

archivemata®

Open-source, standards-based digital preservation

Archivemata is a digital preservation platform designed to:

- Ingest a wide variety of digital holdings
- Perform automated digital preservation actions
- Prepare system independent, standards based Archival Information Packages for long-term storage
- Make access copies available via integrated online access and cataloging systems
- Provide an interface for curation experts to perform analysis, appraisal and preservation planning for digital holdings

Use Archivemata if your organization needs to:

- Ensure that content can be retrieved, rendered and understood over time
- Automate preservation actions to achieve consistency and scalability
- Ensure that content is preserved with no vendor lock-in or proprietary restrictions

Archivemata was developed by Artefactual Systems Inc., a company based in Vancouver, Canada. Artefactual continues to be the lead developer of Archivemata, and offers hosting, technical support, consulting, training and other services related to the development and use of the software.

*This information sheet is designed to provide detailed functional and technical information about Archivemata. For a high-level summary, see **Archivemata digital preservation system summary sheet 2019**, available on the Artefactual Systems website.*

Q&A: Standards	3
Does Archivemata conform to the Reference Model for an Open Archival Information System (OAIS) - ISO 14721:2012?	3
Does Archivemata use PREMIS (Data Dictionary for Preservation Metadata Implementation Strategies)?	3
Does Archivemata implement any other metadata standards?	4
Does Archivemata have its own internal metadata schema?	4
Why doesn't Archivemata also implement ISAD(G), DACS, RAD, EAD, etc.?	4
Does Archivemata conform to [XYZ standards] for storage systems?	4
Q&A : Workflows	4
How does content get ingested into Archivemata?	4
Can I ingest a directory with subdirectories?	5
Can I ingest content from external media directly into Archivemata?	5
What kinds of content can be ingested?	5
Can I ingest zipped or other compressed packages?	5
Does Archivemata crawl websites?	6
Does Archivemata ingest email accounts?	6
Does Archivemata handle forensic disk images?	6
How do I add descriptive metadata to my content?	6
How do I add rights metadata to my content?	6
Can I import custom metadata into Archivemata?	6
Does Archivemata assign identifiers to ingested objects?	6
Does Archivemata assign checksums to ingested objects?	6
Does Archivemata perform virus and malware scanning on ingested objects? What happens if a virus is found?	7
What methods and tools are used to identify and characterize the formats of ingested files?	7
What methods and tools are used to convert ingested files to preservation file formats?	7
How do I manage and modify workflows?	7
Can I modify or remove content after ingest but before the AIP is prepared?	9
What happens if there are errors during processing?	9
What types of errors are reported?	9
Does Archivemata maintain an audit trail of actions that occur during processing?	10
What does an Archivemata Archival Information Package (AIP) consist of?	10
How do I find a stored AIP?	11
Can I modify the contents of an AIP?	12
Can I delete an AIP?	12
Does Archivemata perform periodic fixity checks on stored AIPs?	12
Does Archivemata support records classification and retention scheduling?	13
What does an Archivemata Dissemination Information Package (DIP) consist of?	13
How is a Dissemination Information Package (DIP) displayed in AtoM?	13
Does Archivemata monitor formats and provide preservation risk assessments?	13
Does Archivemata provide statistical reports? Can these be customized?	14

Archivematica digital preservation system Information sheet 2019

Q&A: Scalability	14
Are there any physical limits on the size or number of files that can be ingested at one time?	14
Can Archivematica accommodate multiple ingests?	14
Can ingest be automated to run continuously?	14
Is there a maximum AIP file size or a maximum number of files that may be included in an AIP?	14
Q&A: Integrations	15
What kinds of storage protocols and systems are supported?	15
Can Archivematica ingest content from Microsoft products such as SharePoint and Exchange?	15
What access systems does Archivematica integrate with?	15
Q&A: System administration	15
Does Archivematica require any third-party applications in order to function correctly?	15
Do any required third-party applications require additional license fees or custom integration?	15
Where are system logs stored and how are they kept secure?	15
What monitoring tools can be used with Archivematica?	16
Can I create different kinds of user accounts with their own sets of permissions?	16
Can Archivematica be integrated with LDAP and Shibboleth for single sign-on?	16
How is Archivematica protected against intrusion?	16
How can my data be recovered in the event of a system failure, failed update etc.?	16
Can Archivematica be run as a multi-tenanted platform?	17
About Artefactual Systems Inc.	17
Find out more	17

Q&A: Standards

Does Archivematica conform to the *Reference Model for an Open Archival Information System (OAIS) - ISO 14721:2012*?

Archivematica implements the functional areas described by OAIS, including Ingest, Preservation Planning, Archival Storage, Administration and Dissemination. Archivematica ingests digital content and packages it into Submission Information Packages, Archival Information Packages and Dissemination Information Packages (SIPs, DIPs and AIPs). We therefore say that Archivematica is based on OAIS, but since OAIS is a reference model and not a software specification, we don't use the terms "conforms to" or "complies with".

Does Archivematica use PREMIS (Data Dictionary for Preservation Metadata Implementation Strategies)?

Yes. Archivematica has a very robust PREMIS implementation, serialized to XML and packaged in METS (Metadata Encoding and Transmission Standard). Detailed technical and administrative metadata and a comprehensive audit trail of preservation events are generated automatically; the user can also use a metadata entry template to add rights metadata if desired.

Does Archivematica implement any other metadata standards?

Yes, METS, as described above, and Dublin Core.

Does Archivematica have its own internal metadata schema?

No. When an AIP is generated, a PREMIS in METS file, which contains all of the metadata pertaining to the contents of the AIP, is generated and packaged into the AIP along with the other AIP contents. No metadata that is relevant to the AIP or which needs to be preserved over time is kept in the database, and no future export or conversion of the metadata to standardized formats is required.

Why doesn't Archivematica also implement ISAD(G), DACS, RAD, EAD, etc.?

Archivematica isn't a cataloging or access system. Instead it sends Dissemination Information Packages to access systems such as AtoM, where rich metadata enhancement can be undertaken for discoverability and access purposes.

Does Archivematica conform to [XYZ standards] for storage systems?

Archivematica does not store AIPs, but instead integrates with storage systems. Types of storage systems include local file-based systems; cloud-based storage such as Amazon S3, Microsoft Azure and OpenStack Swift; and specialized storage tools and services such as LOCKSS, DuraCloud and Arkivum. For hosted Archivematica instances offered by Artefactual, we use third-party hosting infrastructure and storage services such as DuraCloud, Amazon, OVH and Microsoft Azure. These services have ISO 27001 and other standard cloud computing certifications.

Q&A : Workflows

How does content get ingested into Archivematica?

Users can manually ingest content via the Transfer tab in the web-based dashboard (fig. 1), or can use automation tools to automatically ingest content from designated source locations such as a folder on a network drive or from systems such as DSpace, Islandora and Dataverse. Transfer locations are configured using the Storage Service module.

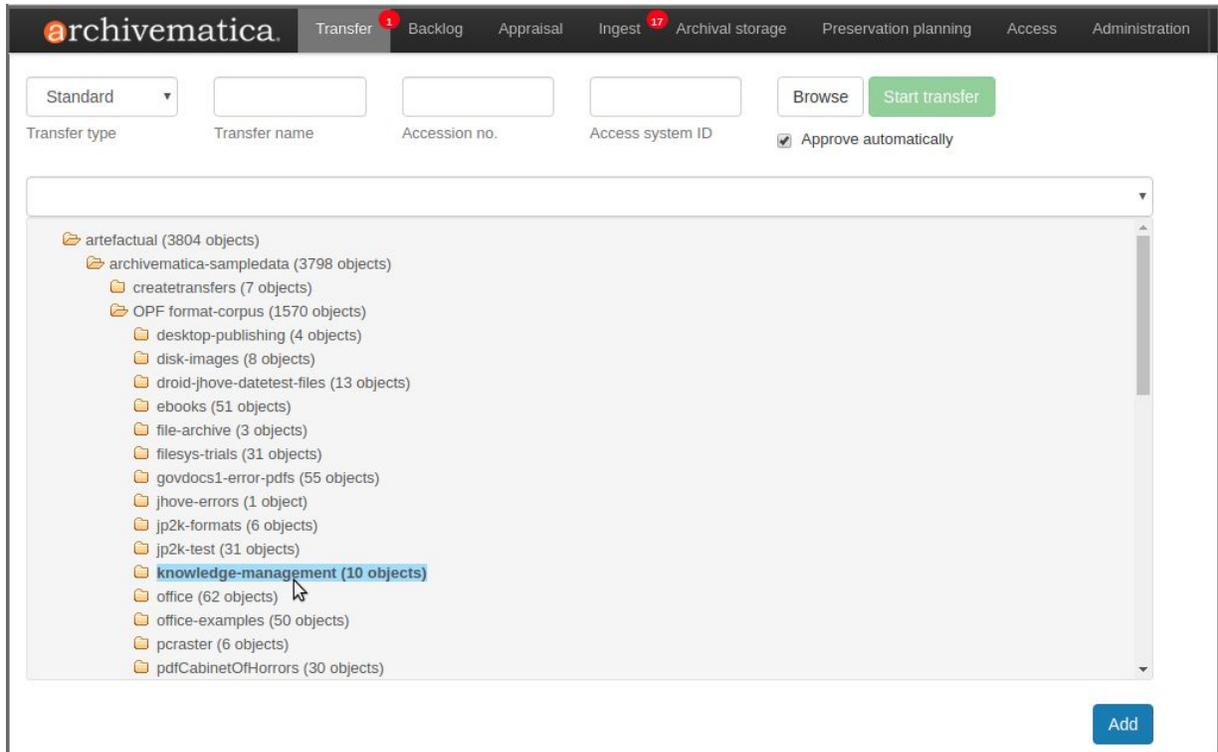


Fig. 1 Using the web-based dashboard to browse content for ingest.

Can I ingest a directory with subdirectories?

Yes. There is no known limit to the number of levels in a folder hierarchy that can be ingested. Archivematica will recognize and record the original folder structure of the hierarchy, even if the user subsequently rearranges the folders and their contents.

Can I ingest content from external media directly into Archivematica?

No. Content on external media must be transferred to a location accessible to Archivematica (such as a designated directory in a network file system) prior to ingest.

What kinds of content can be ingested?

Any content that can be placed in a directory can be ingested, including images, text files, office documents, publications, audiovisual files, databases and research datasets. Archivematica performs digital preservation actions such as identifier assignment, checksum generation, virus scan, format identification and metadata extraction on all ingested objects. For certain types of files, specialized tools (such as ImageMagick for raster images, ffprobe and ffmpeg for AV files, Ghostscript for PDF files, Inkscape for vector images, and bulk_extractor and sleuthkit for forensic disk images) are used for metadata extraction, format normalization and other preservation actions. New tools, rules and commands for preservation actions can be added to the workflow using the Format Policy Registry (FPR) via the Preservation Planning tab in the web-based dashboard.

Can I ingest zipped or other compressed packages?

Yes. Archivematica by default automatically extracts the contents of packages and processes the extracted contents like any other files. A record of the extraction is captured as a PREMIS Event. The user can easily

Archivemata digital preservation system

Information sheet 2019

configure the system not to extract packaged files, or to extract them and keep the original package in addition to the extracted files.

Does Archivemata crawl websites?

No. Archivemata does not include a web crawler, although it can ingest warc (Web ARChive format) files.

Does Archivemata ingest email accounts?

Yes. Archivemata can ingest mbox files, a common backup format and emerging digital preservation standard for email accounts. However, this is a relatively underdeveloped area in the digital preservation field and the full curation and preservation of email may require the use of other tools, such as ePADD, an open-source software tool being developed by Stanford University's Special Collections and University Archives. ePADD supports archival processes around the appraisal, ingest, processing, discovery, and delivery of email archives, and its progress is being closely watched by the Archivemata development team.

Does Archivemata handle forensic disk images?

Yes. Tools such as BitCurator or FTK are often used with write-blockers to transfer content from external media into disk images that Archivemata can process. Archivemata has specialized workflows and data entry templates to perform extraction, analysis and characterization of disk images.

How do I add descriptive metadata to my content?

Dublin Core descriptive metadata can be added to a SIP via a metadata entry template in the dashboard. The metadata entered in that way are for the SIP as a whole, rather than for individual files. Metadata can also be imported as CSV files, and those metadata can be for individual files as well as for whole SIPs.

How do I add rights metadata to my content?

PREMIS rights metadata can be added to a SIP via a metadata entry template in the dashboard. This is a full implementation of the rights module introduced in version 2.2 of the PREMIS Data Dictionary. As with descriptive metadata, rights metadata entered in that way are for the SIP as a whole, rather than for individual files, but the metadata can also be imported at both the SIP and the file level as CSV.

Can I import custom metadata into Archivemata?

Yes, you can include metadata files with your transfer; they will be saved in the AIP in a designated metadata directory and referenced from the METS file. However, unless they are Dublin Core they will not be mapped to a particular standard - if they are imported as CSV metadata they will be identified in the AIP METS file as "custom" metadata.

Does Archivemata assign identifiers to ingested objects?

Yes. A UUID (Unique Universal Identifier) is assigned to each digital object automatically during ingest. Archivemata can also be configured to make requests to a Handle System HTTP API so that files, directories and entire AIPs can be assigned externally-generated persistent identifiers (PIDS) and derived persistent URLs (PURLs).

Does Archivemata assign checksums to ingested objects?

Yes. Archivemata by default assigns SHA-256 checksums to all ingested objects.

Does Archivematica perform virus and malware scanning on ingested objects? What happens if a virus is found?

All ingested content is automatically scanned for viruses and malware using clamAV, a well-established open-source software tool that comes bundled with Archivematica. Virus definitions are updated on a regular basis. If a problem is found, Archivematica will show an error message in the web-based dashboard, send an email notification to specified users, and cease processing the transfer. The user must then remove the infected file(s) from the transfer or fix the errors before attempting to ingest the content again. If desired, users can set a quarantine period on ingested objects prior to scanning, in order to give ClamAV time to update its virus definitions.

What methods and tools are used to identify and characterize the formats of ingested files?

Archivematica runs micro-services to identify, validate and extract metadata from different types of files. Bundled tools for these micro-services include FIDO and Siegfried for format identification (both of which are linked to the PRONOM format registry), JHOVE and MediaConch for format validation, and numerous tools for metadata extraction (FITS, exiftool, ffprobe, ImageMagick, bulk_extractor and others). An administrator can customize tools and commands using the Format Policy Registry.

What methods and tools are used to convert ingested files to preservation file formats?

File format conversion, or normalization, is automated through the use of preservation commands entered into the Format Policy Registry, which is accessible to the user in the Preservation Planning tab of the web-based dashboard. Bundled tools for transcoding include ffmpeg, Inkscape, Ghostscript and ImageMagick. The FPR comes with hundreds of format-specific commands which can be edited by the user (although the majority of institutions use the system defaults, which are extensive). New normalization tools can be integrated into the FPR by an administrator if desired.

How do I manage and modify workflows?

Users can select workflow options during processing by interacting with the web-based dashboard (fig. 2). Decision-making can also be automated for multiple ingests using the workflow configuration menu in the Administration tab of the dashboard (fig. 3).

Archivematica digital preservation system

Information sheet 2019

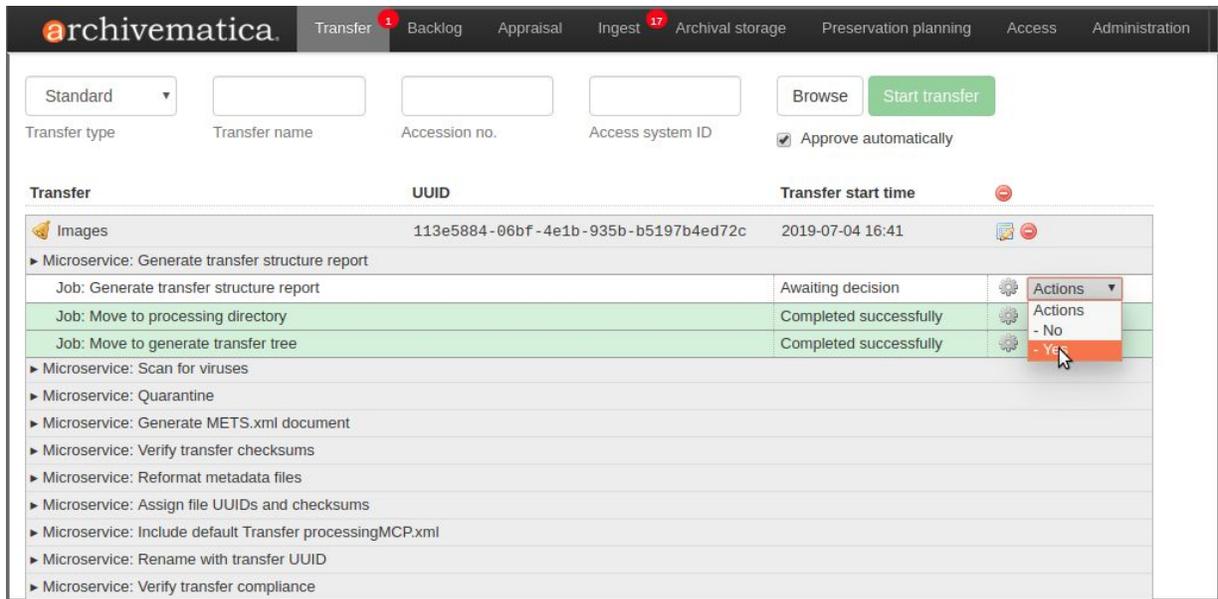


Fig. 2 A user makes a processing choice using the web-based dashboard. This choice can also be set as a default for automated processing.

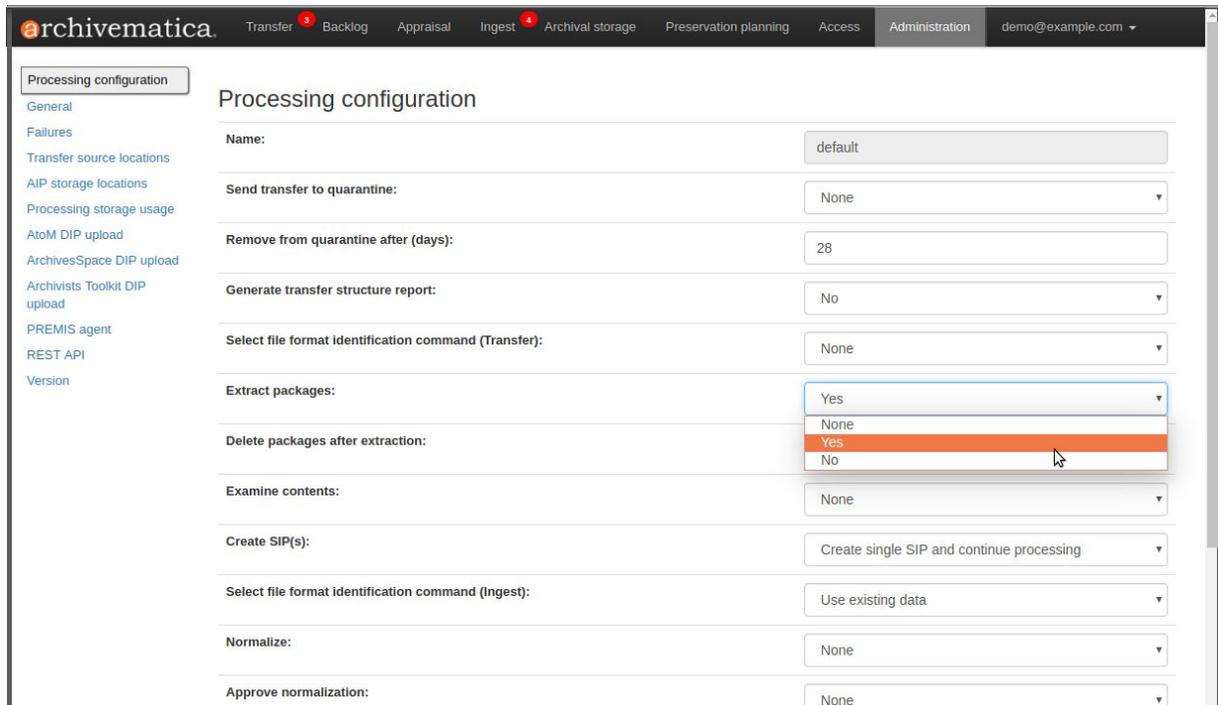


Fig. 3 Archivematica's workflows are easily configured in the Administration tab of the dashboard.

Can I modify or remove content after ingest but before the AIP is prepared?

Yes. Ingestion starts as the transfer of content into Archivematica. After initial micro-services are performed on all content (identifier assignment, checksum generation, virus scan, package extraction, format identification and metadata extraction) transfers can be placed in a virtual “backlog” before further processing if desired. The user can then retrieve one or more transfers from the backlog via the web-based dashboard, analyze the contents of the transfers, and arrange one or more transfers into one or more Submission Information Packages (SIPs). Not all of the content of the transfers needs to be added to the SIPs - some may be left in the backlog for permanent deletion or deaccession. A full record of the original content of a transfer is recorded as a METS file and a copy of the METS file is automatically placed into each SIP into which the transfer contents are arranged. Note that the files themselves cannot be modified in any way during this process.

What happens if there are errors during processing?

Errors are reported in the web-based dashboard as they occur (fig. 4), and certain types of errors trigger an email notification to registered users. A list of error reports is also maintained in the Administration tab of the dashboard.

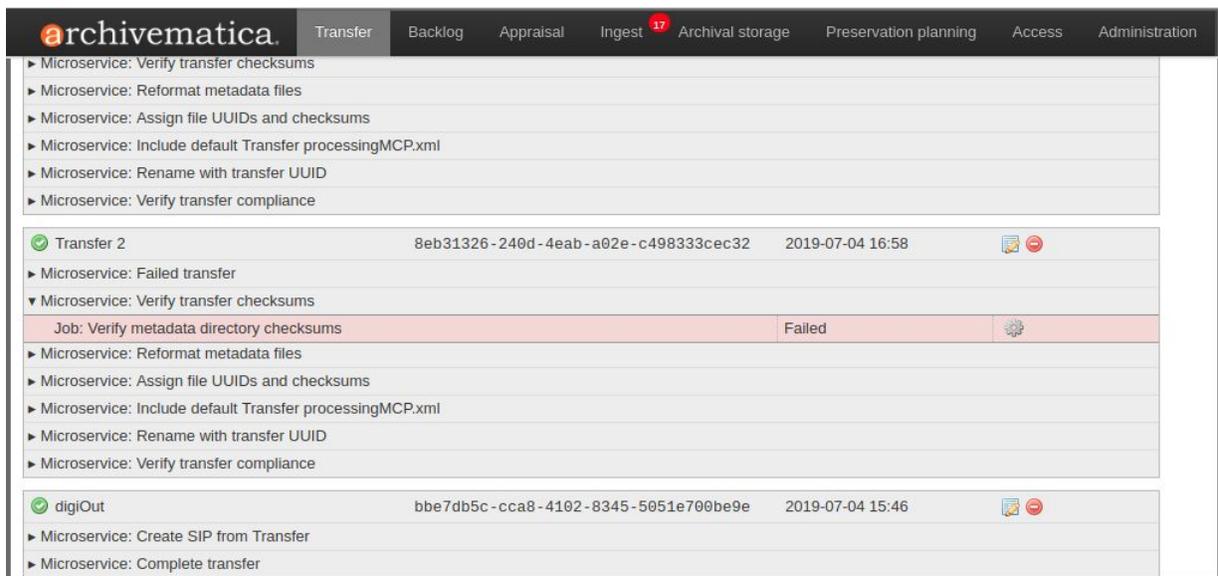


Fig. 4 A checksum verification error in the dashboard.

What types of errors are reported?

These are examples of errors that may occur:

- Fixity check: A fixity check may be performed on an ingested file that has a checksum from the originating system; if the checksums don't match, the error is reported and processing stops. This type of error will also be reported if files are ingested in a bag (i.e. packaged in the Library of Congress BagIt format) and the bag contents don't match the bag manifest.
- Virus scan: If a virus or other malware is detected an error is reported and processing stops.
- Format identification: If Archivematica can't match an ingested file format to the PRONOM format registry, it will report the unknown file in the dashboard.

Archivemata digital preservation system

Information sheet 2019

- Normalization: The dashboard will report normalization errors when normalization is attempted but fails or when no normalization is attempted and the file is not in a recognized preservation or access format.
- DIP (Dissemination Information Package) upload: If the DIP cannot be uploaded to the designated access system an error message will appear in the dashboard and the user will be asked to try again.

Does Archivemata maintain an audit trail of actions that occur during processing?

Yes. Actions are logged and recorded as PREMIS Events in the AIP METS files. Standard Events include fixity check, message digest calculation, unpacking (package extraction), virus scan, format identification, format validation, normalization, compression (if the AIP is compressed for storage) and encryption (if the AIP is encrypted for storage). For each PREMIS Event three PREMIS Agents are recorded: the name of the institution, the name of the software platform (i.e. Archivemata 1.x) and the name of the logged-in user.

What does an Archivemata Archival Information Package (AIP) consist of?

Fig. 5 shows an Archivemata AIP as it appears in a standard file browser. The AIP consists of a README file describing the AIP's structure and content; the ingested files; any preservation masters generated during processing; thumbnails for searching and viewing purposes; log files generated by some of the bundled software tools; and technical, preservation, audit and descriptive metadata encoded in METS XML. The METS file contains a robust PREMIS implementation, including the Object, Event, Agent and Rights entities, and Dublin Core is used as a descriptive metadata standard. The AIP contents are packaged into the Library of Congress BagIt format and can be compressed for archival storage or stored uncompressed.



Fig. 5 The Archivemata Archival Information Package (AIP) consists of a simple directory structure and can be

Archivematica digital preservation system

Information sheet 2019

opened in any standard file browser.

How do I find a stored AIP?

AIPs and their contents are searchable in the Archival Storage tab of the dashboard, using keyword/phrase searching with standard boolean operators. Fig. 6 shows a simple browse view of stored AIPs, and fig. 7 shows search results for individual files within AIPs.

The screenshot shows the Archivematica web interface. The top navigation bar includes: archivematica, Transfer (with a red notification bubble), Backlog, Appraisal, Ingest, Archival storage (active), Preservation planning, Access, Administration, and demo@example.com. Below the navigation bar is a search area with a text input, a dropdown menu set to 'Any', another dropdown set to 'Keyword', checkboxes for 'Show files?' and 'Show AICs?', and a 'Search archival storage' button. Below the search area is a link for 'Add New'. The main content area is titled 'Browse archival storage' with a subtitle 'Total size: 10.42 MB | Files indexed: 24'. Below this is a table with columns: AIP, Size, UUID, Date stored, Status, and Actions. The table contains six rows of data for 'Transfer1' through 'Transfer6'. At the bottom of the table area, it says 'Found 6 entries. Showing 1 to 6.'

AIP	Size	UUID	Date stored	Status	Actions
Transfer1	7.55 MB	d9a46c2e-781d-4cb2-be76-208547425a5b	2018-01-03 10:15	Stored	View
Transfer2	0.19 MB	a49de27a-9bfe-41d7-8dd5-17f82816554a	2018-01-03 10:21	Stored	View
Transfer3	0.93 MB	4278af6a-42d3-478d-b62a-ca28d870d716	2018-01-03 10:21	Stored	View
Transfer4	0.19 MB	09c08374-4b11-4f98-9365-0f010a1b9a9b	2018-01-03 10:22	Stored	View
Transfer5	0.63 MB	7a12ebdc-70be-413c-8206-ac9dc3ec90c5	2018-01-03 10:26	Stored	View
Transfer6	0.93 MB	c9e32815-dfe7-48c8-8f6f-d0d8d8d57ecb	2018-01-03 10:26	Stored	View

Fig. 6 The Archival Storage tab allows the user to browse and search for AIPs. Clicking on “View” brings up more details and allows the user to download or reingest the AIP.

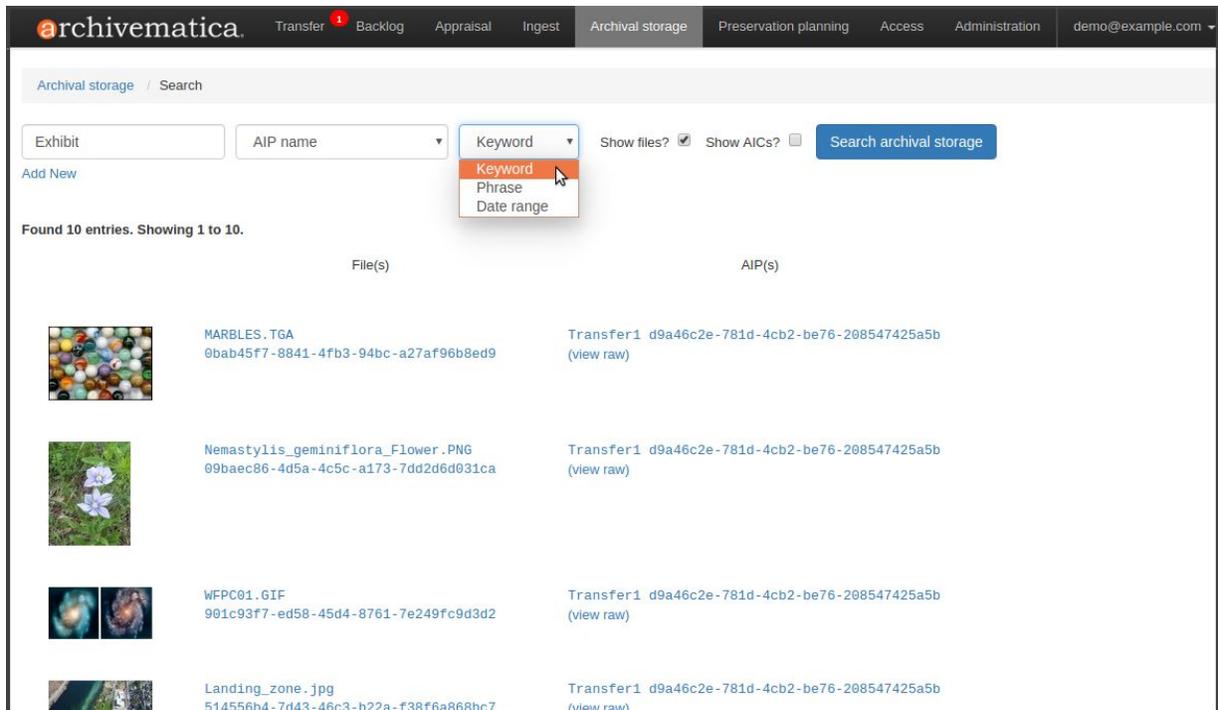


Fig. 7 A user has searched for files in the Archival Storage tab. The user can click on a thumbnail to download the file, or click on the AIP UUID on the right to retrieve the AIP as a whole.

Can I modify the contents of an AIP?

Yes. An AIP can be re-ingested via the Archival Storage tab for metadata enhancement or re-running of preservation micro-services such as format identification, metadata extraction and normalization. The AIP METS file is updated and all changes are versioned. Note that the ingested files themselves cannot be modified in any way.

Can I delete an AIP?

Yes. Deleting an AIP is a two-step process, to ensure that AIPs are not deleted by accident. The end-user requests a deletion and enters a rationale for the request. A Storage Administrator then approves the deletion request and performs the deletion using the Storage Service module, which generates a deletion log.

Does Archivematica perform periodic fixity checks on stored AIPs?

Archivematica itself doesn't perform fixity checks, but a separate command-line application does and reports results back to the Storage Service module. The Archivematica Storage Service provides a REST API endpoint for checking fixity of AIPs. This tool provides a simple command line interface to that functionality, in a way that's simple to automate.

Does Archivemata support records classification and retention scheduling?

No. Records scheduling and disposition are records management functions which are not supported by Archivemata. However, integration with records management systems is possible through the Archivemata Storage Service REST API.

What does an Archivemata Dissemination Information Package (DIP) consist of?

The DIP consists of access copies of the ingested digital objects and the METS file. The access copies have identifiers that link them back to the ingested originals and preservation masters in the AIP. Some of the metadata in the METS file, such as descriptive and rights metadata and information about the intellectual order of the objects, may be read and displayed by an access system.

How is a Dissemination Information Package (DIP) displayed in AtoM?

AtoM displays the DIP contents in the context of their descriptive hierarchies. The nature of those hierarchies (such as a collection with items or an archival fonds with series, subseries and files, etc.) is up to the AtoM users. For logged-in users, AtoM also displays a small amount of technical metadata about a digital object, as well as the persistent identifiers of both the object and its AIP.

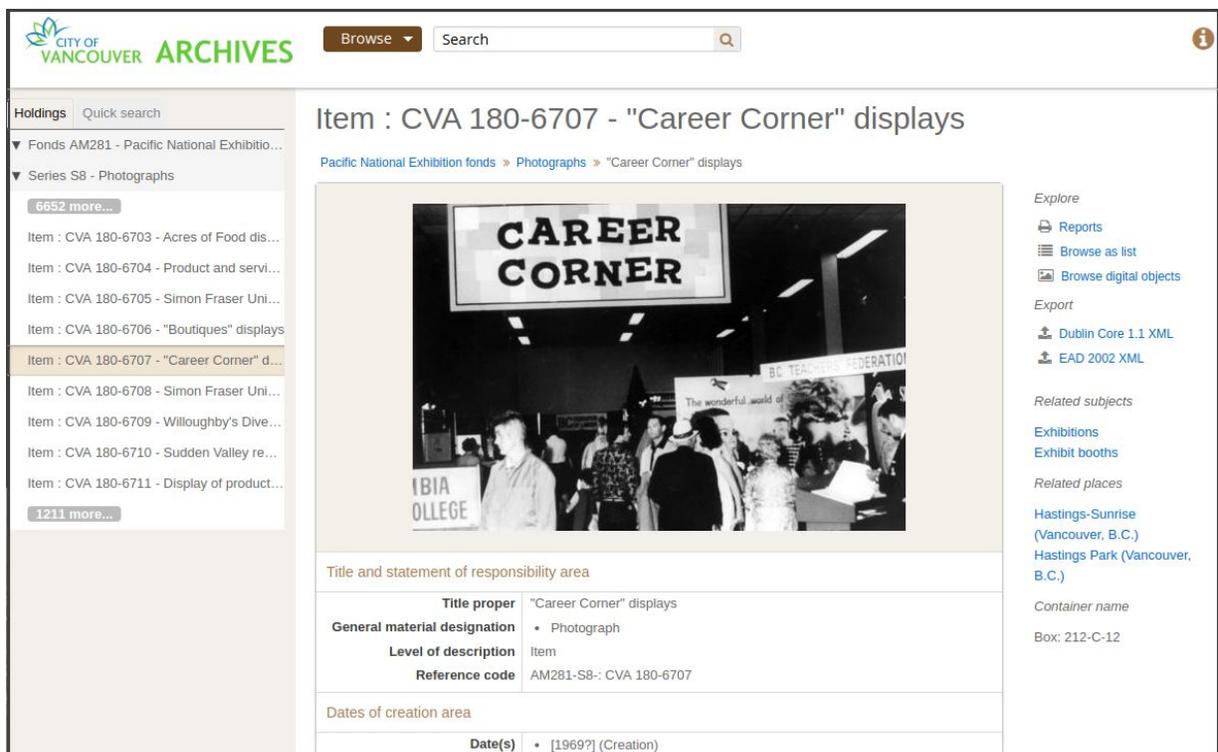


Fig. 6 An access copy of a preserved digital image displayed on the City of Vancouver Archives' AtoM site.

Does Archivemata monitor formats and provide preservation risk assessments?

This is not supported as out-of-the-box functionality. However, Archivemata captures format information as standardized PREMIS Object and Event metadata and can make these metadata available to reporting applications

Archivemata digital preservation system Information sheet 2019

via an API. The system also supports re-ingest of existing AIPs in order to implement updated preservation plans, including new normalization tools and commands.

Does Archivemata provide statistical reports? Can these be customized?

Out-of-the-box reporting in Archivemata is limited. For a given AIP storage location, the Archival Storage tab shows numbers of AIPs and files in storage, total volume of content stored, and dates placed in storage. The Storage Service module shows storage space use and availability for all storage locations. The ability to generate customized reports is not supported. Archivemata has APIs that can be used to supply information from the METS files to external reporting applications but no integrated reporting applications are available as out-of-the-box functionality.

Q&A: Scalability

Are there any physical limits on the size or number of files that can be ingested at one time?

Physical limits are dependent on compute resources allocated to the virtual machines on which the software is installed and on other infrastructure components such as bandwidth. Archivemata has been tested successfully with files of 1 TB+ in size and with transfers of 100,000 files. Minimal production requirements are considered to be 2 CPU cores, 4GB+ memory and disk space that is three to four times the size of the largest SIP to be processed. For example, for a SIP 200 GB in size the required disk space may be 600-800 GB. Note that this is processing space, not permanent storage, and that different filetypes might have different requirements (for example, highly compressed audiovisual files may need considerably more disk space for processing than research datasets).

Can Archivemata accommodate multiple ingests?

Yes. The Archivemata workflow engine implements Gearman, an open source application framework designed to distribute tasks to multiple computers, allowing for micro-services to be run in parallel or distributed to dedicated machines. Additional control of asynchronous processing is possible using optional add-on components such as the Archivemata Automation Tools. For large-scale processing, installation of multiple processing pipelines (Archivemata instances) is recommended. There is no known limit on asynchronous ingests.

Can ingest be automated to run continuously?

Yes. A suite of automation tools can be used to automate ingest using pre-configured settings. These tools can be configured to run as cron jobs or to run when a given task is completed (for example, ingest is triggered by placement in archival storage of the AIP produced by the previous ingest cycle).

Is there a maximum AIP file size or a maximum number of files that may be included in an AIP?

Archivemata has been tested with AIPs of up to 100,000 files. However, the desirable number of files may be based on considerations such as intellectual arrangement and amount of descriptive metadata for each file, the types of rights associated with groups of files, the format of the files and other considerations. Artefactual can assist the client to determine the optimal arrangement of files into AIPs. Additionally, Archivemata allows the user to create AICs (Archival Information Collections) which allow multiple AIPs to be aggregated into collections. Archivemata also supports different package transformation methods (compression, splitting/chunking) to allow large AIPs to be handled within the limitations of different storage subsystems.

Q&A: Integrations

What kinds of storage protocols and systems are supported?

Archivematica includes a component called the Storage Service module that provides a mechanism for connecting the digital preservation platform to disparate storage systems, using standard protocols such as CIFS, NFS, FTP and HTTP. Archivematica has been deployed in environments using file-based storage systems, MS Azure, Amazon AWS, LOCKSS, Duraspace, Chronopolis, Arkivum, DSpace, iRODS and OpenStack Swift. Additional storage platforms and protocols can be added by developing Storage Service plugins.

Can Archivematica ingest content from Microsoft products such as SharePoint and Exchange?

A suite of automation tools is available to ingest content from different transfer source locations and systems into Archivematica. Ingest from MSFT Sharepoint could be accomplished with customization of one of those tools; this could be a relatively trivial task, depending on the level of integration desired - for example, a simple hand-off between systems would be more easily implemented than one that included extensive metadata mapping. MS-Exchange Server can be integrated using a plugin or an automation tool, or through the use of external tools such as offlineimap to back up email accounts as maildir or mbox before transferring to Archivematica.

What access systems does Archivematica integrate with?

Archivematica has out-of-the-box integrations with AtoM (AccessToMemory), Dspace, ArchivesSpace and CONTENTdm. Configuration and workflow instructions are available in the online user manual. Archivematica can also be used with Islandora, Dataverse and Samvera, but some customization may be required for setup.

Q&A: System administration

Does Archivematica require any third-party applications in order to function correctly?

Archivematica bundles numerous open-source software tools to perform preservation micro-services such as format identification, format validation, metadata extraction and normalization. These tools are kept up to date with each public release of the software. Archivematica also requires the following: a mysql database server, which can be locally installed or provided as a service; a web server such as Nginx; gunicorn as an application container; Elasticsearch for indexing; and the Gearman message queue for asynchronous communications. The Archivematica dashboard and Storage Service module are accessed by users through standard web browsers with javascript enabled, such as Firefox, Chrome or MSFT Edge.

Do any required third-party applications require additional license fees or custom integration?

No.

Where are system logs stored and how are they kept secure?

Archivematica sends logs to a configurable location accessible only to authorized systems administrators. Logging can be configured to follow local security conventions; some institutions use external services such as Graylog to manage Archivematica logs, others employ an ELK stack (ElasticSearch, Logstash, Kibana) for this purpose.

What monitoring tools can be used with Archivemata?

Archivemata provides REST APIs that can be used to monitor and control the status of active transfers. Reference implementations of how to interact with these REST APIs is provided through freely available automation tools. The Archivemata dashboard provides a GUI that can be used to monitor and control active transfers as well. Tools such as Zabbix, Nagios, Zenoss or Pandora can be used to monitor standard performance metrics like cpu, memory and disk usage; login names and dates; number and volume stored of AIPs; Gearman job worker status; and other types of system information.

Can I create different kinds of user accounts with their own sets of permissions?

The dashboard provides a simple cookie-based user authentication system using the Django authentication framework. Access to the dashboard is limited to logged-in users only, and a login page will be shown when the user is not recognized.

Users can also be created, modified and deleted from the Administration tab. Only users who are administrators can create and edit user accounts.

Can Archivemata be integrated with LDAP and Shibboleth for single sign-on?

Yes, both LDAP and Shibboleth integrations are supported.

How is Archivemata protected against intrusion?

Archivemata is not a public facing application; the default configuration therefore assumes that no general access is required and uses whitelisting for exceptions. The application components (i.e. the dashboard application and Storage Service module) each have their own sets of users, allowing access to be restricted by function. Archivemata can be deployed using configuration management scripts that are explicit about details such as firewall rules, and enabling services are used to provide "defense in depth" along with a perimeter firewall. Archivemata explicitly manages its own dependencies, with a focus on compiling binaries in a build environment and deploying as packages to production servers, which reduces the amount of software required in production environments. For hosted instances of Archivemata, all software is hosted and all data stored in third-party data centres, and information security is governed by the policies of service providers such as Amazon AWS, OVH and MSFT Azure, all of which have robust, well-documented security protocols.

How can my data be recovered in the event of a system failure, failed update etc.?

In Archivemata all metadata are written to the AIP METS file. In the event that Archivemata is unavailable or corrupted, a new Archivemata instance can be installed and the AIP METS files re-indexed to restore the system in full. Since the AIPs are packaged in a standardized format they can also be opened in any file browser, or moved to another system capable of reading and parsing the METS files.

Archivemata uses MySQL and SQLite databases, which store information about ingested content that is still being processed. They also contain user information, configuration settings, storage locations and other ephemeral content that is not stored in the AIP. To ensure that an existing Archivemata instance can be restored (as opposed to re-indexing stored AIPs in a new instance) those databases should be backed up regularly.

Can Archivematica be run as a multi-tenanted platform?

Artefactual does not recommend running a single Archivematica instance as a shared service. User permissions are not granular enough to allow fine-grained access to specific institutional holdings, and multiple saved workflow configurations can be run simultaneously only by using automation tools.

About Artefactual Systems Inc.

Artefactual Systems was founded in January 2000 as a digital preservation consulting company, and expanded into open-source software development starting in 2006. The company now offers software development, technical support and hosting services, data migrations (from legacy systems into AtoM), training and consulting. The company currently has more than 90 hosting/technical support clients and at any given time has around 12- to 15 software development, data migration and consulting clients. The company has 25+ personnel, a diverse group which includes archivists and librarians, software developers, systems administrators, solution architects and technical services managers.

Find out more

End-user and administrator documentation: <https://www.archivematica.org/en/docs/archivematica-1.9/>

AtoM (AccessToMemory) project website: <https://www.accesstomemory.org/>

Code repositories: <https://github.com/artefactual>