

# The Archivematica Project

*Meeting Digital Continuity's Technical Challenges*

**Peter Van Garderen,<sup>1</sup> P. Jordan,<sup>2</sup> T. Hooten,<sup>2</sup> C. Mumma<sup>1</sup> and E. McLellan<sup>1</sup>**

<sup>1</sup>Artefactual Systems

<sup>2</sup>International Monetary Fund

## Abstract

*Archivematica is a free and open-source digital preservation system developed by Artefactual Systems in part through funding from UNESCO. This interactive session will present the history of Archivematica and an overview of the system's design, features and technical architecture. It will also provide examples of how the software is being used in projects worldwide, such as the International Monetary Fund, the University of British Columbia Library and the City of Vancouver Archives. The presentation will conclude with a discussion of future directions for the system, including the development of new features; the open-source business model; training and support services; and ongoing challenges to implementing Archivematica in developing countries.*

## Authors

Peter Van Garderen is president of Artefactual Systems Inc., which he launched in 2001 to provide technology analysis and implementation services to archival organizations, in particular those that want to develop and implement open-source solutions. Since that time Peter has worked with a wide range of clients, providing services that range from writing strategy reports, analyzing system requirements, designing technology architectures, developing software, and managing open-source projects. He is a Doctoral Candidate in Archival Science at the University of Amsterdam and a Distinguished Alumnus of the University of British Columbia's Master of Archival Studies (1997). Peter also holds a Certificate in Software Engineering (2001) from UBC. He is a regular conference speaker on the topics of digital preservation and archives technology.

Paul Jordan worked for four years as the Digital Archivist for the International Monetary Fund. Prior to that he worked as an archivist for both Occidental College and the Mary Pickford Foundation. He received his Master of Science in Information from the University of Michigan, and is currently working as the Digital Asset Administrator for Kaplan University.

Evelyn McLellan is systems analyst and consultant on Artefactual's digital preservation projects. She provides end-user support, user documentation, quality assurance testing and system requirements management for Artefactual's Archivematica project. Evelyn is a graduate of the University of British Columbia's Master of Archival Studies program (1997) and has over 10 years experience as an archivist and records information analyst. She was a co-investigator on the International Research on Permanent Authentic Records in Electronic Systems (InterPARES Project) and served as an Adjunct Professor at the University of British Columbia's School of Library, Archival and Information Studies. Evelyn has written several articles and delivered a number of presentations and workshops on the topics of digital preservation and open technologies for archives.

Courtney C. Mumma is a systems analyst who provides end-user support, user documentation, quality assurance testing and system requirements management for Artefactual's Archivematica project. As a digital archivist at the City of Vancouver Archives, she helped to develop and implement their digital

archives system while managing the acquisition of the hybrid digital-analog 2010 Winter Games archives. She has been a researcher and co-investigator on the International Research on Permanent Authentic Records in Electronic Systems (InterPARES 3 Project), researcher on the UBC-SLAIS Digital Records Forensics Project, and a member of the Professional Experts Panel on the BitCurator Project. Courtney has been published in *Archivaria* and has delivered a number of presentations and workshops nationally and internationally on the practical application of digital preservation strategies.

## 1. History of the Archivemata Project

In June of 2007, Kevin Bradley of the National Library of Australia with Junran Lei and Chris Blackall of the Australian Partnership for Sustainable Repositories, published “Towards an Open Source Repository and Preservation System: Recommendations on the Implementation of an Open Source Digital Archival and Preservation System and on Related Software Development” for the UNESCO Memory of the World Programme Sub-Committee on Technology. Bradley et al. advocated building sustainable systems instead of expecting some permanent storage media to solve digital preservation challenges. Their report defined open source software requirements for the implementation of a digital archival and preservation system that would consider all aspects of a digital repositories as defined by the ISO 14721 Open Archival Information System (OAIS) functional model;<sup>1</sup> Ingest, Access, Administration, Data Management, Preservation Planning and Archival Storage, including storage media and management software. Further, the report claimed that digital preservation solutions for simple digital objects were well understood, and that “what is needed are affordable tools, technology and training in using those systems.”<sup>2</sup>

The Sub-Committee identified existing gaps and made recommendations for the development and packaging of an Open Source Digital Preservation System. Ultimately, they concluded that what was needed was an affordable, sustainable approach that could leverage the expertise of larger institutions with more resources to innovate and share solutions with the digital preservation community at-large. Such collaborative innovation could look to the open source software development community for a model of “how a sustainable archival system might work, be sustained, be upgraded and be developed as required.”<sup>3</sup> Most significantly, the report recommended that UNESCO support “the aggregation and development of an open source archival system, building on, and drawing together existing open source programs.”<sup>4</sup>

---

<sup>1</sup> ISO 14721:2003, Space data and information transfer systems – Open archival information system – Reference model (2003).

<sup>2</sup> Kevin Bradley of the National Library of Australia with Junran Lei and Chris Blackall of the Australian Partnership for Sustainable Repositories, published “Towards an Open Source Repository and Preservation System: Recommendations on the Implementation of an Open Source Digital Archival and Preservation System and on Related Software Development” for the UNESCO Memory of the World Sub-Committee on Technology.

<sup>3</sup> *Ibid.*, 3.

<sup>4</sup> *Ibid.*, 8.

Concurrently, Artefactual Systems, Inc., was busy developing their generic Qubit<sup>5</sup> information toolkit software as AtoM (Access to Memory), an open source, web-based archival description software based on International Council on Archives (ICA) standards. The UNESCO report coincided with the Artefactual team, some of their clients and the digital preservation community at large realizing there was a need for an open-source, sustainable, OAIS-based digital preservation system.

Therefore, the Archivemata project had its beginnings as the back-end digital preservation system for ICA-AtoM, and was originally referred to as “Qubit-OAIS”. Over time, though, the development team recognized that the direct association with ICA-AtoM may be too exclusive, obscuring the larger goal to allow Archivemata to integrate with other systems. Therefore, qubit-oais became Archivemata, an open-source digital preservation system designed for standards-based, long-term access to digital materials.

## **2. City of Vancouver Digital Archives Project**

With UNESCO, the City of Vancouver Archives was one of the first institutions to allocate resources for Archivemata development. The objective of their project was to establish a prototype digital archives environment and provide direction on the management framework within the City of Vancouver Archives (CVA) to implement and sustain a digital archives. The CVA is responsible for permanently preserving archival records created by the City of Vancouver and its various boards and agencies. It is also responsible for acquiring the archives of private-sector individuals and organizations within the constraints imposed by its acquisitions mandate. Increasingly, many of the records created by these various bodies exist only in digital form. The CVA recognized their responsibility to ensure that it had adequate policy infrastructure and technical capacity in place to be capable of permanently preserving and providing access to authentic and reliable digital records. To meet this responsibility it partnered with Artefactual Systems, Inc. and launched the Digital Archives project.<sup>6</sup>

The Digital Archives project focused on problems related to preserving municipal digital records created within the City's Electronic Records and Document Management System, called VanDocs, as well as digital records created outside of the VanDocs environment, in particular, records created and/or maintained by individuals and organizations from the private-sector in recordmaking and recordkeeping systems that the Archives had no control over. The diversity of records and recordkeeping systems in this prototype project were ideal towards developing a system that could adapt to a multitude of memory institutions with different mandates and acquisition policies. Digital preservation goals may be similar industry-wide, but different types of digital objects and workflows are unique and plentiful.

## **3. Digital Preservation – An Overview**

In modern-day institutions, daily operations and communications are managed through the creation and exchange of digital information (e.g. business records, email, technical drawings). However, unlike paper records, which can sit untouched in boxes or filing cabinets for years or even decades without harm, digital records require specialized actions to manage and preserve them. In fact, the long-term accessibility, usability and authenticity of digital materials are at risk due to the inherent fragility and

---

<sup>5</sup> Artefactual Systems, Inc. website, Qubit.

<sup>6</sup> Vancouver Digital Archives wiki. [http://artefactual.com/wiki/index.php?title=Vancouver\\_Digital\\_Archives](http://artefactual.com/wiki/index.php?title=Vancouver_Digital_Archives). Last accessed September 2, 2012.

complexity of digital objects and to technological incompatibilities or obsolescence at the level of file storage, application software, metadata and file formats.

Digital records can easily be lost, deleted or modified; sometimes maliciously but more often through mishap (e.g. storage media failure, lack of proper backups in case of accidental deletion) or a simple lack of understanding that they are records of the organization that should be handled with as much care and attention as paper records have been in the past. Over time, some formats can no longer be read when the software that created them is upgraded or discarded. This can result in serious productivity issues or lost business opportunities to re-purpose and re-use digital assets.

Even if they can be read, the rendering of digital files may not be reliable; the "look and feel" may be altered or there may be data loss due to the fact that they are being rendered using different software or a newer version of the software that created them. Moreover, the records can easily become detached from their context: that is, they can be separated from their metadata or lose their links to other records that were originally created and maintained as part of the same business process. This means that even if an electronic record can be retrieved and read it may have compromised its authenticity and its evidentiary value in legal and regulatory proceedings.

For these reasons, the Archivematica project focused on maintaining accessibility, usability, and authenticity of digital information objects over space, time, and technology. To accomplish the task, they set about to build their system in compliance with the ISO-OAIS functional model and other digital preservation standards and best practices.

#### **4. OAIS Functional Model Analysis**

It was with the aforementioned digital preservation goals in mind that Artefactual Systems and the CVA began building what the UNESCO Sub-Committee had imagined. In late 2008, Artefactual and the CVA project team began conducting a comprehensive requirements analysis to establish minimal baseline functional requirements, policies and procedures for a digital archives system based on accepted standards. The initial round of requirements gathering started with the development of use cases based on the ISO-OAIS model.<sup>7</sup>

The OAIS is the de-facto standard for designing digital archives systems. Many digital preservation systems or projects claim to be "OAIS-compliant" and this was also a goal for the Vancouver Digital Archives project, but at the time it was difficult to trace requirements between the OAIS standard and systems that claim to be "OAIS-compliant". The detailed OAIS requirements analysis with its use case methodology to establish what the system requirements are for Digital Archives to be "OAIS-compliant" was an attempt to build traceability into the Archivematica project.

A simple use case methodology<sup>8</sup> was established to structure the use cases. Use cases were clustered around the same broad categories as the OAIS Functional Entities. Like the latter, use cases were organized into hierarchy with high-level scenarios broken down into more specific tests (sub- and sub-sub-cases). The use cases attempted to present plain language descriptions of what a Digital Archives should accomplish.

---

<sup>7</sup> City of Vancouver Digital Archives wiki, Requirements Analysis. [http://artefactual.com/wiki/index.php?title=Requirements\\_Analysis](http://artefactual.com/wiki/index.php?title=Requirements_Analysis). Last accessed September 2, 2012.

<sup>8</sup> Ibid., Use Case Methodology, [http://artefactual.com/wiki/index.php?title=Use\\_Case\\_Methodology](http://artefactual.com/wiki/index.php?title=Use_Case_Methodology)

Functional, metadata and technology requirements were derived from the use cases and from an open-source technology evaluation. The functional requirements specified what Archivemata should be able to do. Metadata requirements stipulated what data attributes had to be captured for each step. Technical requirements stipulated specific technical features, formats or protocols that had to be implemented. Policies and procedures were also derived from the use cases, in that they are developed to support all the steps the use cases contain. For example, a use case may state: "System implements disaster recovery policies such as duplication to remote storage facility"; successful completion of this step naturally requires the development of such policies.

The functional requirements were expressed as UML Activity Diagrams. A first set of these was based directly on the OAIS use cases without any additional interpretation. These were then revised as a second set of CVA-specific Activity Diagrams based on a business process and IT architecture analysis carried out by the project team as well as the ongoing technology and tools evaluation<sup>9</sup> and software integration and development work.<sup>10</sup>

In the course of the requirements analysis, the project team had an opportunity to become a part of the InterPARES 3 Project.<sup>11</sup> The team consulted with the InterPARES 3 Project to conduct a gap analysis between OAIS and the InterPARES 1 Project's Chain of Preservation (COP) Model.<sup>12</sup> Review of the model, along with consultations with archivists about processing analogue records, revealed that appraisal occurs in a few different stages during archival processing. This gap analysis led to use cases and UML Activity Diagrams which addresses appraisal requirements for Archivemata.

## 5. Archivemata

The thorough use case and process analysis by CVA and Artefactual identified workflow requirements to comply with the OAIS functional model. The resulting Archivemata system uses a micro-services design pattern to provide an integrated suite of free and open-source software tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAIS functional model. It allows digital preservation professionals to process digital transfers (accessioned, simple and complex digital objects), arrange them into Submission Information Packages (SIPs), apply media-type preservation plans and create high-quality, repository-independent Archival Information Packages (AIPs). Archivemata is designed to upload Dissemination Information Packages (DIPs) containing descriptive metadata and web-ready access copies to any access system (e.g. Dspace, ContentDM, ICA-AtoM, etc.). Users monitor and control the micro-services via a web-based dashboard.

Through deployment experiences and user feedback, including the gap analysis conducted with the InterPARES 3 Project, Archivemata has expanded beyond OAIS to address analysis and arrangement of transfers into SIPs and allow for archival appraisal at multiple decision points. The Archivemata micro-services implement these requirements as granular system tasks, which are provided by a combination of

---

<sup>9</sup> Ibid., Technology/Tools Evaluation. [http://artefactual.com/wiki/index.php?title=Technology/Tools\\_Evaluation](http://artefactual.com/wiki/index.php?title=Technology/Tools_Evaluation).

<sup>10</sup> Ibid., Software Integration/Development. [http://artefactual.com/wiki/index.php?title=Software\\_Integration/Development](http://artefactual.com/wiki/index.php?title=Software_Integration/Development).

<sup>11</sup> InterPARES 3 Project. [http://www.interpares.org/ip3/ip3\\_index.cfm](http://www.interpares.org/ip3/ip3_index.cfm). Last accessed May 21, 2012.

<sup>12</sup> InterPARES 2 Project, Chain of Preservation (COP) Model, [http://www.interpares.org/ip2/ip2\\_model\\_display.cfm?model=cop](http://www.interpares.org/ip2/ip2_model_display.cfm?model=cop). Last accessed May 21, 2012.

Python scripts and one or more of the free, open-source software tools bundled in the Archivematica system.

Archivematica uses METS, PREMIS, Dublin Core and other recognized metadata standards. The media type preservation plans it applies are based on an analysis of the significant characteristics of file formats.<sup>13</sup> Archivematica supports emulation preservation plans by preserving original bitstreams, and it will support migration preservation plans by monitoring at-risk file formats and providing a process to migrate them at a future date. Nevertheless, Archivematica's default preservation strategy is to normalize digital objects into preservation formats upon ingest, in order to make best use of the limited time that organizations will have to process and monitor large, diverse collections of digital objects. Building normalization paths into the software requires choosing target formats and integrating open-source tools to perform the migrations. The choice of preservation formats is based on four basic criteria, which will be familiar to many of those who have experience with digital preservation:

1. The specification must be freely available.
2. There must be no patents or licenses on the format. Archivematica's preservation formats are all open standards.<sup>14</sup>
3. Other established digital repositories should be using or have endorsed the format.
4. There should be a variety of writing and rendering tools available for the format.

Selection of preservation formats has been an iterative process of researching best practices, testing normalization tools, and, as far as possible, comparing before and after results of conversions by measuring significant properties. The choice of access formats is based on the ubiquity of viewers for the file format as well as the quality of conversion and compression.

Archivematica prepares a METS file for each SIP and packages it with the AIP. The purpose of the METS file is to capture, in a standardized way, information about all the objects that are being preserved. The METS file lists all of the objects in the AIP, categorizes their role (original, preservation copy, submission documentation, etc.), and allows an original object to be intellectually linked to its preservation copy. The METS file also includes a robust PREMIS (Preservation Metadata Implementation Strategies) implementation which provides highly detailed technical information about each object, an audit trail of actions taken on the object since it was ingested, and detailed and granular rights information.

## **6. Open-Source Software and Agile Development Methodology**

All of the software, documentation and development infrastructure are available free of charge and released under AGPL3 and Creative Commons licenses to give users the freedom to study, adapt and re-distribute these resources as best suits them. Archivematica development is led by Artefactual Systems, a Vancouver based technical service provider that works with archives and libraries to implement its open-source solutions as part of comprehensive digital preservation strategies. All funding for Archivematica development comes from clients that contract Artefactual's team of professional archivists and software

---

<sup>13</sup> Archivematica wiki, Significant characteristics evaluation.

[https://www.archivematica.org/wiki/Significant\\_characteristics](https://www.archivematica.org/wiki/Significant_characteristics). Last accessed August 3, 2012.

<sup>14</sup> Wikipedia definition of open standards [http://en.wikipedia.org/wiki/Open\\_standard](http://en.wikipedia.org/wiki/Open_standard).

developers to assist with installation, integration, training and feature enhancements. The majority of Archivemata users take advantage of its free and open-source license without additional contracting services.

Archivemata follows an agile software development methodology. Its micro-services model is malleable enough to allow for a rapid release cycle and iterative, granular updates to the requirements documentation, software code and end-user documentation. Artefactual clients and the Archivemata user community help to prioritize new features and bug fixes for each release.

## **7. Pilot Project Prototyping and User Experience at the International Monetary Fund**

One of the first Archivemata project clients was the International Monetary Fund (IMF). The IMF joined Artefactual and the City of Vancouver Archives in real-world testing of Archivemata. The IMF's experiences as users, as well as the confidentiality issues of digital archives the IMF is working through, have been invaluable to Archivemata's improvement.

Based in Washington, DC, the IMF is focused on analysing and reporting on world economic conditions, and providing loans to member countries when needed. The IMF Archives provides the institutional memory of the Fund. Their paper collections date back to the Bretton Woods Conference in 1944, which created the Fund and continue up to the present day. However, the IMF's digital archives is much newer. Though the Fund's network drives contain documents dating back to 1980, and external media up to and including boxes of punch cards have been found while processing paper collections, it's only been in the last few years that the Archives has been able, from a funding and expertise perspective, to begin to work on how to bring those records into the fold, preserve them, and make them accessible.

As Digital Archivist, Paul Jordan took the lead in the hands-on gathering and testing of collections. One of the main tools he used in his investigations was Archivemata, which was first installed in December 2009 and has been updated regularly ever since. IMF Archives used it in both prototype and pilot projects with a variety of source systems and file types, including legacy shared drives, current email mailboxes, and external media. All of the documents used came from actual collections brought in from various Fund departments.

Installation was simple since Artefactual uploaded the entire Archivemata platform to the Ubuntu repository, and so theoretically all you need is an Ubuntu machine and an internet connection, and a few commands will download the entire thing. In fact, the CVA got a couple of older computers which were on their way to the recycling centre after a computer refresh, sat them on a table, installed Ubuntu, and then downloaded Archivemata. After adding ethernet cables to link them together, they had a digital processing cluster on a table. Unfortunately, the IMF does not work that way. The IMF is a very security paranoid organization, and with good reason. No software is allowed into our IT environment, not even the development environment, without first putting it through a security accreditation process. This provided a challenge for the IMF Archives and Artefactual, since the software had not matured enough, and the project schedule was too short, to accommodate full testing.

Ultimately, IMF Archives created an isolated sandbox, a couple of virtual computers on the IMF's virtual server farm that were completely separated from the rest of the IMF's network, with the sole exception of one link out to a single network share which could be used to load files into and out of the sandbox. This worked fine from a security perspective, but it makes installation a bit cumbersome. One of the things that is very specifically blocked was any kind of internet access, which meant staff could not simply download Archivemata from the repository. Archivemata and all of its dependencies had to be

loaded onto external media and then moved into the sandbox. However, during the most recent installation, IT was able to temporarily open a port to the internet, and then shut it down as soon as the installation was done. That installation took about half an hour.

Over the course of our pilot projects, the IMF Archives focused on three source systems: departmental network drives, external media, and email. The network drives allowed for the widest variety of content, with files dating back to 1980, many of them in formats that were difficult to identify even with tools like Jhove and Droid. Files from the external media were much the same, though complicated by balky and sometimes corrupt media. The emails came from a departmental shared mailbox and were the most sensitive, as well as being the largest files. There were hundreds of files from the shared drives, but individually few of them were more than a couple hundred kilobytes, and gigabytes of email.

From its earliest days, Archivemata has been an invaluable tool for donor outreach. Even when it was still crude, it provided the IMF Archives with a concrete reason to reach out to departments in the Fund with collections of interest. The IMF Archives is a corporate archive without any external donors. All of their collections come from departments within the Fund. While this means that many departments already know the Archives exists, donor relations and networking is no less critical. Policies, procedures, and contacts for paper records are well established. Those for digital have barely begun, and while the existing Archival contacts form a firm groundwork, the specifics of transfer are still to be decided.

But once the IMF Archives had a pilot project with an actual working system that needed sample collections, it gave the Archives the perfect opportunity to talk to departments of interest. Outreach was overall successful. Some potential donors were extremely interested, and the Archives made friends in more than one department. People are starting to become aware of digital preservation issues, or at least worry about losing access to files and email, and they seem very willing to work with the Archives.

One hazard, however, is promising too much, too soon. The Archivemata version installed when these interviews occurred was not yet production ready. The Archives were not ready, either. It was a pilot project, and it was important to stress that to the people we were working with. It made for a very interesting tightrope to walk: trying to convince them that the Archives knew what they were doing, and that their records would be safe in their hands, without going so far as to take formal custody of the items and promise full archival processing and then access; things they did not yet have the procedures, the software, or the personnel to deliver.

For small and medium-sized institutions, Archivemata is a platform that ties together many smaller services, each with its own set of tools. When digital archivist Paul Joran got into the field of Digital Archives, he spent an entire summer trying to install and get working a few of the individual tools contained within Archivemata. However, he had extremely limited IT assistance and was not able to make much headway. In contrast, the IMF was able to acquire a single package that contained all of the software he had struggled with and more, all working together in the same direction already. IMF Archives was able to test the OAIS model against what they had and what they wanted to do with version 0.5 of Archivemata.

Archivemata is also a very flexible system that can support whatever workflows an organization might have. For the IMF Archives, one of those workflows that requires a great deal of effort and that is only just beginning to be addressed in the digital archives is classification.

The IMF is a very security-conscious organization. The Archives has a full-time declassification archivist whose role is the identification, removal, and processing of classified materials within the paper

archive, and none of the collections can be made available until this screening has been completed. The same thing will be true for digital objects.

Up to this point, the Archives has focused on records already open to the public. This has primarily been digitized archival collections: the entire repository of scanned documents from the Executive Board that are open to the public, some of the archival country files, and an oral history collection. Everything that has been digitized has already been screened and declared open; the Archives does not scan anything still confidential. No born digital collections have been made available yet, because it is in those extremely large legacy collections that will likely cause problems. When IMF Archives first started using Archivematica, it was not set up to handle both public and non-public documents, because Artefactual had never worked with a partner that had addressed requirements for such a need. A lot of the features around classification and document review were suggested by the IMF. Fortunately, the sandbox approach allowed the Archives to work with classified documents in isolation and to do some analysis.

Some of the documents at the IMF already have security classifications assigned to them. The Fund's document management system tracks classification, and there is also an Outlook add-in that does the same for email. Those will be the easy ones to identify. It will also be easy to determine which of the documents within that time period are public, or subject to automatic declassification, because lower classification statuses are also tracked. Older records from before these systems were implemented are more of a problem. Email will usually have classification in the header, but documents, for instance off of shared drives, will generally have to be opened to determine whether or not they're classified. Also, many emails have attachments that are themselves classified, and the classification of the two does not always match.

Therefore, one of the Archive's steps during appraisal will be a macro review of classification status. The hope is that based on provenance, archivists will be able to get a general idea of the quantity of classified documents, which can then factor into a collection's processing priority. It can also help divide SIPs for ingest; if a processing archivist can determine that everything outside a particular sub-directory is open, the classified subdirectory can be sequestered and everything else made available. Once the macro declassification appraisal has been completed, the collection will be appraised and processed normally. Only once processing is complete will archivists go back and do a second, item-level classification screening.

One of the things that may ease the process, and Archivematica is planning to implement for their 1.0 release, is full-text indexing of incoming documents. This will allow archivists to search for classification keywords and phrases. There are a large, though finite, number of terms that can classify documents; if staff can identify those documents that contain those words, they will be able to weed out a significant number of open documents. However, they will still need someone to go through the documents that have been flagged and determine whether each really is a "secret document," or whether it's an email where the sender is talking about his kids wanting a "secret moon base."

## **8. Archivematica 0.9 Beta Release Features and 1.0 Development Roadmap**

Beyond the IMF and CVA, Artefactual clients include the University of British Columbia Library, Simon Fraser University Archives, and the Rockefeller Archives Center. Based on their input, Artefactual's own research and goals, evolving best practices and requirements for digital preservation systems and the input from our user community at-large, our recent release included numerous improvements on the previous iterations. Working with pilot project implementers, the Archivematica team has gathered requirements

for managing a backlog of indexed digital acquisitions transfers, creating a SIP from a transfer or set of transfers, basic arrangement and description, preserving email, and receiving updates about new normalization paths via a format policy registry (FPR). After creating workflows that would account for real-world archival processing needs, these requirements were added to our development roadmaps for 0.9, 1.0 and subsequent Archivemata releases.

The first Archivemata beta release, Archivemata 0.9, became available for download from the Archivemata website in early September of 2012. In addition to fixing bugs and enhancing features, release 0.9 includes the following new features:

- An update to the Ubuntu 12.04 LTS as the base operating system.
- The web browser dashboard interface has replaced most of the file browser functionality.
- DIPs can be uploaded to CONTENTdm.
- All AIP metadata can be indexed and searched using a tool called ElasticSearch.
- The rights module is updated to the most current PREMIS implementation, PREMIS 2.2.
- Email handling is improved and there is a prototype ingest of maildir.
- User accounts can be created.
- Automatic restructuring of transfers for compliance.
- In the dashboard, jobs are grouped into micro-services.
- Ingest of Library of Congress Bagit format.
- Nightly backup of MCP MySQL database.
- Scalability enhancements.

Users and clients will continue testing and processing digital records using the 0.9 release, all the while informing the 1.0 release in early 2013. So far, there is a base set of features and enhancements for that release on the Archivemata wiki. Proposed features for 1.0 so far include:

- Develop a Format Policy Registry (FPR) and upload/download of format policy information between FPR and Archivemata instances.
- Upgrade file identification used as the basis to trigger format policy actions (aka 'preservation plans').
- Include a manual normalization workflow.
- Improve email handling.
- Add ability to edit format policies from preservation tab in the dashboard.
- Add ability to add/change format policies from FPR updates.
- Add a workflow for applying updated format policies to pre-existing AIPs.
- Include advanced search screens for searching AIP contents in the dashboard.
- Generate DIPs from the access tab in the dashboard.
- Include visualization of transfers.
- Include file-level Dublin Core and rights metadata entry.
- Include field validation in rights templates.
- Index transfers and identify/flag personal information. Evaluate BitCurator tool to determine how much functionality/data can be integrated/re-used prior to Archivemata ingest.
- Customize statistical reporting

- Improve AIP retrieval (whole or part) and delivery.
- Possibly remove packaging/compression for AIPs .
- Sync metadata between DIP and AIP for CONTENTdm, AtoM via OAI-PMH.
- Include AIP versioning (METS file updates).
- Include an enterprise ID service to connect AIPs and DIPs ([dns/uuid](#)).
- Enhance CONTENTdm DIP upload.
- Allow DIP upload to XTF.
- Transfer metadata from Archivist Toolkit.
- Ingest of TRIM exports.
- Automatic ingest from DSpace using OAI-PMH - OAI API for Archivemata dashboard.
- Management of persistent MCP metadata that does not end up in AIP.
- Make MCP processing workflows editable from the administration tab.
- Improve multi-node processing.
- Further scalability testing/prototyping.

## 9. Next Steps

The Archivemata project analysis and development described in this article are driven by practical demands from our early adopter community, including the initial vision from the UNESCO Memory of the World Sub-Committee on Technology. The alpha release prototype testing sponsored by our contract clients and shared by a growing community of interested users from the archives and library professions and beyond has provided the opportunity to spearhead the ongoing evolution of digital preservation knowledge in the form of a software application that is filling a practical need for digital preservation professionals.

At the same time, the digital curation community is also evolving and maturing. New tools, concepts and approaches continue to emerge. The Archivemata technical architecture and project management philosophy are designed to take advantage of these advancements for the benefit of Archivemata users and the digital curation community at large.

The free and open-source, community-driven model provides the best avenue for institutions to pool their technology budgets and to attract external funding to continue to develop core application features as requirements evolve. This means the community pays only once to have features developed, either by in-house technical staff or by third-party contractors such as Artefactual Systems. The resulting analysis work and new software functionality can then be offered at no cost in perpetuity to the rest of the user community at-large in subsequent releases of the software. This stands in contrast to a development model driven by a commercial vendor, where institutions share their own expertise to painstakingly co-develop digital preservation technology but then cannot share that technology with their colleagues or professional communities because of expensive and restrictive software licenses imposed by the vendor.