
REVISION OF DIFFERENT IMPLEMENTATIONS FOR DIGITAL PRESERVATION: TOWARDS A METHODOLOGICAL PROPOSAL FOR PRESERVING AND AUDITING IR RELIABILITY

REVISIÓN DE DISTINTAS IMPLEMENTACIONES PARA PRESERVACIÓN DIGITAL: HACIA UNA PROPUESTA METODOLÓGICA PARA LA PRESERVACIÓN Y LA AUDITORÍA DE CONFIABILIDAD DE RI

REVISÃO DE DIFERENTES IMPLEMENTAÇÕES PARA A PRESERVAÇÃO DIGITAL: PARA UMA PROPOSTA METODOLÓGICA DE PRESERVAÇÃO E AUDITORIA DE CONFIANÇA DE RI

¹Marisa Raquel De Giusti, ¹Gonzalo Luján Villarreal

¹Universidad Nacional de La Plata

Correspondência

¹Marisa Raquel De Giusti
Universidad Nacional de La Plata Porto
Buenos Aires, Argentina
Email: marisa.degiusti@sedici.unlp.edu.ar
ORCID: <http://orcid.org/0000-0003-2422-6322>

Submetido em: 01/02/2018

Aceito em: 19/12/2018

Publicado em: 19/04/2018



JITA: JH. Digital preservation

RESUMEN: Este trabajo relata la experiencia inicial de prueba de una estructura apta para la preservación de documentos digitales en un archivo o repositorio. Se reconocen numerosos antecedentes de estructuras similares y, entre ellas, se describen brevemente tres experiencias exitosas dedicadas a conectar un repositorio con herramientas capaces de asegurar la preservación digital de los contenidos siguiendo el modelo OAIS, norma ISO 14721 (2012). Tras la descripción de estos tres modelos considerados más relevantes, se relata un prototipo en prueba en los repositorios gestionados en PREBI-SEDICI (UNLP) con las herramientas DSpace, Archivematica y ArchivesSpace, en el que el repositorio en DSpace está encargado del ingreso y la entrega de los contenidos digitales mientras que la estructura de Archivematica realiza las actividades de preservación digital a través de la implementación de un conjunto de microservicios, que actúan sobre una estructura conceptual asimilable al paquete de información (IP) en sus distintas versiones. La estructura física resultante del paquete de información en sus diferentes versiones (SIP, AIP, DIP) incluye archivos, checksum, logs, documentación de la transferencia y metadatos en una estructura XML. Este trabajo no tiene más pretensiones que mostrar los antecedentes y el inicio de un trabajo de investigación con el objetivo de generar consultas y reflexiones en el contexto latinoamericano, donde estas temáticas son incipientes.

PALABRAS CLAVE: Preservación de documentos digitales. Repositorio institucional. Herramientas de administración de repositorios. Paquetes de información para archivistas. Modelo OAIS. Auditoría.

ABSTRACT: This work introduces the initial experience of an infrastructure for digital documents preservation in archives or repositories. Prior backgrounds of similar infrastructures are recognized in this work, and among them three successful experiences are described. These experiences are all aimed to connect a digital repository with different software tools able to ensure digital preservation of repository contents according to OAIS ISO 14721 standard (2012). After the description of the three models, we describe a prototype under development in the repositories supported by PREBI-SEDICI (UNLP), which uses the software tools DSpace, Archivematica and ArchivesSpace. In this prototype, DSpace handles the ingest and delivery of digital contents, while Archivematica performs all the required digital preservation activities. This is achieved through a set of microservices applied to a conceptual structure similar to the information package (IP) in its different versions (SIP, AIP, DIP). The resulting structure of the IP includes checksums, original files, logs, transfer documentation and XML metadata. The main purpose of this work is to show the background activities already carried out in institutions around the world, and to start a research project aiming to generate ideas and thoughts in the Latin American context.

KEYWORDS: Preservation of digital documents. Institutional repository. Repository management tools. Information packs for archivists. OAIS model. Auditing.

RESUMO: Este trabalho relaciona a experiência inicial de testar uma estrutura adequada para a preservação de documentos digitais em um arquivo ou repositório. Numerosos antecedentes de estruturas similares são reconhecidos e, dentre eles, três experiências de sucesso dedicadas a conectar um repositório com ferramentas capazes de garantir a preservação digital de conteúdos seguindo o modelo OAIS, ISO 14721 (2012) são brevemente descritas. Depois de descrever estes três modelos considerados mais relevantes, um protótipo é contada em evidência em repositórios gerenciados em PrEBi-SEDICI (UNLP) com DSpace, Archivematica e ferramentas ArchivesSpace no repositório DSpace está a cargo de renda e entrega de conteúdo digital enquanto a estrutura do Archivematica realiza atividades de preservação digital através da implementação de um conjunto de microsserviços, que atuam sobre uma estrutura conceitual assimilável ao pacote de informações (IP) em suas diferentes versões. A estrutura física resultante do pacote de informações em suas diferentes versões (SIP, AIP, DIP) inclui arquivos, soma de verificação, logs, documentação da transferência e metadados em uma estrutura XML. Este trabalho não tem mais pretensões do que mostrar os antecedentes e o início de um trabalho de pesquisa com o objetivo de gerar consultas e reflexões no contexto latino-americano, onde essas questões são incipientes.

PALAVRAS-CHAVE: Preservação de documentos digitais. Repositório institucional. Ferramentas de gestão de repositórios. Pacotes de informação arquivísticas. Modelo OAIS. Auditoria.

OVERVIEW

Digital preservation (DP) aims to ensure long-term access to digital content. There are many ways to achieve this goal, including format monitoring, file integrity control, migration and emulation of environments. Digital repositories, conceived as spaces to host and disseminate large quantities of digital objects (DO), require a deep knowledge of the different aspects pertaining to DP and take adequate actions to ensure long-term access to hosted DOs. For this purpose, each organization needs to establish an internal structure capable of performing the different activities to digitally preserve their objects, which requires deploying computer systems, adapting technological infrastructures and establishing adequate processes throughout the DOs life cycle, in order to better analyze and transform when required.

Currently, there is no consensus on how to implement an adequate structure to ensure DP. However, there are good practices, computer systems and workflows that have proven to be very effective to perform most preservation tasks, and which are also key to enable the assessment and audit of an institutional repository (IR).

The main goal of this work is to analyze a limited number of recommended structures to preserve and ensure long-term access to DOs stored in a repository or in structures suitable to host them. The three proposals analyzed here have been tested by other recognized institutions, including successful projects undertaken jointly by multiple institutions, repositories and archives, interested in digital preservation. Throughout the sections herein, the reason for the proposed configuration to be tested in the CIC Digital and SEDICI repositories, managed by the PREBI-SEDICI working groups, will be clearer.

Regarding the conclusions, suffice it to say that after analyzing the tools, and having the implementations a set of basic shared features, such as aggregation of special metadata to enable preservation and follow-up of the DOs life cycle, the choice relied on the most common implementation of IR in Latin America based on the open source tool DSpace. Multiple environments could be used for testing, but in this case, the specific operation of managed repositories was taken into account since it also offers a simpler structure.

BACKGROUND: SUCCESSFUL STRUCTURES AND THEIR DEPLOYMENT

1. STRUCTURE USED IN THE SCAPE PROJECT

Scalable Preservation Environments (SCAPE) is a project coordinated by the Austrian Institute of Technology (AIT), funded by the European Union. The project was launched in 2012 and was completed, according to the plan, at the end of 2015. It brought together experts from institutions that host cultural heritage, data centers, laboratories, universities and industries to study the technological and organizational aspects pertaining to digital preservation. The highlights of the project were:

- Analysis of repository file formats
- Formal description of preservation plans and policies
- Automation and vitalization of tools and scalable processes
- Quality control of preservation processes

The reference architecture proposed by SCAPE includes, throughout the life cycle of preservation: 1) a repository instance, which can be generated on DSpace, EPrints or RODA and in the implementation of the reference structure on RODA; 2) a monitor for the different internal and external processes, which also notifies risks and opportunities, especially in relation to aspects of preservation management, implemented on SCOUT; 3) a detailed process of preservation planning carried out through PLATO, which leads to a preservation plan that is ingested in the repository and 4) a workflow management system that allows the execution and synchronization of complex tasks, including characterization and migration. These tasks are performed by different tools, integrated into the workflow of the TAVERNA Workflow Management System. The reference implementation of SCAPE follows a scheme which is similar to the one shown in Figure 1.

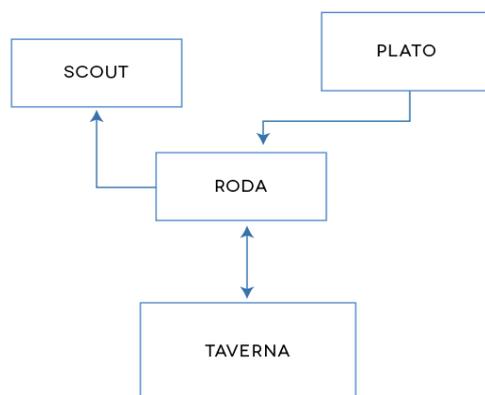


Figure 1 SPE (SCAPE Preservation Environment). Reference implementation

Miguel Ferreira and other (2014) present the SCAPE project architecture and implementation and the results of its evaluation in relation to ISO 16363, the standard for auditing and certification of reliable repositories. Besides providing a detailed model description, the paper states that in fact the scheme/model developed can comply with most of this standard's requirements -it is worth noting, however, that the metrics related to the organization that supports the repository and its procedures exceed the technological capabilities of a tool. With this vision, the analysis sets aside the metrics of the Standard section under the title Organizational Infrastructure and mostly those in the section Organizational Structure & Staffing, except those referring to repository policies, data transparency and integrity, since it is able to partially support the monitoring of intellectual

management, rights and restrictions; however, the partial aspect is related to the human problem posed by lacking permissions or setting them wrongly when the AIP is set up. The model allows and gives full account of the functions specified in sections 4.2 and 4.3 of the standard regarding intake, AIP creation, preservation planning and management, and access management. The processes related to institution dependent procedures are excepted, including those related to the actions to be taken on the AIP, and even procedures that exceed documentation and relate to the technology -hardware and software- to meet the requirements of the designated community or those posed by risk management.

The work is thorough and rigorous and its reading is highly recommended for further information on the issues of reliability. Within this work, however, the most important aspects are those related to studying structure operation and, naturally, the functionality of the entities proposed by ISO 14721, as well as the distribution of such functionality in the SPE.

Although the project has already been completed, all the materials generated and the information collected over four years is available for free on its website.

ABOUT RODA

RODA was developed to be a complete digital repository and it delivers functionality for all the main functional units of the OAIS reference model (Faria et al., 2009). RODA implements the entire ingest workflow in which it not only validates the SIP but also handles the negotiation process between the file and the producer of information. For the access process, it provides different search and navigation possibilities through metadata, as well as DO visualization and download functions. Administration components were also developed to allow archivists to modify descriptive metadata and define rules for preservation actions, such as planning integrity controls on all stored DOs, initiating a migration process, or control by user/groups authorized to execute actions within the repository.

RODA's content model is atomistic and very PREMIS-oriented. Each intellectual entity is described by an EAD-component (Encoded Archival Description) (Pitti, 1999) of metadata record. These records are organized hierarchically in order to constitute a full archival description, but are kept separately within the Fedora Commons content model. (Fedora, 2017). These EAD components are created using Fedora's own RDF linking mechanism and each node, hierarchical tree leaf is linked to a representation object (Figure 2) -e.g., a Fedora object that includes all the files and bit-streams that compose the digital representation. Logical relations are maintained between all these objects, by a set of PREMIS entities (PO nodes) that provide information about the digital object's provenance and history.

The preservation events that take place are recorded as new preservation-event nodes. Some special events, like format migrations, establish additional relationships between two

preservation-representation nodes (linking events). Each preservation event is executed by an agent, which can be a system user or an automatically triggered software application. As expected, the information of the agent that triggered the event is also recorded in the PO agent node.

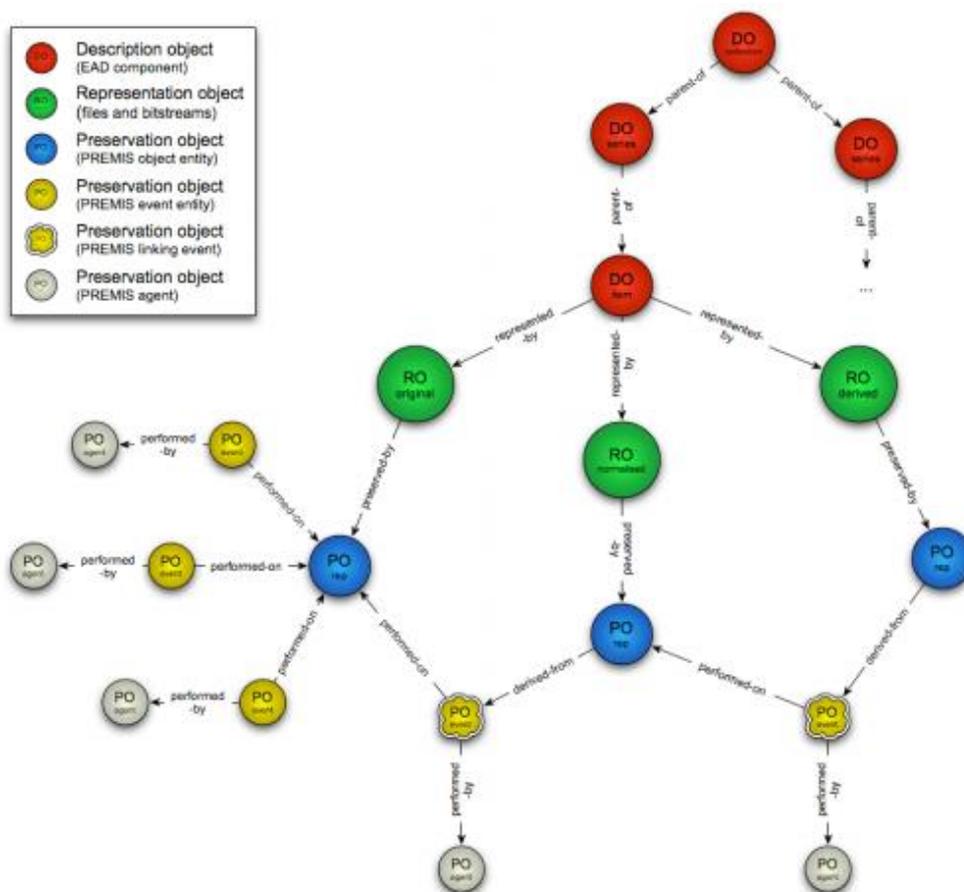


Figure 2 RODA's content model with EAD-components, digital representations and PREMIS objects.
 Source: Faria et al. (2009)

2. STRUCTURED USED BY BRITISH COLUMBIA UNIVERSITY

Lori J. Ashley (2016) considers the technological risks which digital assets are exposed to, analyzes the limits and scope of a given set of strategies over time and provides a brief description of the OAIS reference model, the ISO 16363 standard and some technological topics which are key to preservation, including updating, mirroring, migration, emulation and format standardization. Her work acts as a theoretical introduction to Bronwen Sprout and Sarah Romkey (2016), who describe the experience of University of British Columbia Library, including the implementation of the institutional repository. This repository was implemented in DSpace, supported by Artefactual Systems, creators of Archivematica. After analyzing the preservation practices that were carried out both on digitally born materials and

those resulting from digitization processes, Artefactual observed the deficiencies, made a diagnosis and proposed a structure to be tested in a pilot project, shown in Figure 3.

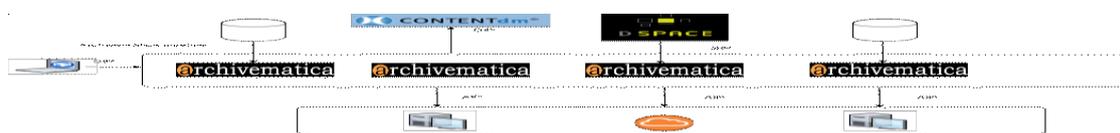


Figure 3 General flowchart of an Archivemática pipeline.
Source: Sprout y Romkey (2016)

The flowchart in Figure 3 shows the different mechanisms (there could be other) for SIP entry in terms of OAIS, that is, how a file enters the pipeline, and the execution by Archivemática of the functions described by the abstract OAIS structure that allow generating an AIP (a preservable package) and also a DIP (a deliverable package) for any configuration, e.g., the image includes CONTENTdm, a tool focused on the digital asset storage and management. Each instance of Archivemática within the pipeline can process content for different applications that will use it in one way or another. The preservable package can be stored in different virtual spaces, such as local servers or a cloud, by the different instances of the pipeline.

This work (Sprout and Romkey, 2016) describes the hands-on implementation experience by the University of British Columbia, which has a DSpace repository, known as cIRcle. In this configuration, DSpace acts as a deposit and access tool (SIP and DIP), but does not generate the AIP preservable package, which is done by one of the instances of Archivemática in the pipeline. It is important to note that this method does not affect the user interface or their experience: the repository is the entry point for their files and from the repository, the user receives the answers to the information requests entered.

ABOUT DSPACE

DSpace is an open source development which enables the implementation of a repository, access to files in multiple formats, adding metadata for cataloguing, storage, mirroring, dissemination and delivery upon request by repository users. While DSpace includes some of the requirements under the OAIS Model, it is not easy to comply with most of the functions that OAIS describes within the entity called Preservation Planning, mainly because that entity supposes an evolutionary behavior; DSpace can perform transformations, such as data migration. The functions that can be carried out from the administration, similar to those of the Administration entity, which is the most complex within the OAIS model, are far from complying with all the functions required to ensure preservation (De Giusti et al., 2012). In particular, it is difficult to separate agents and events according to the description

provided by PREMIS data dictionary, which would be important when recovering DOs or executing any DO transformation task. The possible events do not have an adequate description, either, so it is difficult to follow up the DO's life cycle and ensure the necessary preservation actions are taken so as to guarantee access and readability over time.

This difficulty posed by the software presents when it comes to digital preservation (De Giusti, 2016), is one of the reasons behind the analysis about structures such as the ones reviewed in this work.

ABOUT ARCHIVEMATICA

Archivematica is an open-source digital preservation application based on recognized standards designed to ensure long-term access to digital files. Developed by Artefactual Systems, Archivematica consists of a set of integrated applications and open-source tools that allow users to process DOs from their entry (ingest) and their storage until delivery (access) as per the ISO-OAIS model. The functionalities of Archivematica will be described in greater detail in section 4.3.

3. STRUCTURE USED BY BENTLEY HISTORICAL LIBRARY (MICHIGAN UNIVERSITY)

After reviewing various open-source tools for digital asset management software, the University of Michigan Library decided to integrate the functionalities of ArchivesSpace, Archivematica and DSpace to achieve a workflow with the structure shown in Figure 4.

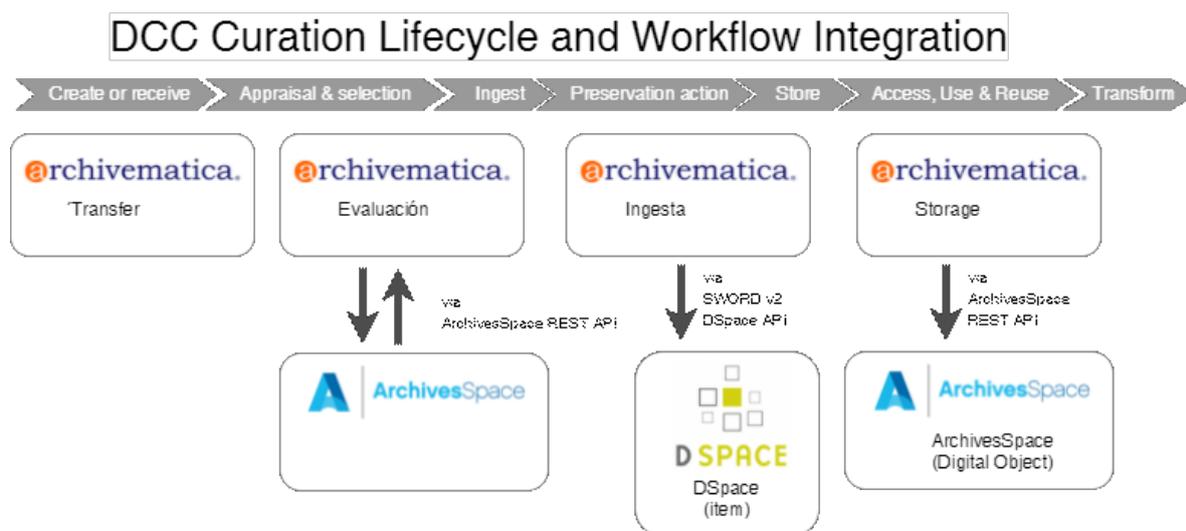


Figure 4 Structure used by Bentley Historical Library, Michigan University
Source: Eckard, Pillen & Schallcross (2017)

The proposed structure meets preservation and long-term access needs for digitally born objects as per the institution's criteria. Within the implemented structure, distributed functions are performed according to the needs of the institution:

- Facilitating creation/reuse of descriptive and administrative metadata in preservation and management systems.
- Simplifying content intake and storage in a preservation repository.

Finding solutions for Bentley, which can be extrapolated to other institutions. Sharing all the code and documentation with archives and digital preservation communities.

ABOUT ARCHIVESPACE

ArchivesSpace is an open source file management software that allows institutions to track access sessions, manage collection and generate an EAD (Encoded Archival Description) description. Basically, a document is a "description tool" encoded using EAD, which consists of three segments: one that provides information about the description tool itself -including title, compiler and compile date- a second component that includes the elements required for the formal publication of the description instrument, and a third component that provides the description of the archival material, as well as the related contextual and administrative information. ArchivesSpace helps to perform the basic functions of any file, including material description, document authority management and management of documents and issues related to them (such as the number of visualizations), and even metadata editing the application itself.

This tool offers multiple functions, including:

- Incorporating new records
- Publishing materials
- Authority management
- Location management
- Rights management
- Reference service
- Reporting
- Generating EAD, MARCXML, MODS, Dublin Core and METS metadata
- Exporting

4. PROTOTYPE SELECTED FOR TESTING IN THE IRS MANAGED IN PREBI- SEDICI

Based on the review of the three models above and, in particular, the analysis of workflows and implementation of managed repositories (SEDICI and CIC Digital) in DSpace, a structure similar to the one used by the University of Michigan was chosen. The test tasks have only just began: at the moment the Archivematica application and the TRAC (Trustworthy Repositories Audit and Certification) functionality have been installed on a

server. Since these elements are part of the experience to be shared, the aspects pertaining to Archivemática are extensively described below.

4.1. Installation

The Archivemática installation requires a GNU/Linux server -for the time being, only Ubuntu Server 14.04.5 and CentOS 7.3.1611 are supported, both 64 bits-, MySQL -also supports Percona and MariaDB-, an HTTP server -Nginx or Apache- and Elasticsearch. The installation guide available on the Archivemática website (Archivemática, 2017a) recommends the minimum hardware requirements for both test environments and production environments, with the steps required to perform a typical installation, both in Ubuntu Server and in CentOS. The Archivemática installation also includes the installation of StorageService, a web software for managing available storage spaces to enable access from Archivemática. In this software, a local directory was created for the Transfer process (Archivemática process prior to Ingest). System users with read and write permissions were also generated through SFTP in that directory. This allows to speed up the loading process for files and complete directory structures, and enables tests from any computer with a running SFTP client.

4.2. The case for selection - pros and cons

One of the key features of Archivemática is related to its design, which includes proven tools to perform the various functions recommended in the OAIS abstract model. This was considered a competitive advantage -as well as a difference- against models such as SPE, since expected preservation functions are performed in a single architecture; on the other hand, this makes it difficult to track the different steps.

A particularly important issue in relation to the goals herein, and which communicates the needs of the repositories managed by the PREBI-SEDICI group, is verifying the capabilities of SEDICI and CIC Digital repositories in terms of reliability, as per the Audit and Certification of Trustworthy Digital Repositories (CCSDS, 2011) which is then included in ISO 16363. In this sense, Archivemática is also suitable, since it incorporates the tool developed by MIT in the project aiming to provide curation and preservation services, known as TRAC Review Tool, whose installation is independent from Archivemática.

4.3. Basic features of Archivemática

Archivemática offers an integrated suite of free and open source software tools that allow users to process DO from ingest to storage, archiving and access in compliance with the ISO-OAIS functional model and other digital preservation standards and good practices.

The structure of Archivemática¹ is based on two key and complementary elements: microservices and *Foss tools*, which allow compliance with the OAIS model. These tools are embedded in the different platform modules and they can be updated and configured individually. These tools enable normalization of the different file formats used throughout the workflow, which begins with information transfer to the system.

The Archivemática microservices are processes used to execute jobs, actions and transformations during the processing of information packages in all the stages of digital file management, including transfer, ingest, storage and access. The administrator can customize and distribute them throughout the workflow as needed. Some of these actions carried out on the archives are automatic, while others may require the intervention of the repository manager, who will have to make decisions, often strategic ones.

Communication between Archivemática and the administrator in the various processes is done through a *Dashboard* that displays the microservices and in some cases requests approval or attention from the administrator. This Dashboard layout is shown in Figure 5:

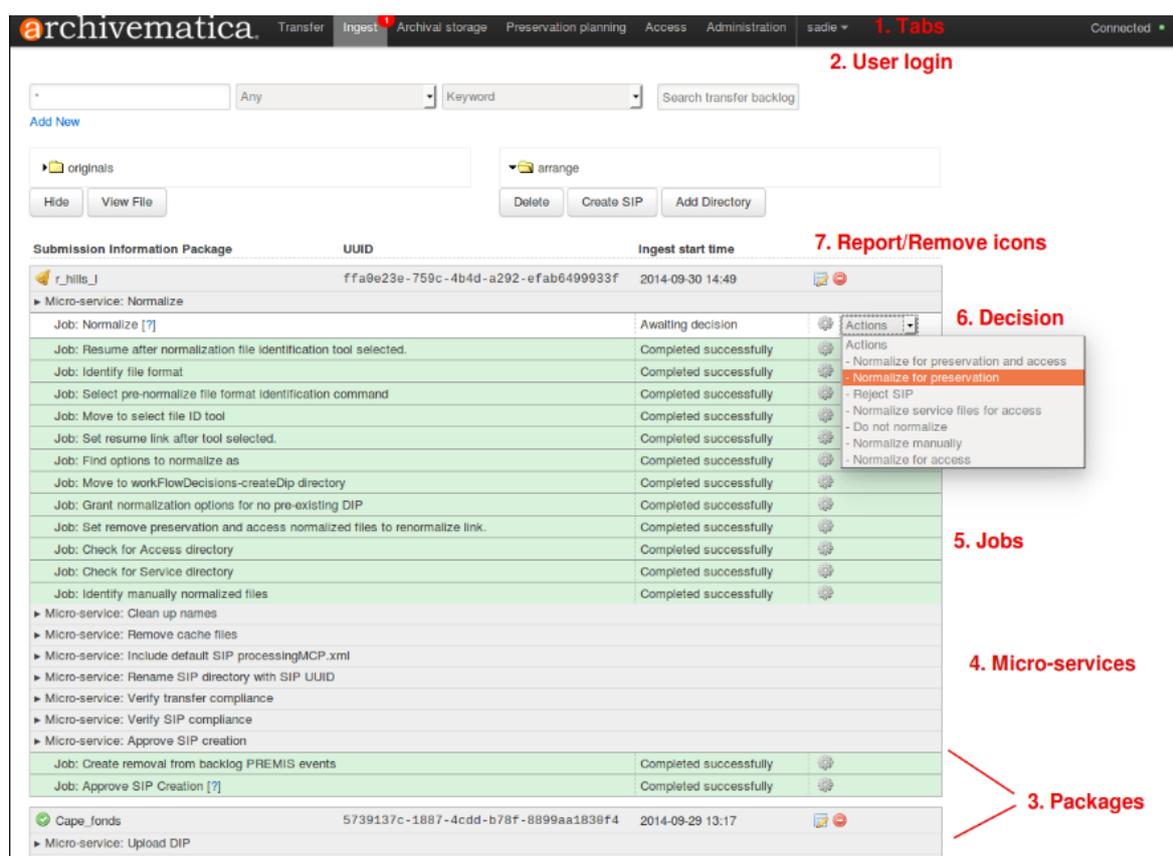


Figure 5 Archivemática Dashboard

Source: <https://www.archivematica.org/es/docs/archivematica-1.5/ images/Dashboard.png>

¹Archivemática offers extensive documentation on its website; based on the version installed for testing, we have used version 1.5 of the manual, available at <https://www.Archivematica.org/es/docs/Archivematica-1.5/>

Archivematica supports PREMIS, METS and Dublin Core metadata, but it also supports import of other type of metadata added to the DO by the administrator. Archivematica implements preservation plans for different content types. Upon installation, it connects with the Format Policy Registry (FPR) to update its local database. This registry allows users to define the policies for the different file formats.

The Archivematica Dashboard allows to monitor the actions that happen in the different processes and microservices, as well as to track any events, states and errors.

Unlike the OAIS model, whose workflow begins at ingest, Archivematica highlights the previous process, called Transfer, which is the process of transforming, checking and validating any set of digital objects and/or directories into a SIP (Submission Information Package, in terms of OAIS); these DO/directories can originate in a DSpace repository or from other applications as seen in the two proposed models. The administrator chooses the appropriate option to start the transfer process -e.g., selecting a directory with the contents to be submitted to the Archivematica curation process; these contents may include documents as agreed with the content providers, in which case the administrator needs to create the necessary directories to organize them. To carry out the first tests, a limited set of DO is preferable, with some obsolete formats or malformed files, in order to acquire familiarity with the reports delivered by the tool. Transfer is approved by a manual process, once the administrator triggers delivery validation (first microservice) in the Dashboard

The processes executed in this testing phase include:

- Approving delivery
- Checking requirement compliance
- Renaming the files by adding a unique identifier
- Creating checksums and verifying them
- Creating METS XML files
- Quarantining files when necessary
- Identifying file formats
- Extracting packaged files
- Checking for viruses
- Moving files to the deliverables directory
- Creating the SIP²

A variety of tools perform the processes above. For instance, Bagit is used for packaging and storing digital objects, Fido and Siegfried for extracting metadata, Jhove and

²Alternatively, the files can be sent to the backlog for further processing and even creating reports, as can be seen in the Archivematica documentation, available at <https://www.archivematica.org/es/docs/archivematica-1.5/user-manual/transfer/manage-backlog/#retrieve-from-backlog>

FITS for extraction and validation. This greatly simplifies the administrator's work, because although the progress of the system can be seen through the different steps being executed, the administrator does not need to master all the tools.

Then, the Ingest function, which is found in the Archivematica menu, is activated. Ingest runs SIPs through several microservices, namely:

- Normalize
- Add metadata (this can be before or after normalization)
- Add PREMIS rights

Normalizing is the process of converting ingested digital objects to preservation and/or access formats. It is worth noting that the original objects are always kept along with their normalized versions. Normalize for preservation and access creates copies in order to have a preservable object (AIP) and a deliverable object (DIP). Once normalization is approved, the SIP runs through a number of microservices, including submission documentation processing, METS file generation, indexing, DIP generation and AIP packaging.

As with each step, the results can be reviewed to verify that everything is correct. Once normalization is completed, the user can save the AIP, publish the DIP, and even review the AIP if desired (see Figure 6).



Figure 6 Screenshot of Archivematica Dashboard and AIP review.
Source: <https://www.archivematica.org/es/docs/archivematica-1.5/user-manual/ingest/ingest/#store-aip>

The AIP can be reviewed if necessary. The user manual recommends reviewing and storing the AIP before uploading the DIP since any problem with the AIP would require having to locate the DIP and delete it. Archivematica supports DIP uploads to AtoM, ArchivesSpace, CONTENTdm and Archivists's Toolkit. AIP reingest is also supported for the purpose of adding metadata.

Archival Storage functions do not require much detail; however, some notes are worth mentioning. Archivematica uses a directory tree structure to locally store AIPs. The tree is based on AIP 16-digit alphanumeric unique universal identifiers (UUID); it also supports

multiple storage sites either locally or remotely, with LOCKSS locations. The integration procedure details can be found in the Archivemata wiki (Archivemata 2017b).

The Dashboard's Archival Storage tab shows a table with information about the stored AIPs and the administrator can sort them out or copy them. Each AIP is identified by its name and the identifier assigned during SIP creation. Figure 7 shows a view of two files with their identification:

Browse archival storage

Total size: 55.56 MB Total files: 22 indexed

| AIP | Size | UUID | Date stored |
|--------|----------|--------------------------------------|------------------|
| NewDir | 27.03 MB | 2e267ee7-19ab-4fe1-830a-23cb7e223899 | 2014-08-15 10:33 |
| Test | 28.53 MB | 94c2f288-ac61-4e38-ae5c-863bb282ec5e | 2014-08-12 09:07 |

Figure 7 Dashboard screenshot with file list

Source: <https://www.archivemata.org/es/docs/archivemata-1.5/images/ArchStorTab1.png>

Regarding the Preservation Planning module, one of its main goals (as mentioned in the Ingest process) is to normalize files to preservation and access formats. Upon the first connection to the FPR server, Archivemata is able to exchange data on the Agent and its Identifier, as well as the UUID and the IP address of the host, as well as the time of the event.

When creating a new format version, a text describing the format will be required, saved in METS files; also required are the version number for this specific format version, the PRONOM ID, that is, the specific format version's unique identifier in PRONOM, the UK National Archives's format registry and an indication specifying whether this format is suitable as an access format and/or for preservation. While Archivemata supports a wide range of formats, it does not always normalize all formats, as is the case with MS Word³.

The Archivemata Preservation planning module (Borthwick Institute for Archives, 2017) has a key element, namely, the Preservation Planning tab which displays the local Format Policy Registry (FPR) and where the administrator can add formats or edit existing ones upon the first connection. Figure 8 shows the identifier for PDF/A format.

³This is specified in Archivemata (2017c) documentation, "some formats such as Microsoft Word documents are not necessarily in the best preservation format but are still so ubiquitous and well-supported that they need not be normalized at the present time. In these cases, the Archivemata default is to keep them in their original formats."

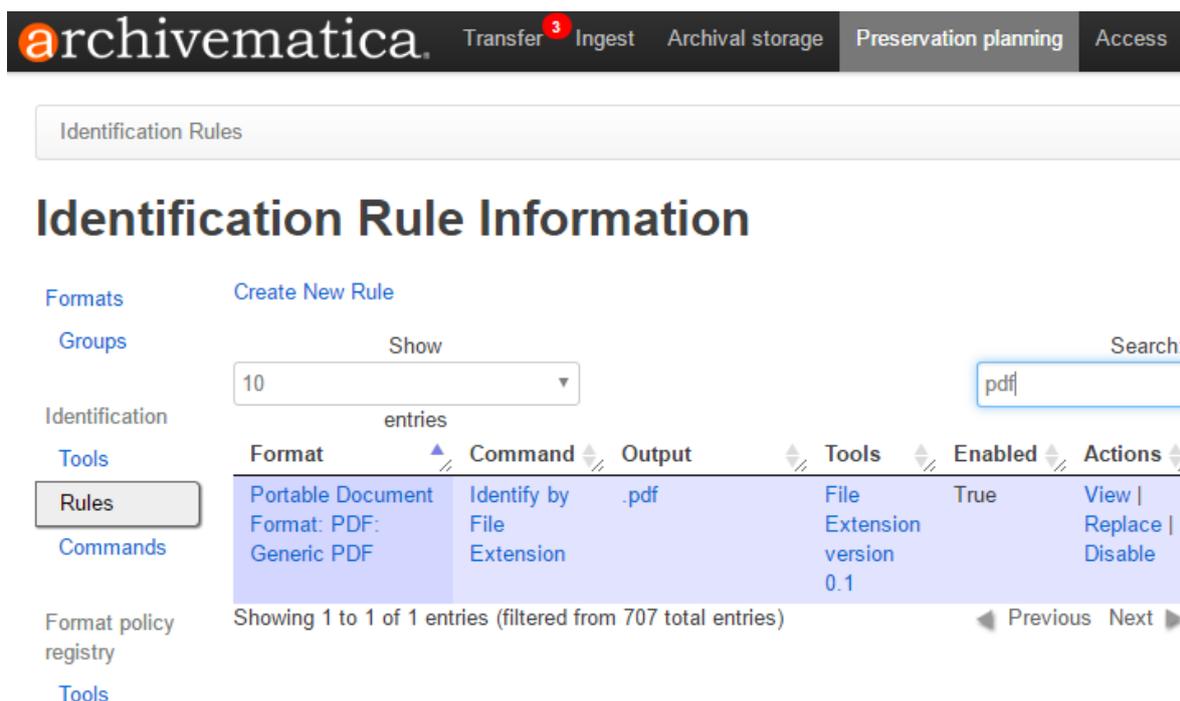


Figure 8 Screenshot of the Preservation Planning Dashboard

FPR rules can be updated at any time by the Administrator. This module also has an identification tool which controls format identification and their association with the FPR; for this purpose, several tools are used, such as FIDO (Open Planets Foundation), which identifies files based on the IDs in PRONOM; a *script* which identifies files by their file extension; and Siegfried, which is also based on PRONOM ID Version 1.5 has five format characterization tools, including FITS. The validation process is based on Jhove.

TRAC REVIEW TOOL

TRAC Review⁴ tool was developed by the MIT, based on the content manager Drupal, very useful for an organization seeking to implement a trusted digital repository and particularly for institutions using Archivematica for this purpose. TRAC provides an overview of compliance -or lack thereof- with the requirements of the CCSDS checklist that was approved as ISO 16363 (2012) and is based precisely on the requirements set forth by Trustworthy Repositories Audit and Certification (TRAC). This self-assessment allows to demonstrate good practice and conformance as a trusted digital repository to its designated communities. TRAC proposes many responsibilities for compliance which in many cases are distributed throughout the organization, with specific units and committees with responsibilities for certain requirements.

⁴Installation requirements are available in <https://www.archivematica.org/en/docs/archivematica-1.5/user-manual/getting-started/trac/>

JOBS AND PENDING TAKS

When considering a digital preservation system, it is important to be familiar with the experience of other organizations and the potentially successful tool combinations, not only because these tools meet the technical requirements for which they were designed, but also because they can be combined with other tools in order to implement complex systems that adapt to specific organization requirements. The task carried out so far in PREBI-SEDICI has been becoming familiar with the tools that make up the selected preservation structure and describing three internationally recognized success stories: the SCAPE project, the architecture proposed by the British Columbia University and the structure used by the University of Michigan Bentley Historical Library.

This is a work in progress and hence conclusions related to the best implementation should not be considered definitive. In many aspects, Archivemática complexities need to be further analyzed, since it is used in two of the three proposals considered herein, and, additionally, the generation of connections with ArchivesSpace and a test repository implemented in DSpace need to be assessed, since a connection under the SWORD 2 protocol will have to be enabled. So far only isolated tests have been carried out on Archivemática and there is still much work to be done. However, the analysis of these tools is a great step forward for the future of managed repositories and the purpose of these notes is, as mentioned before, sharing these first steps in the search for the best solution to the growing problem of digital file preservation.

REFERENCES

ARCHIVEMATICA. **Installation**: documentación (Archivemática 1.6): open-source digital preservation system. 2017a. Disponible en: <<https://www.archivematica.org/es/docs/archivematica-1.6/admin-manual/installation/installation/>>. Acceso en: 12 dic. 2017.

ARCHIVEMATICA. 2017b. Disponible en: <https://wiki.Archivematica.org/LOCKSS_Integration>. Acceso en: 12 dic. 2017.

ARCHIVEMATICA. **Preservation planning**. 2017c. Disponible en: <<https://www.archivematica.org/fr/docs/archivematica-1.4/user-manual/preservation/preservation-planning/>>. Acceso en: 12 dic. 2017.

ASHLEY, Lori J. Theory: Creating a preservation strategy. In: BANTIN, Philip C. (Ed.). **Building trustworthy digital repositories**: theory and implementation. Rowman & Littlefield: London, 2016.

BORTHWICK INSTITUTE FOR ARCHIVES FILLING THE DIGITAL PRESERVATION GAP. **Report on Archivemática for research data now available**, 2017. Disponible en <<https://www.york.ac.uk/borthwick/projects/archivematica/>>. Acceso en: 12 dic. 2017.

Consultative Committee for Space Data Systems (CCDS). **Audit and certification of Trustworthy Digital Repositories**. The Magenta Book, 2011. Disponible en: <<https://public.ccsds.org/pubs/652x0m1.pdf>>. Acceso en: 12 dic. 2017.

DE GIUSTI, Marisa R. LIRA, Ariel; VILLARREAL, Gonzalo; TEXIER, José. Las actividades y el planeamiento de la preservación en un repositorio institucional. In: CONFERENCIA INTERNACIONAL BIREADIAL-ISTEC 2012, 2012, Barranquilla (Colombia). [**Annales electrónicos...**]. Barranquilla: Universidad del Norte, 2012. Disponible en <<http://sedici.unlp.edu.ar/handle/10915/26045>>. Acceso en: 12 dic. 2017.

DE GIUSTI, Marisa R.. Las dificultades de la preservación digital: problemas, desafíos y propuestas para los repositorios. In: CONFERENCIA INTERNACIONAL BIREADIAL-ISTEC 2016, 2016, San Luis Potosí. [**Annales electrónicos...**]. San Luis Potosí, México: Universidad Autónoma de San Luis Potosí, 2016. Disponible en <<http://sedici.unlp.edu.ar/handle/10915/56288>>. Acceso em: 12 dic. 2017.

ECKARD, Max; PILLEN, Dallas; Schallcross, Mike. Bridging Technologies to Efficiently Arrange and Describe Digital Archives: the Bentley Historical Library's ArchivesSpace-Archivematica-DSpace Workflow Integration Project. **Code(4)Lib Journal**, n. 35, p. 1-30, 2017. Disponible en: <<http://journal.code4lib.org/articles/12105>>. Acceso en: 12 dic. 2017.

FARIA, Luís et al. RODA: a service-oriented repository to preserve authentic digital objects. In: INTERNATIONAL CONFERENCE ON OPEN REPOSITORIES (OR 2009), 2009, Atlanta. [**Presentations of...**]. [S.l.]: [s.n.], 2009. 71 Slides PPT. Disponible en: <<http://repositorium.sdum.uminho.pt/handle/1822/9408>>. Acceso en: 12 dic. 2017.

FEDORA. **Fedora Commons**. 2017. Disponible en: <<http://www.fedora.info/>>. Acceso en: 12 dic. 2017.

FERREIRA, Miguel Et al. **SCAPE: Report on compliance validation**. Whitepaper, 2014. Disponible en: <<http://hdl.handle.net/1822/30689>>. Acceso en: 12 dic. 2017.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO) .ISO 14721. **Space data and information transfer systems - Open archival information system (OAIS) - Reference model**, 2012. Disponible en: <<https://www.iso.org/standard/57284.html>>. Acceso en: 12 dic. 2017.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 16363. **Space data and information transfer systems - Audit and certification of trustworthy digital repositories**. 2012. Disponible em: <<https://www.iso.org/standard/56510.html>>. Acceso em: 12 dic. 2017.

PITTI, Daniel V. Encoded Archival Description. An Introduction and Overview. **D-Lib Magazine**, v.5, n.11, Nov., 1999. DOI: 10.1045/november99-pitti. Disponible en: <<http://www.dlib.org/dlib/november99/11pitti.html>>. Acceso en: 12 dic. 2017.

SPROUT, Bronwen. ROMKEY, Sarah. Implementation: Building a preservation strategy around Archivematica. In: Bantín, Philip C. (Ed.). **Building trustworthy digital repositories: theory and implementation**. Rowman & Littlefield: London, 2016.