
AtoM 3 Proof of Concept Proposal and request for funding

Prepared by:

AIM 25 (UK)
Artefactual systems (Canada)
Docuteam (Switzerland)
Imagiz (UK)
Nothing Interactive (Switzerland)
Zazuko (Switzerland)

September 20, 2018

Table of Contents

Executive summary	3
How this document is structured	4
1. Why replace AtoM 2?	5
2. Design principles for AtoM 3	5
3. What is a Proof of Concept?	7
4. Building a Proof of Concept: detailed description	7
4.1 Example target users	8
4.2 Use case 1: converting existing archival descriptions to Linked Data	8
4.3 Use case 2: enabling improved archival description	9
4.4 Use case 3: federating and exposing data	10
4.5 Tasks for the Proof of Concept	12
4.6 Evaluating the Proof of Concept	13
5. Proposed technical architecture (high-level)	14
6. Proof of Concept cost estimate	15
7. Beyond the Proof of Concept: developing AtoM 3	15
Phase 1 - Proof of Concept	15
Phase 2 - Prototype development	15
Phase 3 - Beta development	16
Phase 4 - Final development and production release	16
Phase 5 - Maintenance	16
APPENDIX A: THE CASE FOR LINKED DATA	17
Better sharing of information	17
Easier aggregation and updates in multi-repository portal sites	17
More context, more meaning, more nuance	18
Better support for standardization	19
Meeting the need with AtoM 3	19
APPENDIX B: PROPOSED TECHNICAL ARCHITECTURE (DETAILED)	20

Executive summary

This document presents a proposal for building a Proof of Concept for a software application that would serve as a successor to Access to Memory 2.x (AtoM 2). AtoM 2 is an open-source, web-based, multilingual application designed to enable description and online discovery of archival holdings.¹ Built around the International Council on Archives' descriptive standards (ISAD(G), ISDIAH, ISAAR (CPF) and ISDF²), AtoM 2 has been implemented around the world as both single-institution sites and multi-repository portals at the regional and national level.

From August 23 to 25, 2018, a group of organizations dedicated to developing and supporting open-source software met at King's College in London to discuss feature and technical requirements for what is currently being called AtoM 3. This document is the output of that three-day workshop and is being presented as a request for funding of the Proof of Concept (PoC). Representatives of the following organizations took part in the workshop and jointly authored this document:

AIM25 (UK)

AIM25 is a charity which supports access to more than 150 archival institutions in London, including local authorities, universities and learned and scientific bodies such as the Royal Society and Wellcome Library. It aggregates and makes cross-searchable collection level descriptions of around 17,000 archives held by its members, and also manages the UK Archival Thesaurus. AIM25 has led or partnered in four Linked Data projects using archival data. <https://aim25.com/>

Artefactual Systems (Canada)

Artefactual specializes in developing and supporting open-source software for the cultural heritage sector. The company is the lead developer of AtoM 2 and, with 10 archivists currently on staff, has deep experience in archival practice and implementing archival descriptive standards. <https://www.artefactual.com/>

Docuteam (Switzerland)

Docuteam supports a wide range of organisations in their information management and archiving activities, covering the whole life cycle of information. The company also develops open source software for digital preservation in archives and research libraries. <https://www.docuteam.ch/en/>

Imagiz (UK)

Imagiz is a web design and development company with extensive experience helping cultural and academic organizations improve the management, availability and presentation of their

¹ <https://www.accesstomemory.org/>.

² <https://www.ica.org/en/public-resources/standards>.

archival materials. Imagiz specializes in AtoM data migrations and technical support, creating and managing large online collections for a number of high-profile UK institutions.

<https://imagiz.com/archios/>

Nothing Interactive (Switzerland)

Nothing Interactive is a web design company that supports clients in user-centered design to optimise the experience of the client's digital product. Nothing Interactive strives for a holistic approach and offers expertise in concept, design and development. Among others, Nothing Interactive works with the Swiss parliament to promote education and helps Swiss publishers like Klett in their digital transformation. Nothing Interactive also supports and contributes to different open-source technologies that enable better collaboration and/or eases content management. <https://www.nothing.ch/en>

Zazuko (Switzerland)

Zazuko, founded in 2014, is a consulting company for Semantic Web and related technologies. The company consults with various Swiss government agencies like the State Secretariat for Economic Affairs, the Federal Archive, the Federal Statistical Office and Federal Office of Topography and with enterprise customers in the communication, shipping and life sciences industries. Zazuko is co-creator of the open-source RDF.js stack for JavaScript, in collaboration with MIT CSAIL (under Sir Tim Berners-Lee) and IDLab at Ghent University. <http://zazuko.com/>

How this document is structured

The authors of this document propose that AtoM 3 be designed as a platform capable of supporting Linked Data while maintaining the functional capabilities of AtoM 2. "Linked Data" is a way of structuring data so that disparate sources of information can be linked together in ways that can be understood by computers. This allows for much more robust searching, data aggregation and data visualization, making multi-repository portals easier to update and allowing archives to make use of existing online resources such as authority records and subject terms. See [Appendix A](#) for more information about Linked Data and [section 2](#) for proposed design principles for AtoM 3.

The authors are seeking funding to build a Proof of Concept for AtoM 3. A Proof of Concept, or PoC, is a project designed to demonstrate the feasibility of developing a software application to meet the needs of defined stakeholders. It does this by testing that a set of desired software functions will work as expected when implemented with the chosen technical architecture and toolset. See [section 3](#) for a more detailed definition of a PoC.

The proposed PoC is described as a set of user personas, use cases, user stories and tasks. [Section 4](#) describes all of these aspects of the PoC in detail, and also defines the role of testers in evaluating the success of the PoC. [Section 5](#) proposes a high-level technical architecture designed to achieve the functional capabilities envisioned for AtoM 3, which will be tested

during the PoC; [Appendix B](#) provides a more detailed look at the proposed architecture and technology stack.

We are estimating that a functional PoC would require effort in the range of 600-750 hours and have provided a cost estimate for this in [Section 6](#). [Section 7](#) goes on to describe how the PoC could serve as the first stage of development for AtoM 3.

1. Why replace AtoM 2?

AtoM 2 is a re-architected version of ICA-AtoM, the development of which started in 2006. ICA-AtoM was introduced as a production system in 2008 and is still in use at some institutions around the world. AtoM 2 was designed to support greater scalability, better search capabilities and more robust data interoperability, but it was not a complete redesign of the system. ICA-AtoM and AtoM 2 are thus software applications that were designed more than 12 years ago, a long time in the software development world. Since then, new development frameworks, libraries and tools have been introduced that affect the design of software applications developed for the web, and the basic technology stack that makes up AtoM 2 has become deprecated and challenging to maintain.

In addition to advances in technology, changes in basic concepts of how data are created, displayed and exposed via search interfaces have led to fundamental changes in how an application like AtoM could serve the needs of archives. In 2016 the International Council on Archives' Experts Group on Archival Description released Records In Contexts (RiC)³, a Linked Data conceptual model for archival description which is designed to incorporate and be a successor to the ICA's existing archival description standards. The Experts Group is currently working on an ontology for the model. The release of RiC, and developments in other standards and ontologies for describing resources, points to fundamental shifts in the way archivists think about the ways in which records are described and made available to users. Software applications designed for archival description need to adapt to these shifts.

2. Design principles for AtoM 3

A chief design principle is that AtoM 3 will incorporate AtoM 2's current features while providing a better user experience and enabling organizations to share information through the use of Linked Data. AtoM 3 must also ensure that AtoM 2 data can be imported into AtoM 3 with relative ease. Moreover, although Linked Data is a new paradigm for archival description, the writers of this proposal are acutely aware that most archivists do not necessarily want to

3

<https://www.ica.org/en/records-in-contexts-ric-a-standard-for-archival-description-presentation-congress-2016>

learn new modes of description or abandon previous work in order to use a new software application. With this in mind, we have listed some of the key features of AtoM 2 below, and consider them to be the foundational features that must be supported by its successor.

- **Standards based:** the ability to create archival descriptions using templates based on the ICA's descriptive standards and additional standards such as the Canadian Rules for Archival Description, Dublin Core and MODS
- **Web based:** all end users need is a web browser and an internet connection, meaning they can access and edit their site from any computer
- **Digital object display** within the context of descriptive metadata - the user can navigate from thumbnail to reference display to download (if permitted)
- Editable **hierarchical taxonomies**
- **Multilingual:** ability to coordinate community translations for each release; ability to translate both content and user interface elements
- **Configurable:** many user interface elements can easily be modified (user interface labels, menus, controlled terms)
- **Hierarchical display:** the ability to generate and navigate a full-width treeview of archival descriptions
- Ability to create **custom themes**
- Ability to create **accession records** and link archival descriptions to them
- Ability to separate description and **actor records** and to embed relevant actor information in archival descriptions
- Ability to support **multiple repositories**
- Ability to **bulk import and export** descriptions, authorities and repository records in XML or CSV formats via the user interface

AtoM 3 should support the features described above, as well as the following:

- **Linked Data support:**
 - **Serialization of archival descriptions to RDF⁴**
 - Ability to semantically **link to external resources** such as name authorities, controlled vocabularies and archival holdings at other institutions
 - Ability for users to **dynamically query and display content** from disparate AtoM sites and other Linked Data generators
 - Ability to **expose descriptions to data aggregators** such as Digital Public Library of America, Europeana, ArchivesHub, ArchivesCanada, etc.
 - Ability to integrate with RDF **visualization tools** to support geo-mapping, histogram and other graphic data representations

⁴ RDF stands for Resource Description Framework, a standard model for data exchange on the web using Linked Data triples. See <https://www.w3.org/RDF/>.

- **Continuity:** templates and user interfaces that are familiar to archivists, and the ability to import archival descriptions from AtoM 2 and other EAD authoring tools
- **Data federation:** separation of the front-end and back-end, allowing content from multiple back-end instances to be displayed in a single front-end portal
- **Enhanced digital object handling:** better support for digital object viewers, including IIF⁵ viewers; integration with external systems such as DAMS and digital preservation systems
- **Web accessibility:** support for international standards and best practices for web accessibility to allow individuals with visual impairments to use AtoM 3

3. What is a Proof of Concept?

In software development, a Proof of Concept, or PoC, is a project designed to demonstrate the feasibility of developing a software application to meet the needs of defined stakeholders by testing that a set of desirable software functions will work as expected when implemented with the chosen technical architecture and toolset. To undertake a PoC, the design team must establish relevant stakeholder needs via use cases and user stories. The design team then defines a software function set and technology stack that is hypothesized to satisfy the stated stakeholder needs, but which must be tested by the PoC in order to prove the hypotheses. A PoC includes considerable analysis and design and just enough coding to make the solution testable by the designated stakeholders. A successful PoC will allow the design team to test the feasibility of the proposed technology stack and functionality; it will also allow the stakeholders to assess whether the proposed software design can satisfy their needs.

4. Building a Proof of Concept: detailed description

We aim to focus the AtoM 3 PoC on typical creators and users of archival descriptions. We therefore propose a set of use cases and user stories to meet the needs of the personas described below.

⁵ IIF is a set of specifications designed to enhance the ability of organizations to display their digital content online. See <https://iif.io/>.

4.1 Example target users

Persona	Motivation & frustration
Researcher 	<ul style="list-style-type: none">• My main motivation is to be able to intuitively aggregate different records from multiple sources and create my own collections in order to support my research.• My main frustration is discovering a new interface that looks better but lacks features that help me better search for and aggregate my data.
Archives manager 	<ul style="list-style-type: none">• My main motivation is to see that my organization's work in AtoM 2 is not lost and that the data my organization created can now be used in new ways we couldn't have imagined possible before.• My main frustration is having to allocate resources to learning a new interface or re-creating descriptions done in the past.
Archivist 	<ul style="list-style-type: none">• My main motivation is to be offered a tool that enables me to work with large volumes of data and be able to perform metadata authoring tasks faster with better results.• My main frustration is to feel limited by the tool rather than enabled.
Curator 	<ul style="list-style-type: none">• My main motivation is to have a powerful way of finding data, using precise and specific criteria, and to be able to show it off with a great virtual exhibit.• My main frustration is to be offered a powerful tool that is feature centered and doesn't attract visitors to our site.

4.2 Use case 1: converting existing archival descriptions to Linked Data

Assertions:

1. It will be possible for AtoM 2 users to upgrade to AtoM 3 without having to undertake time-consuming data migrations
2. It will be possible to convert traditional hierarchical descriptions to RDF
3. It will be possible to view the RDF descriptions in hierarchical and non-hierarchical visual representations

User	User story
	As an archives manager, when I implement AtoM 3 to manage and disseminate our archival data, I would like to easily migrate my archival descriptions from AtoM 2 without having to perform data mapping, data clean-up or other expensive and time consuming tasks, so I can protect my organization's investment in our current AtoM 2 data and minimize the costs of upgrading to AtoM 3.

How to meet the needs expressed in the user story
<ol style="list-style-type: none"> 1. Demonstrate that it is possible to import AtoM 2 descriptions into AtoM 3 using scripting and automation 2. Demonstrate that imported AtoM 2 descriptions can be serialized to RDF 3. Demonstrate how data can be represented hierarchically and non-hierarchically on the web

4.3 Use case 2: enabling improved archival description

Assertions:

1. It will be possible for AtoM 3 users to create archival descriptions more quickly and efficiently using automatic linking to external resources
2. It will be possible for AtoM 3 users to create archival descriptions with richer context by linking to external resources
3. It will be possible for AtoM 3 users to create archival descriptions with richer context by incorporating terms from Linked Data ontologies

User	User story
	As an archivist, when I have large volumes of records to describe, I want to be able to create qualified links to existing external resources so I can perform metadata authoring tasks faster, with fewer errors and richer context, so I can produce higher quality metadata and reduce demands on staff time.
	As an archivist, when I am describing archival materials with complex or nuanced provenance, I would like to create any number of qualified relationships to relevant actors, concepts or other descriptions so I can more accurately reflect the complexity of the provenance.



As an archives manager, I want to ensure that the descriptions my staff create are standards-compliant and make use of existing controlled vocabularies and authorities whenever possible, for better data accuracy and better discovery online.

How to meet the needs expressed in the user stories

1. Demonstrate linking and disambiguation using external authorities.⁶ For example:
 - display disambiguation (Kingston, Ontario vs Kingston, Jamaica vs Kingston, London) in a way that lets the user choose the right name
 - show how John Smith in Alberta and Smith, John in Quebec are the same person when both have been linked to external authorities
2. Demonstrate ways to enhance data entry by, for example:
 - auto-completing fields using external authorities and vocabularies
 - dragging and dropping descriptive entities to establish relationships
3. Demonstrate how existing archival descriptions can be enriched by incorporating terms from other ontologies and vocabularies⁷

4.4 Use case 3: federating and exposing data

Assertions:

1. It will be possible to use a single AtoM 3 site to aggregate data from multiple AtoM 3 sites into a data portal
2. Robust search functionality will allow users to retrieve results from one or more AtoM 3 sites and view them in on the web.
3. AtoM 3 will enable better digital object handling to enhance the end user experience
4. AtoM 3 will enable better discovery of archival descriptions on the web

⁶ Examples of external authorities include Library of Congress vocabularies (<https://id.loc.gov/>), SNAC (Social Networks and Archival Context, <http://snaccooperative.org/>) and VIAF (Virtual International Authority File, <http://viaf.org/>).

⁷ Examples of such ontologies include Records in Contexts (RiC-O, <https://www.ica.org/en/records-in-contexts-ric-a-standard-for-archival-description-presentation-congress-2016>), FOAF (Friend of a Friend, <http://xmlns.com/foaf/spec/>) and schema.org (<https://schema.org/>).

User type	User story
	<p>As a archives manager at a large institution, when I need to combine archival data from multiple business units I want to create a single portal site so researchers, archivists, and administrators can easily search across all of the collections held by my institution and view the search results in an integrated list.</p>
	<p>As a digital curator, when creating a virtual exhibit for the anniversary of a significant national event I would like to create a custom collection of archival descriptions and digital objects so I can exhibit the relevant materials from my archives.</p>
	<p>As a researcher, when I am searching for records related to my research topic I would like to search across multiple AtoM sites of various archives so that I can find relevant materials held by those archives.</p>
	<p>As an archives manager, when we publish our archival descriptions I want them to be found easily by members of the public searching via Google or using my parent institution's website, so I can fulfil my mandate to make my repository's archival holdings available to the public.</p>

How to meet the needs expressed in the user stories
<ol style="list-style-type: none"> 1. Demonstrate aggregation of content from multiple AtoM 3 sites into a single front-end 2. Demonstrate how querying data from one or more AtoM sites from a single location can be accomplished 3. Demonstrate how AtoM 3 will be capable of presenting content in IIIF viewers, which can be used to to assemble multi-page digital objects, provide pan and zoom functions, enable OCR text extraction, support annotation and tagging and assemble virtual exhibits 4. Demonstrate how a user can easily implement structured RDF in standardized schemas in order to produce better search results in major search engines

4.5 Tasks for the Proof of Concept

The PoC will consist of implementing AtoM 3 in a development environment, using the technical architecture described in section 5 below, in order to test the use cases. The environment will be limited to testing the use cases in a minimal way in order to demonstrate the feasibility of the proposed architecture. For example, in a real-world, large-scale production environment the AtoM 3 back-end would have numerous data entry templates and a variety of features designed to enhance the user experience; similarly, the AtoM 3 front-end would provide a feature-rich and user friendly experience to public users as they browse, search for and view content. The PoC environment, in contrast, will have very limited data entry templates and end-user browse features, including some that are html mockups or even wireframes instead of functioning forms, in order to demonstrate the functionality without actually implementing it.

The tasks required to successfully run a PoC include the following. Note that each use case builds on the functionality developed for the previous use case.

Use case 1: converting existing archival descriptions to Linked Data

- Provision infrastructure (servers and virtual machines) to host the technology stack
- Install software components
- Select an ontology⁸ for mapping AtoM 2 data to the selected ontology and perform the mapping
- Migrate sample AtoM 2 data to the PoC platform, using two or three selected AtoM 2 datasets
- Create a web interface or modify an existing web interface to query the Linked Data and provide hierarchical and non-hierarchical visualizations

Use case 2: enabling improved archival description

- Create a visual design to show what an AtoM 3 content management interface could look like
- Prepare wireframes for an AtoM 3 back-end to depict data entry templates that would take advantage of Linked Data
- Create a clickable prototype to allow testers to gain an understanding of the data entry experience in AtoM 3 based on the wireframes

⁸ A useful data model for the purpose of the PoC is Matterhorn RDF, developed by docuteam and the state archives of Wallis (Switzerland). Matterhorn RDF is based on existing ontologies with the goal to provide preservation and description metadata for digital holdings in archives.
<https://wiki.docuteam.ch/doku.php?id=docuteam:matterhornrdf>

Use case 3: federating and exposing data

- Create a visual design to show what an AtoM 3 public user interface could look like, including hierarchical and non-hierarchical archival description views
- Prepare wireframes for an AtoM 3 front-end
- Create a clickable prototype of the AtoM 3 front-end
- Set up two or three AtoM 3 back-end sites
- Configure the front-end AtoM 3 prototype to perform a federated search against the three AtoM 3 back-end sites and display the results
- Demonstrate that conducting searches in major search engines such as Google retrieve results from the AtoM 3 front-end prototype

4.6 Evaluating the Proof of Concept

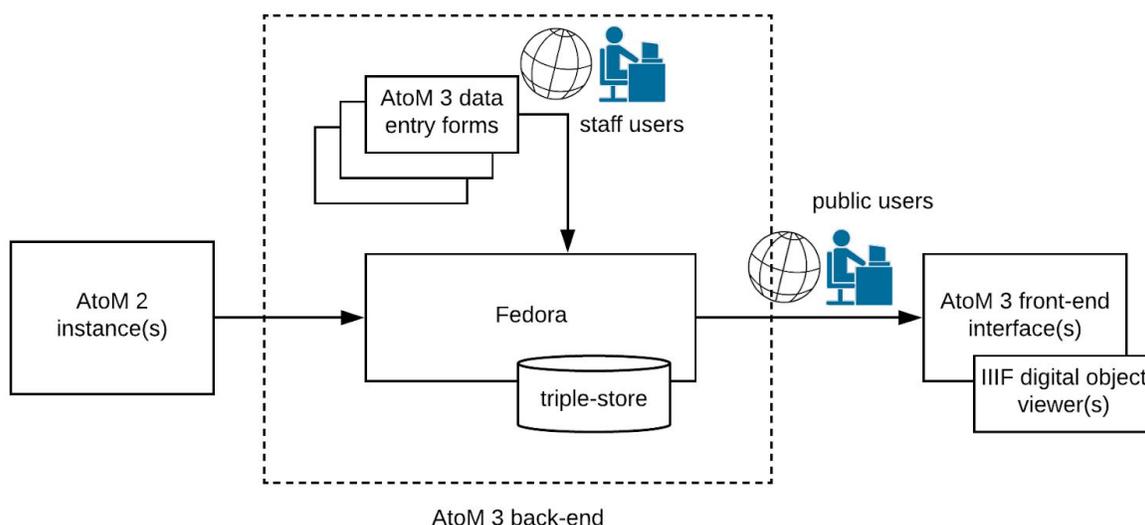
In keeping with best practices for open-source software development, this document and all analysis, documentation, software code and test results related to the PoC should be made freely available online, and should be well publicized in order to keep the archival community aware and informed at all stages of the process.

Iterative analysis, design, prototyping and testing is foundational to the development of a software application that fulfils its intended purpose and meets the needs of its designated users. Members of the community, particularly current AtoM 2 users, should be asked to comment on this proposal and all other aspects of the PoC, and should be recruited as testers. The design team will make hosted AtoM 3 sites available to the testers, will provide guidance on testing the software, will solicit and document feedback received from the testers, and will make adjustments to the PoC design and implementation in iterative development cycles to address the feedback. This approach is very similar to the software development approach for the original ICA-AtoM application, which benefited greatly from a dedicated pool of testers who provided constant feedback on new development beginning in the very early days of the project.

Feedback from testers should also be made available to the PoC funders independently of their interactions with the design team, in order to allow the funders to assess the extent to which the PoC meets user expectations for functionality and usability. If the feedback from testers is negative or uncertain, different approaches to the software design may be in order. This is the point of the PoC - to establish and test certain assumptions about how an archival description and online discovery platform should behave, and to test whether a given design and technology stack are capable of supporting the desired functionality. Proceeding to full development without a well-considered PoC and positive user response to the design and functionality would incur serious risk of failure.

5. Proposed technical architecture (high-level)

Because there are many open-source tools capable of supporting some of the proposed functionalities of AtoM 3, the PoC, and ultimately AtoM 3 itself, need not be built entirely from scratch. In the months before the meetings at King's College, the authors of this document spent considerable time reviewing existing open-source projects to determine what might be usable for AtoM 3, settling on Fedora⁹ as a good candidate upon which to build the platform. The diagram below shows a high-level system architecture which uses Fedora as the back-end repository platform to manage digital objects and link them to their descriptive metadata, which are held in a triple-store. Descriptive metadata (along with related digital objects) are either imported into Fedora from AtoM 2 databases and converted to RDF, or entered via web-based data entry templates directly by archivists and other staff users. For public users, the data are browsed and queried via front-end web-based interfaces using SPARQL query functionality¹⁰ which can display the content hierarchically or non-hierarchically, with digital object display enhanced by viewers based on the IIIF specification. A more detailed architecture diagram and description of the technology stack is available in **Appendix B**.



⁹ Fedora is an open-source digital repository system with native Linked Data support. The Fedora community is supported by the non-profit DuraSpace organization. <https://duraspace.org/fedora/>.

¹⁰ SPARQL is a query language designed to retrieve and manipulate RDF data.

6. Proof of Concept cost estimate

Estimated cost for accomplishing the tasks described in Section 4.5, above, is \$105,000 to \$131,250 CAD. This price is based on an estimate of 600 to 750 hours at a development fee of \$175 per hour.¹¹ The higher end of the range would allow for more detail and functionality (for example, more clickable wireframes and more working code), whereas the lower end is the minimum effort estimated to produce a basic working PoC.

7. Beyond the Proof of Concept: developing AtoM 3

The following is intended to illustrate a possible high-level development plan for AtoM 3, building on the Proof of Concept. The plan is necessarily inexact, and only the PoC phase is given a proposed duration; durations and implementation targets for the following phases will vary considerably based on available funding, community and stakeholder feedback, ongoing requirements analysis and knowledge gained through the development process.

Phase 1 - Proof of Concept

The goal of the PoC phase is to test that the proposed technical architecture and conceptual model are sound. The PoC software should implement an RDF data structure, partial import of data from AtoM 2 for testing, basic public search and discovery, very basic archival description editing, and SPARQL query functionality. Some functionality and user interface elements may be expressed through wireframes or mock-ups rather than functional software. The approximate duration of this phase is four to six months, and it is reasonable to expect to achieve about 5% of the intended functionality of AtoM 3.

Phase 2 - Prototype development

The goal of the Prototype is to integrate the lessons learned from the PoC phase and develop a minimum viable product implementation of AtoM 3. All code developed in the Prototype phase should be used in the final product - in other words, there should be no “throw-away” code. Functional software is the primary goal, and the software should allow a basic but complete workflow for archival description, management and publication, from import of legacy data and online data entry through to public search and discovery. This phase implements approximately 30% of the intended functionality of AtoM 3.

¹¹ \$175 CAD per hour is an average, not a specific price from a specific company. Different organizations from around the world may be involved in the development and exact hourly rates are not known.

Phase 3 - Beta development

The goals of the Beta development phase are to develop all of the major functionality expected in the final 3.0 Release and to enable the community and stakeholders to install and test the software. Functionality should include all features required for archival description and management, from import of legacy data and online data entry through to public search and discovery. Bugs are expected, and some features will require polishing to improve usability and functionality. By the end this phase, 95% of the intended functionality of AtoM 3 should be in place.

Phase 4 - Final development and production release

The Production phase is all about making the AtoM 3 software ready for production use. Extensive bug and usability testing of the software is required to find and address critical bugs and polish the functionality. The documentation likewise must be reviewed, corrected and expanded to maximize clarity and coverage. By the end of the Production phase, the software will include all of the features targeted for the AtoM 3.0 release, the documentation will be complete, and all critical bugs must be resolved. After this phase, the software will have 100% of its intended functionality and will be fully usable for large-scale and multi-repository implementations.

Phase 5 - Maintenance

After the production AtoM 3.0 release, ongoing community funding and engagement will continue to improve and expand the software's functionality and feature set. This ongoing phase includes software maintenance, bug fixes and documentation updates; continued community-funded development; and expansion of vendor support options and services. This is the current state of AtoM 2, a very successful project which in many ways serves as a model for the development of its successor.

APPENDIX A: THE CASE FOR LINKED DATA

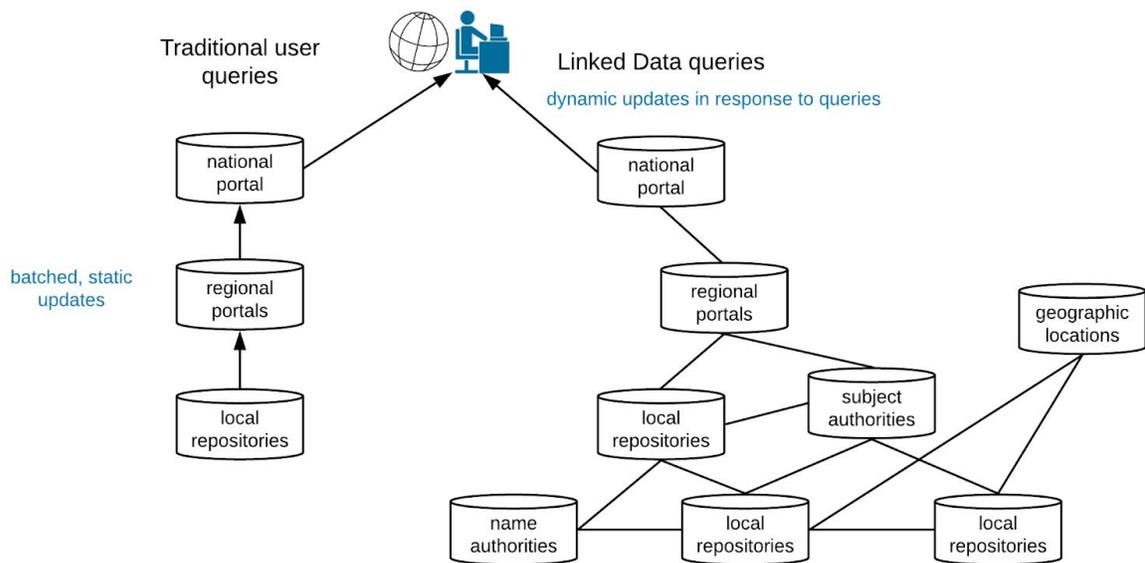
Better sharing of information

“Linked Data” is a way of structuring data so that disparate sources of information can be connected in ways that can be understood by computers. This is very different from embedding hypertext links in web pages, which requires the user to manually open the links in order to understand the relationship between the two resources. For example, in the Linked Data world, a body of records and a name authority can be linked in a meaningful way, so that a search engine can understand that the person or organization in the name authority has a specific relationship to the records, such as creator, accumulator, author, subject, custodian, etc. Linked Data also facilitates information sharing, so that, for example, multiple record sets held by multiple organizations can be linked to a central name authority. Linked Data also allows archival, bibliographic, citation, museum object, images and map data (among many types of resources) to be cross-referenced, bringing the heritage information management communities together. Useful resources for understanding Linked Data for the cultural heritage sector can be found at <http://publish.illinois.edu/linkedspcollections/outcomes/> and <http://lodlam.net/>. See also <http://www.canadiana.ca/pcdhn-lod> and <https://linkedjazz.org/>, two examples of Linked Data projects.

Easier aggregation and updates in multi-repository portal sites

Data aggregation in portal sites is better supported through the use of Linked Data than with traditional hierarchical description. Although AtoM 2 is used by many organizations to describe archival holdings and make them available online, the software has never fully met the need for archives to fully expose their content on the web. Aggregating data from multiple AtoM instances to portal sites such as Archeon and ArchivesCanada, for example, is a somewhat cumbersome process of importing updated data via CSV and EAD import and replacing existing content. The data that reside in multiple locations can easily become out of sync with the portal sites, and the problem is compounded when data are aggregated at multiple levels (for example, local sites contributing to a regional portal; regional portals contributing to a national portal; national portals contributing to international portals). This problem is due in part to the data models involved: updated descriptions must be exported and imported manually or via scheduled OAI harvests in what essentially amounts to recurring data migrations from local instances to data aggregators.

Linked Data, on the other hand, is designed for dynamic querying across multiple information sources, and for assembling results into web interfaces. The differences between using Linked Data and traditional archival description for aggregating data into portal sites is depicted in the diagram below.



More context, more meaning, more nuance

Linked Data is also better at representing complex relationships and contextual information in archival description. Traditional archival description emphasizes ownership and custody of records and the primacy of a single creator for an aggregation of records. However, the real world rarely accords so cleanly with such a model, especially in the digital age. Contemporary archival theorists have increasingly argued that records originate from dynamic interactions between diverse individuals and communities, and that attributing provenance of a fonds, series, file or other unit of description to a single source is often overly simplistic. Descriptive standards such as ISAD(G), ISAAR and RAD (Canadian Rules for Archival Description) prescribe arrangement of aggregations of records into hierarchies, with creators attached to the highest possible level of description and inherited by lower levels. Linked Data schemas, on the other hand, make it possible to capture rich relationships between individuals and communities that intersect with records as creators, custodians and subjects. Rather than a single dominant narrative, a Linked Data approach also allows for a multiplicity of narratives to be captured as statements about a resource, including a multi-provenancial approach to description and arrangement.

In AtoM 2, hierarchical arrangement and limited relationships between agents and records are enforced by the underlying relational database model on which the software is built. In a Linked Data platform, much more expressive relationships are made possible by the use of ontologies designed specifically for that purpose. RiC, for example, contains many possibilities for expressing the complex interactions between individuals, organizations, functions and record sets.

Better support for standardization

Greater standardization in descriptive practices is also more possible with Linked Data, which is designed to allow for linkages with name authorities, controlled vocabularies and other external sources of information. When a data value in a Linked Data platform is a link instead of text, the relationship between the description and the linked external resource can be made meaningful and the linked resource can be displayed together with the data to which it has been linked. For example, an authority record for a school can include a link to an authority record for an individual, and the link can specify whether the individual attended the school or was a teacher or administrator, etc. Other links between the school and individuals can allow search engines to infer relationships between the individuals, and visualization tools consuming the Linked Data can then present those relationships in ways that are not possible if the archivist or researcher is simply following links embedded in free text descriptions. Many organizations can make use of the same name authorities, rather than laboriously writing their own administrative and biographical histories and using those histories strictly for their own archival descriptions.

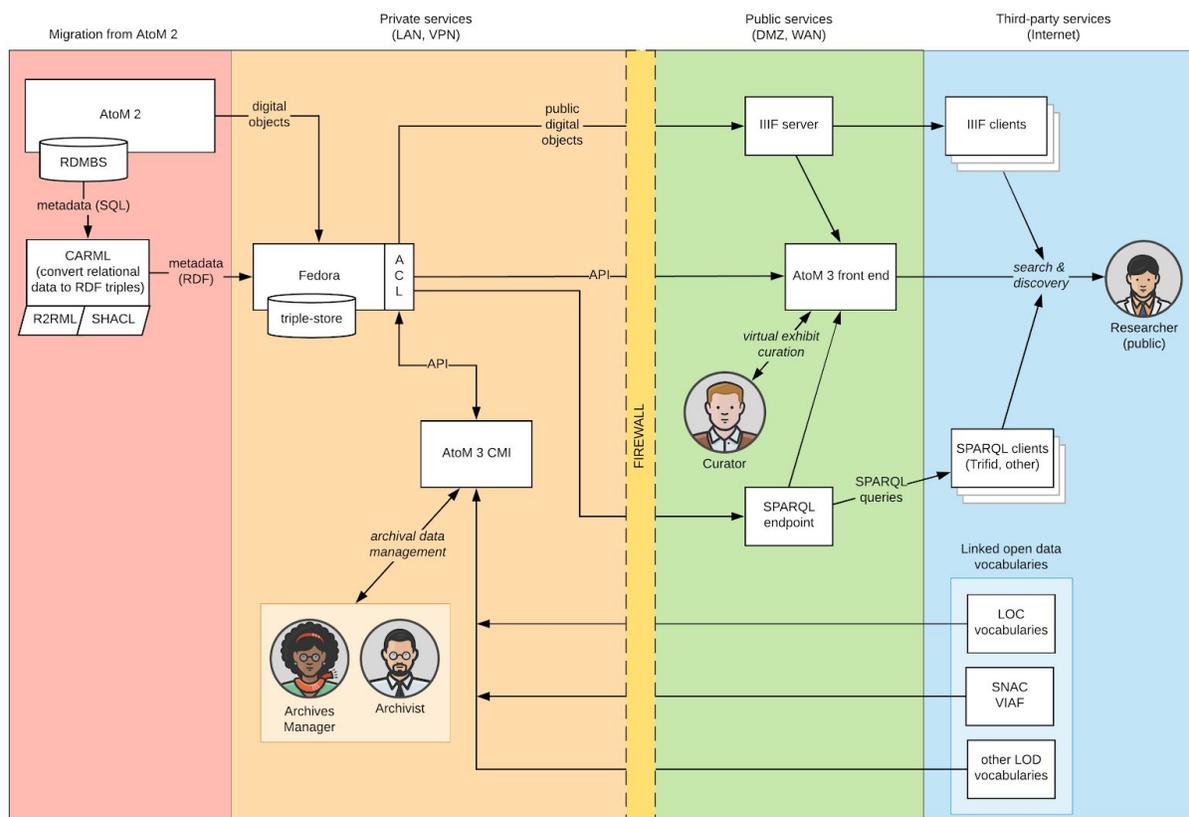
Meeting the need with AtoM 3

The increasing availability of Linked Data services from cultural heritage institutions allows for such linkages to be made only when applications that can consume this data exist. Name authorities, controlled vocabularies and even archival ontologies and data models are rapidly becoming more common and more readily available, but there are few, if any, software applications with a specifically archival focus that can ingest and reuse this data, or contribute back to the Linked Data ecosystem. Linked Data-based applications such as Islandora and Samvera, and data aggregators such as Digital Public Library of America and Europeana, are designed for libraries, which focus their efforts on publications and collections rather than naturally accumulating bodies of records with complex contexts of creation and use.

With the International Council on Archives' shift towards a single, unified archival standard rooted in Linked Data and other changes in the archival description landscape, there is a great need for an open-source, freely available archival description application that can author and expose Linked Data and meet the requirements of small and large institutions alike. AtoM 3 will be designed to meet this need.

APPENDIX B: PROPOSED TECHNICAL ARCHITECTURE (DETAILED)

The architecture diagram and associated narrative below describe the components of the PoC and how they would be used to accomplish the tasks described in the use cases and user stories.



In this diagram:

- An RDF Mapping Language (R2RML)¹² engine such as CARML¹³ converts AtoM 2 MySQL data to RDF triples, and loads the RDF triples to an AtoM 3 back-end based on

¹² R2RML is “a language for expressing customized mappings from relational databases to RDF datasets”. <https://www.w3.org/TR/r2rml/>.

¹³ “CARML - A pretty sweet RML engine” <https://github.com/carmil/carmil>

Fedora¹⁴. The Shape Constraint Language (SHACL)¹⁵ is used to validate the structure of the RDF triples.

- A script imports AtoM 2 digital objects from the AtoM 2 filesystem to the Fedora data store and associates the imported digital objects with the correct RDF metadata.
- Fedora provides metadata storage and management, digital object storage, access control (via WebAC), HTTP REST APIs, resource versioning (via Momento), fixity checking and other repository functionality.
- The AtoM 3 content management interface (CMI) provides data entry templates used to enter metadata directly, and writes the data to Fedora via HTTP REST API. Although the data are captured in Fedora as RDF triples, the data entry templates look similar to those used in AtoM 2.
- When descriptions are being created and edited, the AtoM 3 CMI allows creating qualified relationships using established Linked Data vocabularies (e.g. LCSH, SNAC, VIAF) to enrich the archival metadata and provide additional context.
- Access Control Lists (ACLs)¹⁶ are used to manage group and user permissions for data and digital objects exposed by the Fedora REST APIs.
- The user can upload digital objects to AtoM 3 via the CMI, as in AtoM 2.
- An IIIF¹⁷ server integrates with the Fedora REST API to serve digital object metadata and content to IIIF clients. The digital objects available via the IIIF server are restricted to public records by the Fedora ACL.
- A SPARQL endpoint enables third-party SPARQL clients such as Trifid¹⁸ to run powerful, ad-hoc searches on the AtoM 3 RDF data to facilitate novel queries and aggregations of the data. The content available via the SPARQL endpoint is restricted to public records by the Fedora ACL.
- The AtoM front-end provides public search and discovery of archival data. A user friendly and familiar search interface will query against the SPARQL endpoint to return and display standard search results.
- Detailed archival data will be provided to the AtoM 3 front-end through the Fedora REST API.
- Using SPARQL queries one AtoM front-end can query and aggregate results from multiple AtoM 3 instances.
- Virtual collections can be curated in the AtoM front-end by combining resources from one or more AtoM instances.

¹⁴ Fedora is an open-source repository platform with native Linked Data support.

<https://duraspace.org/fedora/>.

¹⁵ SHACL is a validation language designed specifically for RDF. <https://www.w3.org/TR/shacl/>.

¹⁶ ACL is widely used for managing user groups and permissions, and is used in AtoM 2.

https://en.wikipedia.org/wiki/Access_control_list.

¹⁷ IIIF, or International Image Interoperability Framework, is a well-supported international standard designed to support the rendering of digital objects in viewers. <https://iiif.io/>.

¹⁸ Trifid is an open-source tool developed by Zazuko. <https://github.com/zazuko/trifid>.