expiration dates from qualified digital signatures embedded in PDF files and saves them in metadata.

Digital Signature Expiry Date Extender

The Digital Signature Expiry Date Extender uses a technique called Long-Term Validation (LTV) to ensure the integrity and authenticity of digital objects over an extended period of time.

Risk assessment plugins

File Integrity Verifier

The File Integrity Verifier plugin computes the fixity/checksum information of files inside an Archival Information Package (AIP) and verifies if this information differs from the information stored in the preservation metadata. If so, it creates a new risk and assigns the corrupted file to that risk in the Risk register.

Representation Information Broken Links Verifier

The Representation Information Broken Links Verifier plugin is a valuable tool for verifying the accuracy and accessibility of external links referenced in Representation Information Records.

Incomplete Representation Information Detector

The Incomplete Representation Information plugin is a powerful tool that can help ensure the completeness and accuracy of the representation information for digital files.

Risk Incidence Creator

The Risk Associator plugin associates selected items (AIPs, Representations or Files) to existing risks in the Risk registry (as risk incidences). This action is convenient when the preservation expert wants to associate a set of items to a risk to be mitigated in the near future.

Incomplete File Format Detector

The Incomplete File Format Detector plugin verifies if a file has a complete format information, with a MIME type, PRONOM ID, or a Format designation. If this information is missing, it creates a new risk entry in the Risk Register and assigns the file in question to that risk.

eArchiving plugins

E-ARK AIP Validator

The E-ARK AIP Validator plugin provides a comprehensive evaluation to ensure that AIPs meet the requirements outlined in the E-ARK specification, version 2.x.

E-ARK AIP Manifest Updater

For performance reasons, RODA does not keep updated versions of the METS manifest prescribed by the E-ARK AIP specification. The E-ARK AIP Manifest Updater plugin creates, or updates METS manifest files based on AIP information found in the storage layer.

E-ARK DIP Creator

Create an E-ARK DIP by selecting the metadata and representations we want from the AIP. The result is a RODA DIP which contains a E-ARK DIP as defined in the standard specification.

And many other plugins

Database preservation

Completeness check (inventory report)

Find and Replace

AIP Batch exporter

Index rebuild

Activity log truncator

…

External plugins

Find more actions in the Marketplace, develop your own actions, share with the community.

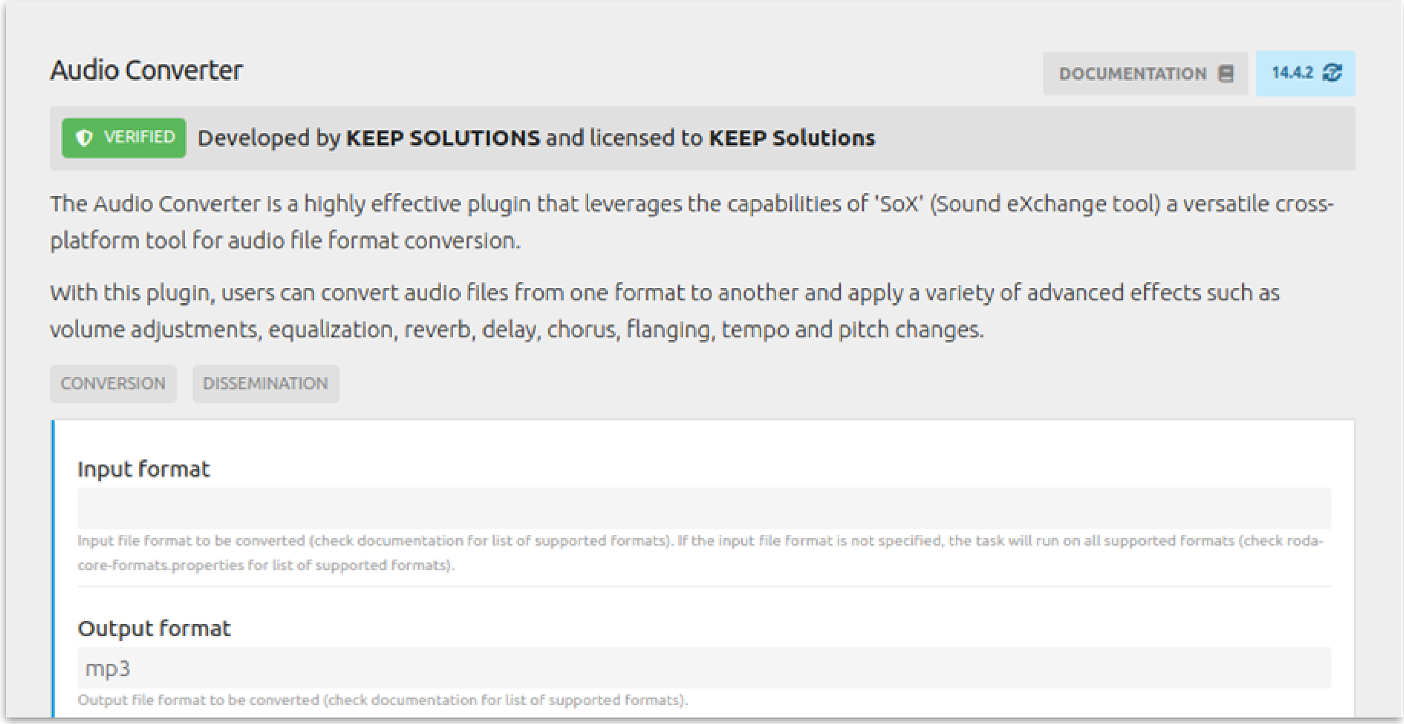
Trusted preservation actions

Digitally signed external plugins

Verified against truststore

License and documentation

Version update indicator



Developed by **KEEP SOLUTIONS** and licensed to **KEEP Solutions**

The Audio Converter is a highly effective plugin that leverages the capabilities of 'SoX' (Sound eXchange tool) a versatile cross-platform tool for audio file format conversion.

With this plugin, users can convert audio files from one format to another and apply a variety of advanced effects such as volume adjustments, equalization, reverb, delay, chorus, flanging, tempo and pitch changes.

Input format

Input file format to be converted (check documentation for list of supported formats). If the input file format is not specified, the task will run on all supported formats (check roda-core-formats.properties for list of supported formats).

Output format

mp3

Output file format to be converted (check documentation for list of supported formats).

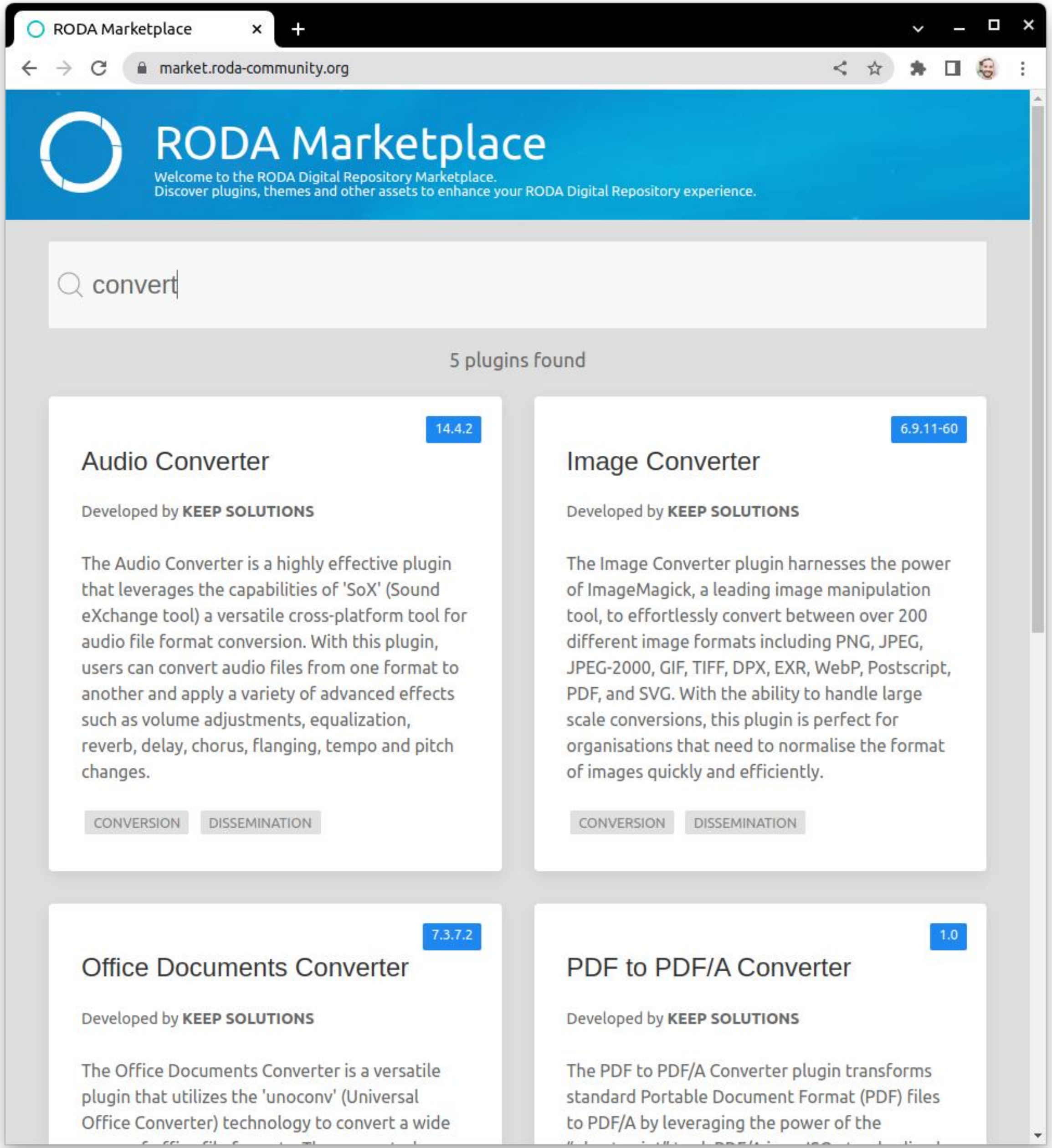
RODA Marketplace

https://marketplace.roda-community.org

Find free and commercial plugins made available by contributors.

Soon to include:

- Components
  External software that integrates with RODA, like external authentication, authorization, monitoring, reporting, etc.
- Services
  Like integration, maintenance and support, hosting, consulting, etc.



RODA Marketplace

Welcome to the RODA Digital Repository Marketplace.
Discover plugins, themes and other assets to enhance your RODA Digital Repository experience.

convert

5 plugins found

14.4.2

Audio Converter

Developed by **KEEP SOLUTIONS**

The Audio Converter is a highly effective plugin that leverages the capabilities of 'SoX' (Sound eXchange tool) a versatile cross-platform tool for audio file format conversion. With this plugin, users can convert audio files from one format to another and apply a variety of advanced effects such as volume adjustments, equalization, reverb, delay, chorus, flanging, tempo and pitch changes.

CONVERSION DISSEMINATION

6.9.11-60

Image Converter

Developed by **KEEP SOLUTIONS**

The Image Converter plugin harnesses the power of ImageMagick, a leading image manipulation tool, to effortlessly convert between over 200 different image formats including PNG, JPEG, JPEG-2000, GIF, TIFF, DPX, EXR, WebP, Postscript, PDF, and SVG. With the ability to handle large scale conversions, this plugin is perfect for organisations that need to normalise the format of images quickly and efficiently.

CONVERSION DISSEMINATION

7.3.7.2

Office Documents Converter

Developed by **KEEP SOLUTIONS**

The Office Documents Converter is a versatile plugin that utilizes the 'unoconv' (Universal Office Converter) technology to convert a wide

1.0

PDF to PDF/A Converter

Developed by **KEEP SOLUTIONS**

The PDF to PDF/A Converter plugin transforms standard Portable Document Format (PDF) files to PDF/A by leveraging the power of the

European Commission

Why create your own RODA plugins?

With plugins you can:

- Support your own SIP formats
- Add your own ingest workflows and ingest steps
- Add your own preservation actions
- Integrate with your own services

How to create your own RODA plugins?

To create new plugins and use them to RODA it is necessary to:

1. Create a new plugin project
   See the RODA plugin template
2. Build the plugin and deploy
   All instructions in the template
3. Publish plugin in market
   Follow instructions to gather and submit external plugin metadata

Any questions?

Next:

- ➔ Live demo

Thank you for joining us

Contact details

https://e-ark4all.eu/

info@e-ark-foundation.com

EU_eArchiving #eArchivingIsBack

https://www.linkedin.com/company/eu-earchiving-initiative

https://www.youtube.com/@e-ark