

## The Choice For A Long Term Digital Preservation System or why the IISH favored Archivematica

*At the beginning of 2017 the IISH decided to use Archivematica as a central system for long term preservation of its digital collections. This decision was made after a thorough comparison between different systems and solutions. This blog gives an overview of this process.*

### Demands for the new digital preservation system

In 2016 the IISH started a project aimed to choose a system for the long term preservation of digital collections. The original project plan provided an overview of the gaps within the IISH digital infrastructure the new system would have to cover. From a bird's eye perspective, the system had to be OAIS ([https://en.wikipedia.org/wiki/Open\\_Archival\\_Information\\_System](https://en.wikipedia.org/wiki/Open_Archival_Information_System)) compliant and would enable the digital repository to reach Trusted Digital Repository (TDR) status (link :<http://www.trusteddigitalrepository.eu/Trusted%20Digital%20Repository.html>) From a more practical point of view the most important requirements for the system were:

- Able to follow the archiving workflow, delivering clear messages of the status of individual processes
- Able to perform appraisal and selection
- Able to use PREMIS (<https://www.loc.gov/standards/premis/>) for administrative and technical metadata
- Able to provide access to born digital and digitized material
- Able to manage access rights, copyrights and permissions
- Able to automatically OCR and index content
- Able to automatically and manually assemble DIP's
- Able to handle both a lot of very small and some exceptionally big files

Some general starting points:

- We aim for a modular system
- To work on the basis of migration, not emulation
- We aim for an open and transparent system – everything must be exportable (no vendor lock-in)
- No preference for open source or closed source software, access to enough technical support is more important
- We want to store born digital material exactly as we receive it
- No longer use too many custom made IISH scripts, but be able to profit from technical support or community input.

These requirements were used to create a first selection of suitable products.

### A first selection

A first selection of suitable preservation systems was made by the using the *Powrr* (Preserving digital Objects With Restricted Resources) *tool evaluation grid* which offers an exhaustive survey of digital preservation tools: <http://digitalpowrr.niu.edu/tool-grid> One of the strong points of this tool grid is that it takes OAIS functionality as a starting point. Based on the *Powrr* grid the IISH selected eight products for further inspection: Archivematica, Fedora, Preservica, Rosetta, Scope, Goobi, Roda and Vital.

The products were investigated on the basis of online documentation, demos, and user reviews. Of these eight choices, a few were taken off the list quite quickly. Goobi was more focused on digitized collections and not suitable for born digital. Roda seemed to be a good product, but lacked a wide user community. Vital and Resetta were too connected to (commercial) products, which we didn't want to use. Scope was out because we couldn't get a good impression of its users or performance. Fedora was longer on our list, being a widely used and a stable system. Fedora was installed but didn't have all the out-of-the-box functionalities that both Archivematica and Preservica had.

### And then there were two: Archivematica and Preservica

So, for us, only Archivematica and Preservica remained as serious candidates. To make a well informed choice between these products both were quite intensely tested and compared. Both companies behind the Archivematica and Preservica – *Artefactual* (<https://www.artefactual.com/>) and *Preservica Digital Preservation* (<https://preservica.com/>) respectively – paid the IISH a visit to better match the institute's requirements with the functionality requirements and the functionality of the products. Both systems were installed on IISH servers to thoroughly test how it would work and interact with other systems. This resulted in a list of selection criteria of which you can find a selection in the - still comprehensive - table below.

Concerning Preservica it is important to note that the IISH tested the enterprise version and not the cloud version. This is because of the IISH collection policy which states that collections have to fall under Dutch jurisdiction and therefore cloud storage outside the Netherlands is impossible. Also important to notice is that this comparison was made in the beginning of 2017 between version 1.4 of Archivematica and 5.7 of Preservica. Some of the conclusions might therefore be outdated regarding new versions of both products.

	Archivematica 1.4	Preservica Enterprise version 5.7
<b>Financial</b>		
	<ul style="list-style-type: none"> <li>- Artefactual development to help implement Archivematica</li> <li>- Optional Artefactual yearly support contract: \$25.000</li> </ul>	<ul style="list-style-type: none"> <li>- Implementation service package (one-time)</li> <li>- Yearly licence costs, including (developer's) support up to €100.000 a year for the enterprise version.</li> <li>- New versions of Preservica are at additional cost</li> </ul>
<b>Policy: open source vs closed source software</b>		
Flexibility	In principle the software is totally adaptable, but of course developments have to be in line with the Archivematica releases of the maintainer (Artefactual).	Even if it is closed source the Preservica SDK offers the possibility to adapt workflows and to connect with other software.
New demands	Open source, costs of development can sometimes be shared with other users and be part of the official release.	New demands can be proposed to the Preservica User Group ( <a href="http://preservica.com/preservica-user-group/">http://preservica.com/preservica-user-group/</a> ). These might or might not be implemented by Preservica. Some functionalities can sometimes be custom developed. But this is only applicable to that which does not belong to the core application. As the customer base is quite large, and growing, the influence of an individual institution is small. So if you want

		to realize some new piece of functionality within the core application you will have to forge coalitions with other institutions to really get the point across.
IISG software policy (mostly open source, importance of having grip on IT processes, not too many custom made IISH scripts/tools)	Archivematica will fit well into the broader IISH infrastructure and will contribute to transparent and well managed IT processes. It might cost a bit more in development costs to adapt the system specifically to IISH needs.	Though the SDK gives the user a lot of the workflows and connection to the IISH infrastructure, there is some doubt as to whether the IISH can truly get a grip on the IT processes within Preservica and on the future development of the product. Preservica is more of a finished product, but it would still need development work to connect it to other IISH applications.
Exit strategy: Is the AIP (including all metadata) independent of the archival software? Can all information be exported?	Yes, Archivematica natively uses the standards METS and PREMIS. The AIP can always be exported.	Yes. Preservica internally uses its own XIP metadata, but this can be exported to METS and PREMIS. This is part of the software so no custom work is necessary.
<b>Technique</b>		
Application code language	Python	Java
Construction of the software	<ul style="list-style-type: none"> <li>- Modular</li> <li>- The workflow is run by the use of microservices; small software tools with a specific task which can be switched on or off and configured individually. The microservices are connected by use of the Gearman application framework.</li> </ul>	<ul style="list-style-type: none"> <li>- Preservica is a so called "Boiler plate" application: standard and logically build, using open source libraries.</li> <li>- Preservica also makes use of the microservice structure and Gearman.</li> </ul>
<b>Performance</b>		
How to deal with the ingest of a lot of small files or big files.	This is dependent on the hardware used during ingest and the smart configuration of the workflow. This has not yet been thoroughly tested.	Idem.
<b>Infrastructure: Integration with IISH systems</b>		
Storage system(s) and backup	The Archivematica storage service connects every "pipeline" to external storage.	Storage adapters connect to a great variety of storage systems.
Connection with IISH acquisition database	Standard connection is possible with Archivespace and Atom, for other systems development is necessary.	N/A. Preservica begins with the ingest of the SIP.
Possibility to connect to PID Handle system	No, development necessary.	No, development necessary.
Metadata systems (Evergreen, "EAD" via X METAL)	<ul style="list-style-type: none"> <li>- Not standard, development necessary.</li> <li>- Archivematica has a standard connection to Archivespace or Atom.</li> </ul>	<ul style="list-style-type: none"> <li>- Can make use of Preservica sync workflow for an update from catalogue to AIP metadata update (catalogue sync).</li> <li>- Preservica has a standard connection to Archivespace.</li> </ul>

Future systems: IIF and rights management system	Not standard, development necessary.	Not standard, development necessary.
<b>Permissions</b>		
How granular are access rights in the system?	By default the only difference that is made is between an "active" and "admin" user.	Different roles (which can be expanded) can be combined with different access rights.
LDAP support	Not standard, development necessary.	Yes
<b>IISH preferred standards used</b>		
PREMIS	Yes	No, but conversion to PREMIS is supported.
METS	Yes	No, but conversion to METS is supported.
<b>Workflows</b>		
Amount of freedom to configure a workflow	Highly adaptable. Tools can be added to the default tools and workflows can be customized.	Workflows are XML based and can be customized. It is not possible to add your own tools.
Is it possible to monitor the workflow and receive messages?	Yes	Yes
How well will the product fit with existing IISH workflows?	Expectation is that existing workflows can be translated to Archivematica without much trouble.	A test with the IISH digitisation workflow made clear that it was relatively simple to translate an existing workflow to a Preservica workflow.
<b>Pre-ingest</b>		
Is appraisal and selection possible?	Yes. Version 1.7 has a special tab for appraisal and selection included.	No. Preservica functionality starts with the ingest of a SIP. This can be created by a separate SIP creator application (made by Preservica).
Transfer of offloaded files to a pre-SIP staging area?	Yes, via the transfer tab.	No
Offers the software any solutions for the transfer of files from the archival donor or scanning service to the archive?	No	No
<b>Ingest</b>		
Is it possible to configure error messages?	Yes	Yes
Is it possible to configure if the ingest process stops after an error message?	Yes	Yes
Is it possible to configure which tools are active for	Yes	Yes

different workflows?		
Is it possible to add your own ingest tools?	Yes	No
Can the system handle long file names and deep directory structures?)	Yes	Yes
Can the system handle non-western languages (Unicode support)	Yes	Yes
Are we able to decide to which file formats we want to normalise?	Yes, via normalization rules and commands (under the Preservation planning tab).	Yes, via the normalization workflow and the migration pathways.
<b>Post-ingest</b>		
Is it possible to supplement or to change an AIP?	No. This comes from the principle idea that an AIP should never change. If, for instance, during a preservation action new preservation copies are made, the AIP with the new files will have to be re-ingested again.	Yes, but only for descriptive metadata. The objects and other metadata cannot be changed. Also, a re-ingest is necessary if files are added to the AIP.
Can an AIP be deleted?	Yes, but only by someone with the correct access rights.	Yes, but only by someone with the correct access rights.
Possibility for reporting on processed AIP's, DIP's, formats, preservation actions, etc	Yes, but only to a certain degree. This was not thoroughly tested.	Yes, Preservica by default can produce all kinds of reports.
<b>Dissemination</b>		
Does the product disseminates the content?	Archivematica itself doesn't offer access to the material. The DIP is placed on a server from which other access applications will have to deal with the dissemination.	Preservica itself doesn't offer access to the material. But Preservica can offer a separate dissemination web application.
Can you give access on the basis of the access rights (specific user, reading room only)	No. This has to be arranged outside of Archivematica.	No. But there is CMIS support.
Is it possible to connect to a rights systems?	Yes, but only after development	Yes, but only after development
CMIS support	Yes, but only after development	Yes
<b>"Nice to have" functionality</b>		
Email archiving	Yes, Archivematica can ingest different email formats as PST and MBOX (and if necessary normalize).	Yes, Preservica can ingest different email formats as PST and MBOX (and if necessary normalize). There is also a separate email processing workflow available.

Web archiving	Not a separate workflow for web archiving, but it can ingest WARC files.	Yes, via a separate web archiving module.
OCR service	Yes	No
<b>Quality of the documentation</b>		
	Not always easy to find and not always up to date.	Exhaustive and up to date documentation.
<b>User friendliness</b>		
SIP creation	Part of the application. Also appraisal and selection is possible.	Happens in another application (SIP creator).
Ingest process	- Start-up and following of the workflow is a relatively intuitive process. - The ingest workflow is not connected to the browser session.	Idem
Preservation planning	A relatively intuitive process.	A relatively intuitive process.
<b>Stability of the software</b>		
Regular updates	Yes. Artefactual and a few of the bigger users are the driving force behind the updates.	Yes. The user group has influence on the roadmap.
Stability company and community	Archivemata has a healthy community of users. The stability of Artefactual is unknown.	Preservica seems to be a healthy company with a big clients and diverse client base. Has recently split off from Tessella as a daughter company.
<b>Research data</b>		
Dataverse support	No, it is planned for 2017.	No
<b>Community</b>		
Forum (lively?)	Yes, seems like a lively discussion forum: <a href="https://groups.google.com/forum/?fromgroups#!forum/archivemata">https://groups.google.com/forum/?fromgroups#!forum/archivemata</a>	Is a closed off online forum where experiences and plugins can be shared. Unknown how lively the user forum is.
User group	There is a user group in the UK: <a href="https://wiki.archivemata.org/Community/Regional_User_Groups">https://wiki.archivemata.org/Community/Regional_User_Groups</a>	Yes.
Nearby users	No, not really. Users in the UK and Germany (Berlin) are the closest.	Yes, National Archives of the Netherlands, Dutch regional archives (RACs - tenants of the NA)
Users from which domain mostly?	Universities	Archives (government, companies)

### There can be only be one: Archivemata

As can be understood from all the criteria listed in the table above the choice was certainly not an easy one. Both products are able to offer complex OAIS preservation workflows, for a wide variety of materials and could both fit within the IISH infrastructure. They both meet the requirements mentioned above. Not included in the above table were the internal arguments concerning how much time our developers would have to invest in the implementation/development and structural technical management of the two systems. As difficult as it was to predict this at the time the feeling was that it would be more or less the same for both.

In the end, and in *essence* what made us choose Archivemata were the following points:

- More functionality in the pre-ingest phase (the transfer tab). As a private archive we have little influence on how archives are transferred. Therefore the extra functionality that Archivematica offers for a first check and appraisal are very welcome.
- Lesser costs: of course Archivematica, as with all open source software, does not come for free. Even without being able to account for the *total cost of ownership* for Archivematica in advance, the difference with the yearly licence costs of the Preservica Enterprise product are considerable. For an organization the size of the IISH this is an important point.
- Preservica may be the more ready and finished product, however the differences are not so big that this could be – for the IISH – enough of an argument to choose Preservica over Archivematica. Besides, both systems will require the same (one-time) development costs to connect them to the other IISH systems. The company behind Archivematica has the advantage that they are willing to do all kinds of development work for us, and also more IISH custom work, while the Preservica company only wanted to work on the Preservica core software. This was an important argument as it would relieve some of the pressure on our small IT department.
- Although the choice for open source software was not a decisive argument in this case, the choice for Archivematica means that any money the IISH invests in new Archivematica functionalities will also benefit the rest of the Archivematica community. As the IISH is a publicly funded organisation this serves as an extra argument for Archivematica.
- On the whole Archivematica gave the feeling that it fit better in our IT infrastructure, that we could have more influence on its development and it would be easier to adapt to our needs.