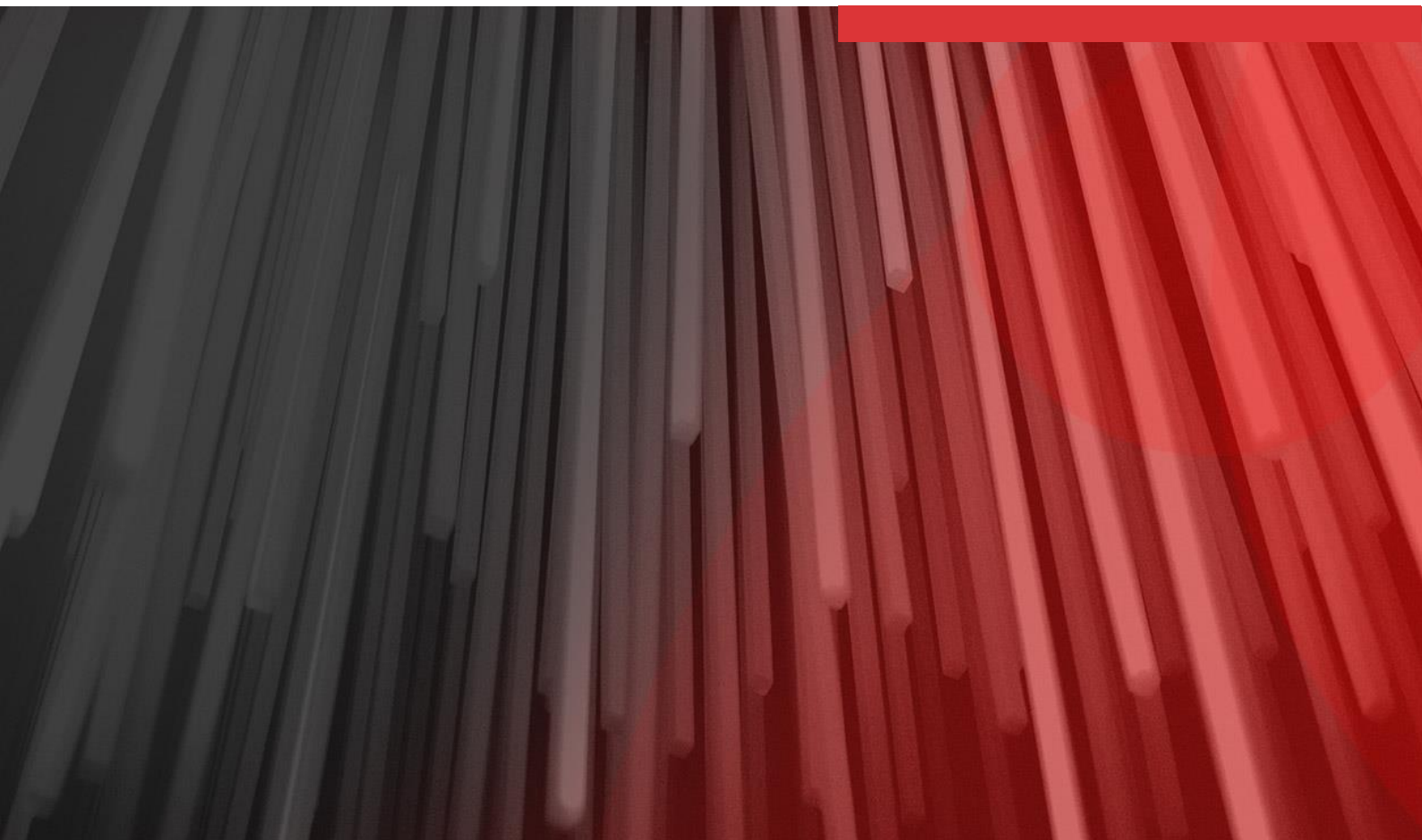


RODA 5 ENTERPRISE



LONG-TERM DIGITAL PRESERVATION

CHARACTERISTICS AND TECHNICAL
REQUIREMENTS

ABOUT THIS DOCUMENT

Identifier	WP232092		
Approved by	Luís Miguel Ferros	Approved on	2023-10-11
Classification	Public		
Distribution	N/A		

REVISIONS

#	Date	Authors	Modifications
1	2023-01-25	Miguel Ferreira	First version of the document
2	2023-02-25	Luís Faria	Plugins descriptions
3	2023-02-28	Miguel Ferreira	Broad review
4	2023-03-01	Luís Faria	Final adjustments
5	2023-03-03	Miguel Ferreira	Updated info about Representation Information Broken Links plugin
6	2023-03-13	Miguel Ferreira	Updated architecture image
7	2023-03-20	Miguel Ferreira	Added cloud hosting option
8	2023-03-21	Miguel Ferreira	Correct the naming of E-ARK specs
9	2023-03-27	Miguel Ferreira	Added the inventory report
10	2023-04-24	Miguel Ferreira	Added new edition
11	2023-04-26	Luís Faria	Final review
12	2023-04-26	Luís Faria	Fixed typo
13	2023-05-02	Luís Miguel Ferros	Fixed the inclusion of plugins on RODA SM edition
14	2023-05-02	Miguel Ferreira	Removed second line of support
15	2023-05-22	Luís Faria	DIP Creator plugin description review.
16	2023-05-22	Miguel Ferreira	Typo
17	2023-05-26	Miguel Ferreira	Renamed Activity Log Truncator
18	2023-06-16	Miguel Ferreira	Revision of number of nodes included in the HA edition, warranty extension and technical support.
19	2023-10-11	Jéssica Lopes, Miguel Ferreira	Fixed typo and added information about backup tool on enterprise editions.

EXECUTIVE SUMMARY

RODA is an open-source long-term digital preservation system that delivers functionalities for all the main functional units of the OAIS reference model. RODA is capable of ingesting, managing, and providing access to various types of digital content produced by large corporations and public bodies.

RODA was developed using open-source technologies and it is supported by standards such as the ISO 14721 - Open Archival Information System (OAIS), ISO 16363 - Audit and certification of trustworthy digital repositories, Metadata Encoding and Transmission Standard (METS), Encoded Archival Description (EAD), Dublin Core (DC), E-ARK Information Package specifications and PREMIS (Preservation Metadata).

This document describes the existing RODA editions and the value propositions associated with each of them. The document also outlines the technical requirements and what is necessary to deploy the software in a production environment.

**PRESERVE AND
PROVIDE
ACCESS
TO DIGITAL
ASSETS
PRODUCED BY
LARGE
ORGANISATIONS**

RODA

RODA (Repository of Authentic Digital Records) is a long-term digital repository solution that delivers functionalities for all the main functional units of the OAIS reference model. RODA is capable of ingesting, managing, and providing access to various types of digital content produced by large corporations and public bodies.

RODA was developed using open-source technologies and it is supported by standards such as the Open Archival Information System (OAIS), Metadata Encoding and Transmission Standard (METS), Encoded Archival Description (EAD), Dublin Core (DC), E-ARK Information Package specifications and PREMIS (Preservation Metadata).

It implements multiple ingest workflows that not only validate Submission Information Packages (SIP), but also checks their content for virus, handle file format identification, extract technical metadata, and migrate file formats to more “preservable” alternatives.

RODA enables users to access digital information in several forms, such as searching and browsing via an online catalogue, as well as providing REST APIs for systems integration. Discovery services are provided over both descriptive metadata and textual content (automatically extracted from an array of document-based formats). Online consultation of ingested objects, preservation formats and dissemination derivatives are also possible via the repository catalogue.

Administration interfaces allow repository managers to maintain representation information updated, manage risks and to execute preservation actions on all digital assets available in the repository.

RODA ensures that ingested data remains authentic by recording PREMIS metadata for every action performed on a digital object. It records provenance information in archival metadata standards such as EAD or Dublin Core and ensures integrity and availability by frequently monitoring data and making sure that it has not been tampered with. All interactions between users and the repository (human and software) are logged for security and accountability reasons.

VALUE PROPOSITIONS

Conforms to **open standards**

RODA is compliant with several open descriptive metadata standards such as EAD, and Dublin Core, PREMIS for preservation metadata and METS for structural metadata.

It can support more standards by means of an advanced templating system that supports validation, indexing, viewing, and editing of descriptive metadata.

SIP, AIP and DIP formats are also based on open specifications managed by the DILCIS Board to avoid technology lock-in (i.e., EARK Information Packages).

Vendor **independent**

RODA is built on top of enduring open-source technologies.

The entire infrastructure required to support RODA is also vendor independent. This means that you may use the hardware and the Linux distribution that best fit your institutional needs.

Because the product itself is open source, you don't have to rely on a single vendor for support. The entire source code of the system is available on GitHub for inspection.

Authenticity

RODA uses PREMIS preservation metadata to create a trust chain between all generations of digital assets.

Preservation metadata, together with the establishment of trust of its surrounding environment (ISO 16363 and ISO 14721) ensures that the repository service is reliable, and digital assets remain authentic over time.

RODA also supports plugins that assess the validity of digital signatures and re-sign archived PDFs when the lifetime of digital signatures is coming to an end.

Support for **multiple formats**

RODA can ingest all sorts of content independent of its format. Format migration plugins are available to cope with decaying formats such as text documents, raster images, relational databases, video, and audio.

Normalization of file formats is possible by setting up adequate ingest workflows. Files are then preserved in formats more adequate for long-term archiving.

Representation information networks can be managed within the repository itself, letting you opt for the right preservation strategy at the right time.

Advanced ingest workflows

RODA supports the ingest of digital material as well as any associated metadata in several distinct formats. Tools and libraries are provided to enable Producers to create packages in the supported Submission Information Package formats (SIP). Ingest workflows can be customised by the user to implement institutional policies and handle special collections of data. Ingest workflows can be enhanced via the use of open source or commercial plugins.

Embedded **preservation actions**

Preservation actions can be executed right from the user interface over any selection of digital objects in the repository.

The task execution engine enables the repository to parallelise the task execution process to take full advantage of the existing CPUs.

Preservation actions include format conversions, checksum verifications, virus checks, various, risk assessment, etc.

Scalable

RODA is ideal for large organisations with millions of digital objects. Its service-oriented nature allows it to be highly scalable, enabling the distribution of load between several servers.

The use of horizontally scalable indexing services enables discovery services to be spread across multiple servers for greater performance.

RODA also takes advantage of all the CPUs available in each server to mass process thousands of objects simultaneously for stellar performance.

Copes with the rapid changing nature of technology

The pluggable architecture of RODA makes it easy to add more functionality to the system without affecting its core functionality.

This includes adding new preservation capabilities such as format converters, risk assessment tools, and more.

The system stores data in a well-documented open AIP format that can be easily inspected by users and processed by third-party repository systems.

Retention and disposal policies

RODA incorporates robust and comprehensive functionality for the definition and implementation of retention and disposal policies.

This feature empowers users to establish clear and concise guidelines for determining how long records need to be kept and how they should be disposed of, in accordance with national legislation or internal regulations.

With RODA's retention and disposal policies, organizations can confidently preserve their valuable digital assets for the long-term while mitigating risks associated with data privacy and security.

Integration with data production systems

RODA exposes all its functionality via well-documented REST API. Convenient Java libraries are available on GitHub to allow developers interact with RODA via its Core APIs. Several tools exist to create and manipulate SIPs and submit them to RODA's for ingest.

In fact, RODA's ability to integrate data from other records management systems include: 1) packaging data available on a filesystem via RODA-in tool, 2) extract data from relational databases using DBPTK, and 3) custom integrations with original systems via APIs or using drop folders.

RODA EDITIONS

RODA is an open-source solution, which means that anyone can download its source code, compile it and have it running on their own institution in a matter of hours. This edition is called RODA COMMUNITY.

RODA Community Edition is designed for mid-sized organizations seeking to implement and customize the product to meet their specific requirements. As its name suggests, this edition is community-driven and provides organizations the necessary tools to manage the product on their own. While this flexibility is a major advantage, organizations may also benefit from enlisting the expertise of IT, software, and digital preservation specialists to tailor the solution to their unique needs by adding specialized features and integrations.

KEEP SOLUTIONS offers three software distributions specially designed to meet the rigorous demands of mid to large-sized institutions with vast collections of digital records, providing the capability to efficiently manage digital assets in large-scale production environments. These software suites are called “Enterprise”, “Enterprise HA” and “Enterprise SM”.

RODA ENTERPRISE

For organizations seeking a long-term digital repository coupled with a comprehensive set of components and services for secure and high-performance production environments.

This edition includes commercial plugins and professional support services to assist organisations with deployment and ongoing operations to ensure optimal performance, stability, and security.

RODA ENTERPRISE HA

For organizations seeking uninterrupted access to digital assets and the ability to scale their preservation environment without causing downtime. RODA Enterprise HA Edition delivers high-availability and horizontal scalability.

This offering includes unlimited scalability and the ability to upgrade the system with minimal disruption, ensuring organizations always maintain seamless access to their critical digital assets.

RODA ENTERPRISE SM

For organizations seeking comprehensive control over their infrastructure and operations, while simultaneously harnessing the benefits of warranty, security patches, commercial plugins, components, and professional services.

Under this model, the organization's IT department assumes the critical responsibility for implementing, monitoring, and maintaining the system. Shadowing support can be enlisted to assist with maintenance and support tasks. Moreover, for additional features, custom development services can be engaged, ensuring that the system continues to evolve to meet the ever-changing demands of the business.

By minimizing reliance on third-party providers for installation, configuration, and ongoing support, this solution effectively preserves data security and system integrity.

COMMUNITY VS. ENTERPRISE EDITIONS

The primary difference between the Community and Enterprise editions lies in the comprehensive set of services and features offered in the Enterprise editions. The Community edition includes the software without any professional services such as branding, deployment, maintenance, and training. Furthermore, it does not encompass any supplementary components deemed essential for large-scale production environments, such as support for Enterprise Identity Providers (LDAP, ADFS, SAML, Azure Active Directory, etc.), infrastructure monitoring or advanced reporting services.

Smaller organizations may not require advanced features such as single sign-on (SSO), remote data replication or advanced reporting capabilities. However, larger organizations have other concerns such as the scalability of the storage layer, ensure uninterrupted access to data, security, and certification, which are critical for governmental and enterprise business operations.

The Enterprise editions of the software are specifically designed for large-scale public and private organizations and come with pre-configured deployment arrangements that are finely tuned to ensure unparalleled standards of compliance, security, availability, and data integrity.

Additionally, the suite of professional services and advanced components included in the Enterprise editions guarantees maximum operational efficiency, allowing organizations to focus on their core operations without worrying about the technical aspects of software implementation and maintenance.

WHAT ARE THE BENEFITS OF CHOOSING AN ENTERPRISE EDITION

If RODA is open source, why pay for an Enterprise edition? The Community Edition serve as an excellent option for those who are experienced with deploying open source software or simply want to explore and experiment. However, if you are seeking to support a service over an extended period, the Community Editions may not be the most ideal option.

Enterprise Editions are tailored to meet the needs and requirements of large public and private organisations. By selecting an Enterprise edition, you will gain access to patches, upgrades, expert technical support, and trained resources. In contrast, the Community edition rely on forum-based support from the community of open source developers, and release cycles that not always align with institutional needs.

Furthermore, with a commercially supported distribution, you will enjoy the benefits of the latest open source innovations along with the stability and support that an enterprise requires. The software's team of engineers works to enhance features, reliability, and security, guaranteeing that your infrastructure remains steady and performs optimally, regardless of your use case and workload.

The following table summarises the set of features included in each RODA edition.

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
Standard features				
ISO 14721 compatibility (OAIS)				
All RODA suites are ISO 14721 (OAIS) compatible, providing support for ingest, archival storage, data management, access and dissemination, administration, preservation planning, risk management, representation information and more.	✓	✓	✓	✓
Online catalogue				
Seamless access to digital assets through an advanced online access catalogue with multi-language support. This catalogue features robust search functionality and integrated viewers for various file formats, including images, text, audio, and video, allowing users to easily find and view the assets they need.	✓	✓	✓	✓
E-ARK SIP and BagIt				
Support for E-ARK and BagIt Submission Information Packages (SIP). This includes thorough validation of SIPs during the ingest process, as well as the ability to transform SIPs into Archival Information Packages (AIPs).	✓	✓	✓	✓
Retention and disposal				
Allows users to establish clear guidelines for how long records need to be kept, and how they should be disposed of, in accordance with national legislation or internal regulations.	✓	✓	✓	✓
Warranty, security updates and backups				
Enterprise Editions come with 1-year warranty for all software components and plugins included. Enterprise editions also include free security updates and bug fixes throughout the entire duration of the warranty. Tools are included that not only help automate the backup process but also provide a comprehensive safety net for all entities housed within the repository.	-	✓	✓	✓
Commercial plugins				
Plugins are specialized preservation actions designed to augment the functionality of RODA. These may include format validators, feature extractors, format converters, risk assessment tools, and more. Although plugins are often supported by open-source tools, commercial plugins exist that have been thoroughly tested and optimized for parallel execution to ensure optimal performance. For a full list of included plugins in each edition of RODA, take a look at the Plugins section below.	-	✓	✓	Sold separately
Scalability				
Support for parallel execution of plugins, which allows for efficient processing and handling of large data sets. For organizations with high-volume processing needs, the RODA can be deployed in a cluster environment, which unlocks the power of horizontal scalability. This allows for unlimited scalability, enabling the system to handle an unlimited number of records, providing a flexible and professional solution for organizations with varying data processing needs.	Vertical Up to 10M records	Vertical Up to 10M records	Horizontal More than 10M records	Configurable
High availability				
RODA can be deployed in a cluster environment, which unlocks the power of horizontal scalability and high availability. The platform is designed to be resilient to failing hardware and software and to allow upgrades to be performed without interrupting the service.	-	-	5 nodes	Configurable

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
Advanced components				
Drop folders				
Facilitates the automated, unsupervised ingestion of submission information packages via shared folders, which is crucial for a smooth integration with other data production systems.	-	✓	✓	✓
Folder scattering				
Ensures that files are evenly distributed across the file system to prevent over-utilization of any specific storage area and allows for the extension of storage capacity without interruption to ongoing operations.	-	✓	✓	✓
Remote replication				
Disaster recovery service that replicates information packages to remote data centres. This ensures that in the event of an unexpected event, such as natural disasters or cyber-attacks, critical data is still available, minimizing the disruption to the business operations.	-	✓	✓	Sold separately
Inventory Report Comparator				
The Inventory Report Comparator is a tool that helps users compare inventory reports from different time periods or produced by different systems. This enables the easy identification of any changes or discrepancies in the repository's content over time or to cross check if the entire content of an ingest batch is has been correctly archived. This tool simplifies the process of validating the completeness and correctness of the repository content by comparing it to previous inventory reports generated during pre-ingest.	-	✓	✓	Sold separately
Identity and Access Management				
Supports the integration of various authentication protocols, including LDAP, SAML, OAuth, Azure AD, EU Login, Citizen Card, and more. This allows organizations to authenticate both users and services in a secure and flexible manner, ensuring that only authorized individuals and systems can access sensitive information.	-	✓	✓	Sold separately
Reporting				
Customizable dashboards for organizations to support informed decision-making and strategic planning. This feature allows for tracking of key performance indicators (KPI), risk analysis, and other important data assessment tasks, providing management with the information necessary to make informed decisions and drive the success of the preservation process.	-	✓	✓	Sold separately
Monitoring				
Load and storage monitoring capabilities with advanced dashboard functionality. Customizable alert notifications allow for proactive monitoring and ensure optimal performance and efficient resource utilization. This provides organizations with valuable insights into the health and performance of their system, allowing them to take proactive measures to maintain high levels of availability and efficiency.	-	-	✓	Sold separately

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
Professional services				
Deployment in production environment				
Includes branding, configuration, and deployment using Docker containers, ensuring a smooth and efficient rollout. The goal of this offering is to provide organizations with a turnkey solution for deploying their system in a production environment, minimizing the time and resources required for deployment and ensuring the success of their operations.	-	✓	✓	-
Technical support (1 year)				
Comprehensive range of support services to ensure the smooth operation of the system. This includes a helpdesk, service level agreement (SLA), ticketing platform, security patches, availability monitoring, and user access statistics. These services ensure that organizations have access to the resources and support they need to effectively manage and maintain their systems, maximizing their investment and ensuring the success of their operations.	-	✓	✓	-
Three-year technical support contracts include an upgrade to a major version within that time frame.				
Training				
Provides operators and administrators with the necessary knowledge and skills to effectively manage and operate the system. The goal of this training is to equip individuals with the necessary knowledge and skills to ensure the success of their operations and maximize the investment in the system. This includes in-depth training sessions, delivered in either English or Portuguese. <u>Please note that traveling costs are not included in this offering.</u>	-	✓	✓	Sold separately
Custom development				
This service is designed to provide organizations with the flexibility and control they need to achieve their business objectives by creating new plugins or integrations that are tailored to their specific requirements.	-	80 hours	160 hours	Sold separately
The service covers the full software development life cycle, from initial design and planning to coding, testing, and implementation.				
Shadowing support				
Shadowing support refers to a type of technical assistance where a support specialist "shadows" a client's support engineer as they work through an issue. This allows the specialist to provide real-time guidance, advice, and troubleshooting to help resolve the issue. The support specialist observes the client's work but does not take control of the process. Instead, they work in collaboration with the client's support engineer to resolve the issue together. Shadowing support is often used when an external support team is not granted direct access to the infrastructure for security and privacy reasons.	-	-	-	Sold separately

PROFESSIONAL SERVICES EXTENSIONS

Enterprise editions of RODA offer a range of professional services to support organizations in their digital preservation efforts. These services may be extended to better meet the needs of individual organizations. This provides organizations with the flexibility and control they need to effectively manage and maintain their digital preservation system, supporting their operations and driving success.

With the option to add additional services, organizations can customize their solution to meet their unique requirements and maximize their investment in digital preservation.

The following table describes the services that may be added to the services already included in the Enterprise editions.

	Includes
Service extensions	
Custom metadata schema Provides organizations with the ability to add support for an additional XML-based descriptive metadata schema. This includes support for validation, indexing, viewing, and editing of the new schema, ensuring that organizations can effectively manage and utilize their metadata. The goal of this offering is to provide organizations with the flexibility and capability to manage their metadata in a manner that best meets their needs, supporting their operations and driving success.	1 schema
Deployment in acceptance environment Includes the deployment of the digital preservation suite in an acceptance environment, ensuring that organizations have a safe and controlled environment for testing and training purposes, safeguarding the success and smooth operation of their production environment.	1 environment
Technical support Additional one year of technical support. The goal of this offering is to provide organizations with the resources and support they need to effectively manage and maintain their system, ensuring its success and maximizing their investment. If the technical support service accumulates to three consecutive years, the service includes an upgrade to a major version of the preservation suite within that time frame.	1 year
Warranty & security updates One additional year of warranty and security updates for all the software components and plugins included in the preservation suite.	1 year
Additional High Availability Nodes Includes the deployment of two additional nodes in the RODA Enterprise HA cluster. It's important to note that additional nodes must be added in pairs to ensure that the cluster always has an odd number of nodes. This is crucial because it enables the cluster to maintain quorum, which is necessary for it to function effectively. Adding these new nodes will enhance the resilience and scalability of the cluster, providing improved performance and reliability for the organization's critical systems and applications."	2 nodes
Custom development package Provides organizations with the ability to customize the digital preservation suite to meet their specific needs and requirements. This service covers the entire software development life cycle, from initial design and planning to coding, testing, and implementation. Organizations will have access to experienced developers who will work with them to understand their requirements and deliver custom solutions that are tailored to their specific needs. The goal of this offering is to provide organizations with the flexibility and control they need to achieve their business objectives, supporting their operations and driving success.	40 hours
Database archiving	SIARD support

Provides organizations with the ability to preserve relational databases and their contents in a long-term, open, non-proprietary, platform-independent, and self-describing format. This includes the implementation of additional features to support the ingest, validation, and access of relational databases in SIARD archival format.

Shadowing support

Shadowing support refers to a type of technical assistance where a support specialist "shadows" a client's support engineer as they work through an issue. This allows the specialist to provide real-time guidance, advice, and troubleshooting to help resolve the issue. The support specialist observes the client's work but does not take control of the process. Instead, they work in collaboration with the client's support engineer to resolve the issue together. Shadowing support is often used when an external support team is not granted direct access to the infrastructure for security and privacy reasons.

40
hours

Digital preservation planning

A digital preservation plan is a document or set of documents outlining the steps and strategies an organization will take to ensure the long-term accessibility and preservation of its digital assets. It typically includes information about the types of digital content to be preserved, the technology and standards to be used, policies for managing and preserving the content, and procedures for regular review and updates. The goal of a digital preservation plan is to ensure that digital information remains usable and understandable over time, despite the rapid pace of technological change. The preservation plan included in this offer only include one system, typically the preservation system.

1
preservation
plan

PLUGINS

RODA plugins consist of specialized actions designed to augment the functionality of RODA. These may include format validators, feature extractors, format converters, risk assessment tools, and more. Although plugins are often supported by open-source tools, they have been thoroughly tested and optimized for parallel execution to ensure optimal performance. This section outlines the main plugins that are included in each edition of RODA.

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
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. This information can be used to determine the appropriate software to open and manipulate the file, as well as to ensure compatibility with different systems and applications.	✓	✓	✓	✓
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. The tool supports numerous file formats, including images, raw image formats from digital cameras from several manufacturers, DICOM metadata, Photoshop tags, metadata from audio and video formats, and metadata from office documents such as Microsoft Word, Excel, PowerPoint, or OpenDocument formats. With this plugin, users can quickly and easily access critical technical metadata, providing them with valuable insights into the contents of their files.	-	✓	✓	Sold separately
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.	-	✓	✓	Sold separately
Optical Character Recognition Extractor				
The Optical Character Recognition Extractor is a powerful plugin designed to extract typed or printed text from digitalized images, making it an essential tool for professionals in various fields, including data analysis, document management, and research. The plugin supports a wide range of image formats and can extract text from images with a minimum resolution of 300 DPIs, low noise, and minimal dilation, erosion, rotation, and skew.	-	✓	✓	Sold separately
While the plugin does not support handwritten text, it offers a range of advanced features for enhancing the quality of the extracted text. These include image pre-processing, specialized dictionaries, word lists, patterns, and trained data. By leveraging these features, users can greatly improve the accuracy and reliability of the text extraction process, ensuring that the extracted text is of the highest quality. Once extracted, the text is then available for search. The ability to extract text from digitalized images enables users to unlock valuable information that may otherwise have remained inaccessible.				

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
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. The goal of a digitization profile validator is to ensure that digital images are of sufficient quality and have the necessary technical attributes for preservation and accessibility over time.	-	✓	✓	Sold separately
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. The PDF/A specification prohibits features that are not conducive to long-term archiving, such as font linking, and encryption.	-	✓	✓	Sold separately

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
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. 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.	-	✓	✓	Sold separately
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. This tool provides a convenient and efficient solution for converting office documents to a desired format.	-	✓	✓	Sold separately
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. The number and type of streams are determined by the container format, and the plugin can enforce these limitations as necessary.	-	✓	✓	Sold separately
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.	-	✓	✓	Sold separately

COMMUNITY

ENTERPRISE

ENTERPRISE
HA

ENTERPRISE
SM

Digital signature plugins

Digital Signature Validator

The Digital Signature Validator performs a comprehensive evaluation of embedded digital signatures within files to ascertain their validity. This task supports verification of digital signatures in the following file formats: PDF, Microsoft Office Formats (such as .docx, .xlsx, and .pptx), and OpenDocument formats (such as .odt, .ods, and .odp).

Upon completion of this process, three significant outcomes are achieved: Firstly, the outcome of the verification is recorded and documented in a PREMIS event. Secondly, the extracted digital signatures are securely stored within the Archival Information Package (AIP) in the designated "metadata/other" folder. Finally, the original files with the digital signatures removed are securely stored in a new representation within the AIP.

-

✓

✓

Sold separately

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. The extracted dates provide a convenient means of sorting and selecting files based on signature expiration date, enabling proactive management of signature validity, and ensuring the ongoing authenticity of the digital content in combination with the Digital Signature Extension Plugin.

-

✓

✓

Sold separately

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. LTV works by verifying the digital signature of a digital object at regular intervals to ensure that the object has not changed since its original creation. The plugin uses the mLTV service provided by Multicert to extend the validity of document's digital signatures by applying timestamps and integrated status information to the content of a PDF document. The mLTV is an external webservice provided by Multicert. As such it requires a service contract from Multicert to be used.

-

✓

✓

Sold separately

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.

By digitally signing the files in the DIP with the repository's digital certificate, the plugin helps ensure that the files are authentic and trustworthy. Depending on regional legislation, the files in the DIP may be considered certified copies of the original digital assets, further enhancing their authenticity and reliability. Using the DIP Digital Signature Creator plugin, users can create a secure and reliable package of digital assets that can be easily disseminated to others. This is especially important for organizations and individuals who need to ensure that their digital assets are trustworthy and can be relied upon for legal or regulatory purposes.

-

✓

✓

Sold separately

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
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. The plugin also creates an incidence linked to the representation if a PREMIS file exists, but the associated file does not. Within the repository, fixity checking is used to ensure that digital files have not been affected by data rot or other digital preservation dangers. By itself, fixity checking does not ensure the preservation of a digital file. Instead, it allows a repository to identify which corrupted files to replace with a clean copy from the producer or from a backup.	✓	✓	✓	✓
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. As an example, if the designated community of the repository provides feedback that a given format under a certain collection is not being displayed properly on the graphical user interface of the repository, then the preservation expert may want to mark these files to be targeted by a preservation action (e.g., generate new representations for access purposes).	✓	✓	✓	✓
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. Representation Information can point to external links for additional information, but these web pages can cease to exist, leading to broken links and potentially lost or inaccessible information.	-	✓	✓	Sold separately
The plugin checks if the links being pointed by the Representation Information Records are still accessible and optionally removes them in case they are broken.				
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.	-	✓	✓	Sold separately
Furthermore, by creating a new risk entry for files with missing format information, the plugin helps organizations track and manage the risks associated with incomplete file format information. This can help organizations prioritize their efforts and resources for addressing these risks and improving the completeness and accuracy of their digital assets' representation information.				
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. The plugin verifies that each file has its format described in the Representation Information Network, and that it is marked as a known or supported format. Additionally, the plugin checks that all configured attributes for Archival Information Packages (AIPs), Representations, and Files are accurately described in the Representation Information Network.	-	✓	✓	Sold separately
By verifying that each file has its format described in the Representation Information Network, the plugin helps ensure that the files can be accurately interpreted and preserved over time.				

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
Maintenance plugins				
Index Rebuild Actions				
Rebuild indexed information from data kept in storage. All indexes can be rebuilt or refreshed from information persisted in storage. This may also be necessary when updating the indexing system. Additional options include clearing index before re-index (full re-index) and to optimize indexes.	✓	✓	✓	✓
Inventory Report Creator				
The Inventory Report Creator plugin automates the generation of a detailed Inventory Report in CSV format for all AIPs and their corresponding files (both data and metadata) within a repository. The report includes technical information such as SIP ID, AIP ID, representation ID, file path, and cryptographic hash values such as SHA-256, MD5, and SHA-1. This information can be used to validate the completeness and correctness of the repository content by comparing it to previous inventory reports generated during pre-ingest.	✓	✓	✓	✓
AIP Batch Exporter				
The AIP Batch Exporter is a powerful plugin that allows you to select a group of AIPs and export them as a single ZIP file or folder. The outcome is saved on the server file system, which can be accessed by users with the necessary permissions.	✓	✓	✓	✓
This plugin is particularly useful when you need to export a large number of AIPs based on specific search criteria. With just a few clicks, you can create a batch export of AIPs and download them in a compressed format. This saves time and effort and ensures that your AIPs are stored and transferred efficiently.				
Audit Log Truncator				
The Audit Log Truncator removes all entries from the audit log that are older than a specified number of days. The log is preserved as external physical files, however older entries will not be displayed in the graphical user interface. To access older log entries, one needs access to the storage layer of the repository server. Audit log truncation automatically frees index space and improves performance of the repository as a whole.	✓	✓	✓	✓
Permissions Manager				
The Permissions Manager plugin enables users to update permissions of selected AIPs by adding or removing groups, users or setting specific permissions such as CREATE, READ, UPDATE, DELETE, GRANT.	-	✓	✓	Sold separately
Find and Replace				
The Find and Replace plugin is a powerful tool for quickly and efficiently updating the descriptive metadata of multiple records at once. Whether you need to make a global change to your metadata or simply correct a typo, this plugin can help you save time and improve the consistency and accuracy of your metadata. With the Find and Replace plugin, you can select all records or a specific set of records in search and replace the content of descriptive metadata attributes. You can choose to replace the entire attribute or search for specific words or phrases to replace. Advanced users can also use regular expressions to match on the content of an attribute, extract groups, and use them in the replacement text.	-	✓	✓	Sold separately
The plugin is highly customizable and flexible, allowing you to fine-tune your search and replace operations to match your specific needs. You can choose to replace all instances of a word or phrase or only those that meet certain criteria, such as occurring in specific fields or having a certain value.				

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
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.	-	✓	✓	Sold separately
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.	-	✓	✓	Sold separately
E-ARK DIP creator				
This plugin creates DIPs containing a selected set of metadata and representations available on the AIP. The outcome is a Dissemination Information Package compliant with the E-ARK specification ready to be delivered to consumers or external applications.	-	✓	✓	Sold separately

	COMMUNITY	ENTERPRISE	ENTERPRISE HA	ENTERPRISE SM
Database preservation plugins				
Database Visualization				
The Database Visualization plugin creates a special purpose DIP and loads SIARD v2 files into an instance of DBPTK for visualisation purposes. This plugin can only be installed if the "Database preservation" Service Extension is also acquired.	-	✓	✓	Sold separately
DBPTK (Database Preservation Toolkit) is an open-source tool for archiving and browsing the contents of relational databases. It is designed to preserve and provide access to databases that are part of a digital archive. The viewing tool also provides a range of features such as database export, saved searches, records concentration diagrams, etc.				

ARCHITECTURE

RODA software is composed of a set of applicational modules, which are represented on image below.

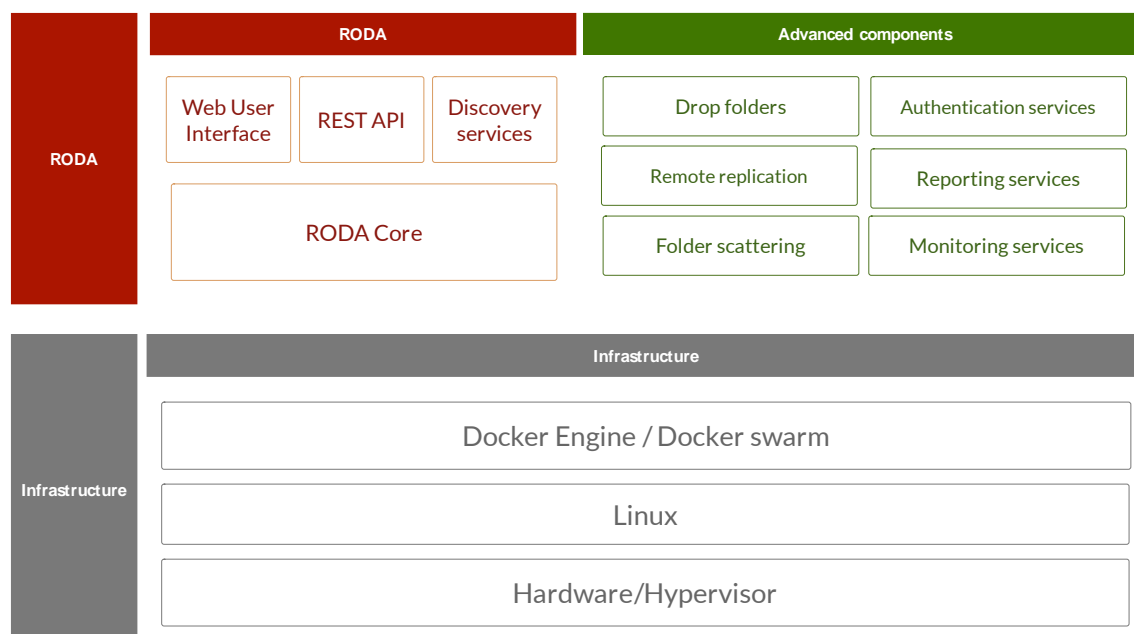


Image 1 – Architecture of RODA

ON-PREMISES VS CLOUD

RODA can be implemented both on-premises, or in the cloud. On-premises implementations involve installing and running the software on servers and infrastructure owned and maintained by the client, while cloud implementations involve using a third-party provider's infrastructure and accessing the software over the internet.

On-premises implementation offers greater control, customization and ownership over the software and its environment, but requires significant upfront investment in hardware, software, and IT staff to manage and maintain the system. Cloud implementation, on the other hand, offers greater flexibility and scalability, as KEEP SOLUTIONS and its cloud partners take care of hardware and software updates and maintenance, but can result in less control and customization options.

Ultimately, the decision to implement RODA on premises or in the cloud will depend on the client's specific needs and resources, including budget, IT expertise, security and compliance requirements, and scalability needs.

TECHNICAL REQUIREMENTS

RODA requires a minimum of two computers to operate: a server and a workstation. The server is responsible for hosting the data and handle all business processes. The workstation is used by end-users to access the system using a Web browser.

The following table outlines the recommended requirements to run the application various scenarios¹. Note that the system may run with lesser requirements, but these are set as a recommendation to support up to 10 million records.

	COMMUNITY	ENTERPRISE	ENTERPRISE HA
Servers	1	1	5
RAM (total)	32 GB	64 GB	128 GB
CPU (total)	8 CPUs	16 CPUs	32 CPUs
Storage ² (total)	512 GB SSD	1 TB SSD	2 TB SSD
Operating system	Ubuntu LTS 22.04+ Debian 11+	Ubuntu LTS 22.04+ Debian 11+ RHEL 7+	Ubuntu LTS 22.04+ Debian 11+ RHEL 7+
Platform	Docker compose	Docker compose	Docker swarm
Network	100 Mbit/s	100 Mbit/s	+100 Mbit/s

¹ Please note that the optimal infrastructure requirements depend on the amount of data to be processed and the number of users accessing the system at any given time. The figures outlined are recommended requirements for each setup.

² Storage requirements do not take into consideration the size needed for archival storage. The necessary storage capacity will vary depending on the amount of data expected to be ingested and preserved.

-  www.keep.pt
-  +351 253 066 735
-  info@keep.pt
-  sales@keep.pt
-  KEEP SOLUTIONS, LDA.
Rua Rosalvo de Almeida, n° 5,
4710-429 Braga
Portugal

KEEP SOLUTIONS

KEEP SOLUTIONS is a company whose mission is to provide advanced solutions for information management and digital preservation.

Our approach consists in providing software and services to allow our customers to make a more efficient management of their information assets.

The company started its activity in 2008, having acquired the status of academic spin-off of the University of Minho, for being a business initiative with strong bonds with research centres and departments from this institution.

Our clients are mostly found in the public sector, more specifically in the areas related to archives, libraries and museums.

We invest in the continuous development of innovative solutions. To support that, we remain active in the production of scientific knowledge while engaging in large-scale R&D projects in cooperation with national and international institutions.