The ICA-AtoM Project and Technology

Peter Van Garderen
Software Release Manager, ICA-AtoM Project
President/Senior Consultant, Artefactual Systems Inc.

Association of Brazilian Archivists, Third Meeting on Archival Information Databases
Rio de Janiero, Brazil. 16/17 March 2009.

INTRODUCTION 1
FREE AND OPEN SOURCE SOFTWARE 3
SOFTWARE FEATURES 6
BETA TESTING 11
PROJECT GOVERNANCE 18
ICA-ATOM AND THE ICA STRATEGIC OBJECTIVES 21
TECHNICAL ARCHITECTURE 23
DATA MODEL 32
CLOSING REMARKS 36

Introduction

Bom dia. I want to thank the Association of Brazilian Archivists for this opportunity to talk about the ICA-AtoM project. I am humbled by the opportunity to address this conference twice this week. I gave another presentation at your annual conference in June so I would like to think that you just love listening to me talk. However, I realize that this invitation is more likely an indication of the growing interest in open-source software within the Brazilian archival community. Before I begin, I would like to pass on greetings and best wishes from the ICA Secretary-General David Leitch. The ICA is very excited that our Brazilian colleagues have taken an early lead role in helping to realize the vision and full potential of the ICA-AtoM software collaboration.

Of course, I myself do not represent the ICA directly. My Canadian company, Artefactual Systems, is the primary contractor that is leading the technical development of the

---

software. My formal role within the project is Software Release Manager and, as such, I work closely with the ICA Secretariat and the ICA-AtoM Steering Committee. However, I should note that I am an archivist myself and that the roots of this project go back to my own personal motivation to provide a free and open-source software application for use by my fellow archivists.

After I graduated from the Master of Archival Studies program at the University of British Columbia in 1997, I worked briefly for a commercial software vendor. There I learned pretty quickly that most archival institutions have very limited resources. I took a call from a sweet elderly lady one morning who was a volunteer at a small community archives. I had spoken to her previously about using our software to manage her archival description project. She told me excitedly that they had a very successful Bingo night and that the archives now had $500 dollars to purchase the software. Unfortunately I had to tell her that the software cost several times that amount. "Oh," she said sadly. "I guess we need a few more bingos." Another time I was on-site with an existing client who had some difficulty getting the software to work. After taking a look at their system I realized that their problem would be solved if they simply implemented an additional module. Unfortunately, that module cost several thousands of dollars. "We've already missed the budget for this year," the archivist noted. "Maybe, we can get the money in the next fiscal year." They were forced to continue using word processing software to write and print out finding aids and eventually migrated their data to another system. These early episodes made quite an impression on me. I had the technical knowledge to help my fellow archivists but that was only if they had the money to pay for access to the required tools.

I soon moved on to another job as the first Project Coordinator for the InterPARES Project back at UBC. Then, in 2001, I launched Artefactual Systems to begin my electronic records and digital preservation consulting practice. Throughout that time I became increasingly interested in the free and open-source software movement, which was characterized by the growing mainstream popularity of the Linux operating system and the Apache web server as well as a number of digital library projects such as Dspace.

Free and Open Source Software

The free software movement began when Richard Stallman released the GNU operating systems in 1983 as a replacement for the UNIX operating system. GNU stands for “GNU is Not UNIX”. GNU still forms the basis of the current Linux operating system. Stallman was frustrated by the restrictions placed on his ability as a computer scientist to study and share the design of software systems. Therefore, he released GNU under his own GNU Public License (GPL) which gave users of the software four basic freedoms:

1. The freedom to run the program for any purpose
2. The freedom to study how the program works, and adapt it to their own needs, meaning that easy access to the source code must be provided
3. The freedom to redistribute copies to help friends, family, colleagues or society in general
4. The freedom to improve the program, and release their own improvements to the public, so that the whole community benefits. Again, easy access to the source code is a precondition for this.

The GPL is now the most widely implemented free and open-source license in use. The ICA-AtoM software is released under version 2 of the GPL. There are also a number of other open-source licenses available. The Open Source Initiative maintains a full list at http://osi.org as well as an open source definition in the form of ten criteria that expand upon Stallman's four freedoms.

It is important to note that these freedoms or criteria do not restrict the ability to charge money for free and open-source software. That said, very few open-source projects actually charge money for their software. Providing software free of charge is certainly one of the more popular characteristics of most open-source projects, and it is what first attracted me and many others. Therefore, there are no fees at all to download, install or revise the ICA-AtoM software. It is completely free of charge. It is like free beer. If I buy you a beer, you can drink it and that costs you nothing. This makes me feel good about myself and it is pretty exciting for you, assuming you like beer. Using the freedom to redistribute free copies, I am finally able to pass on my technical knowledge in the form of software, even to those who could not otherwise afford my consulting services.

However, as exciting as free beer is, the ability for you to study, improve and re-distribute the software is arguably more exciting and was certainly revolutionary when the idea was first proposed by Stallman. This is freedom as in 'free as a dove'. If you choose to use ICA-AtoM or any other open-source software you have the ability to make modifications and add modules as you see fit. You don't have to ask anyone's permission. If you don't have the technical ability to make the modifications you need, you can get the help from a colleague or friend who does or hire anyone with the necessary technical skills at the most attractive price. Better yet, you can collaborate with other users of the same software to pool financial and technical resources to improve the software and redistribute those improvements back to all users of the software. This is the approach taken by the ICA-AtoM project.

Of course, free and open-source software comes with its own set of responsibilities. So in addition to being like 'free beer' and 'free as a dove', open-source software is also like a 'free kitten'. They are cute, cuddly and exciting but they require feeding, a warm place to sleep and you have to let them out every now and then to go to the bathroom. In the end, you build a two-way relationship with the cat which usually rewards both parties with friendship and trust. Similarly, even if the open-source software is free of charge it is not without secondary costs or responsibilities. There are always costs associated with installing and maintaining software systems, although I would argue that the total cost of ownership, which factors in all these costs, is often significantly lower for free and open-source software. Commercial vendors whose business models are threatened by the open-source movement are quick to spread fear, uncertainty and doubt about total cost of ownership estimates. I think each organization needs to do its own cost-benefit analysis before making a choice between proprietary or free, open-source software. UNESCO recently prepared an excellent report to assist with such an analysis.[1]

One good example from my home province of British Columbia in Canada is the provincial Public Libraries Association. After encountering questionable upgrade and pricing practices from some major library software vendors, the association decided to support the migration of all the public library information systems to the open-source Evergreen ILS application[2]. They began by hiring one project manager, system

[1] "Open source and proprietary software"(Sept. 2007) UNESCO Information for All Programme.
[2] http://sitka.bclibraries.ca/

administrator and end-user support staff person. The cost of staff salaries and centrally hosting this free software is still significantly less than each library paying for its own license and technical support from the vendors. Opting in to the central Evergreen service is voluntary but about 80% of the public libraries in the province will likely be using it by the end of 2009. The project manager estimates that this switch will save over $10 million dollars in total costs over the next few years.[3]

Similarly, here in Brazil the federal government has been actively switching to the use of open source for cost reasons. Sergio Amadeu, who runs the government's National Institute for Information Technology, says that "the number one reason for this change is economic." He explains that, for every workstation, the government is currently paying Microsoft fees of around 1200 Brazilian reais (approximately $500 USD). "If you switch to open source software, you pay less in royalties to foreign companies," says Amadeu. "And that can count for a lot in a country like Brazil, which still has a long way to develop in the IT sector." Overall, Amadeu estimates that the government could save around $120 million a year by switching from Windows to open-source alternatives.[4]

Amadeu's last point about using open source as a way to stimulate an ICT sector in Brazil is also interesting. This comes back to the responsibility for taking care of a free kitten. I believe that by getting involved in open-source projects, archival institutions will gain more control and knowledge of the technical infrastructure which they require to manage archival functions. Some would argue that this is not a core function for archives and I would agree to some extent. However, archivists are responsible for preserving and providing access to information. Today, over 90% of the information being created in the world is in digital format. These are tomorrow's archives. In fact, many archives are already struggling with transfers and accruals of electronic records. Archivists can no longer ignore digital technologies simply because they don't fully understand them or because they find the topic overwhelming. Mechanics need to understand automobile engines, doctors need to understand human bodies and archivists need to understand information in its digital form as well as the systems that are used to create, manage and provide access to them. I believe that the focus on open sharing of technical knowledge

[3] Ben Hyman, Manager Policy & Technology. BC Public Library Services Branch, Access2007 Library Technology Conference, October 12 2007, Victoria BC

[4] 'Brazil adopts open-source software' (2 June 2005) BBC News. http://news.bbc.co.uk/2/hi/business/4602325.stm..

and know-how, which is characteristic of healthy open-source communities, will go a long way toward raising the technical capacity of the archival profession.

I also think that the open-source community's willingness to talk about its own faults and shortcomings is a healthy part of that. Archivists should not be embarrassed to ask the 'wrong question' and developers and technical staff should not be embarrassed to admit to bugs, feature gaps or technical mistakes. These exist in all software projects. The difference is that open-source projects such as ICA-AtoM make this information freely available by providing access to the source code repository, online bug list, developer's wiki and discussion list. At the same time we try to actively involve users in working towards solutions. Commercial vendors tend to hide issues or shortcomings and usually find themselves trapped in marketing speak where they exaggerate both the capabilities of their own products as well as the flaws of their competitors. That does not mean that there are not decent commercial vendors providing good products and services to their clients; however, their software license costs are typically out of the price range of most archival institutions. Furthermore, their proprietary technology tends to create a relationship of dependence rather than one which builds sustainable knowledge, capacity and technological autonomy within the archival community.

Software Features

So now that I've explained some of the philosophical foundation for the ICA-AtoM project as well as my own personal motivation to develop the software, I'd like to give a demonstration of the features found in the current 1.0.5 beta version which was released last week on March 11.[5]

Even if you choose not to download and install ICA-AtoM, you are still able to test these features for yourself using the online demo version of the application or by burning a copy of the Demo CD. The online demo is a fully-featured copy of the application which grants you full administrator privileges so that you can experiment with all the application's capabilities. The demo website refreshes every hour with default data so you are free to make any changes you like.

[5] See http://ica-atom.org/docs/index.php?title=User_manual and http://ica-atom.org/demo.html

Another option is the Demo CD. ICA-AtoM is web-based software which requires a web server and database server to operate. However, we created the ICA-AtoM demo CD to make it easy for you to try out the ICA-AtoM software on your own local computer. The Demo CD will run on any computer. It temporarily loads the Ubuntu Linux operating system into memory, along with the necessary web server, database server, Firefox browser and ICA-AtoM application. When you are finished the demo, your computer restarts using your regular operating system and configuration. Of course you can download and make as many copies as you like and I've also brought some Demo CD copies with me to Rio de Janerio.

Researchers are able to search the archival descriptions hosted by ICA-AtoM using a basic search box. Advanced users are able to use the same search box to enter more sophisticated boolean or proximity search criteria. ICA-AtoM uses the Zend Lucene search engine and ranks search results based on where the search term appears in the record. For example, hits in the title, creator and access point fields are ranked higher than the archival history field. These criteria can be edited by developers and we would like to add the ability, possibly by release 1.2, for administrators to configure the algorithm ranking themselves.

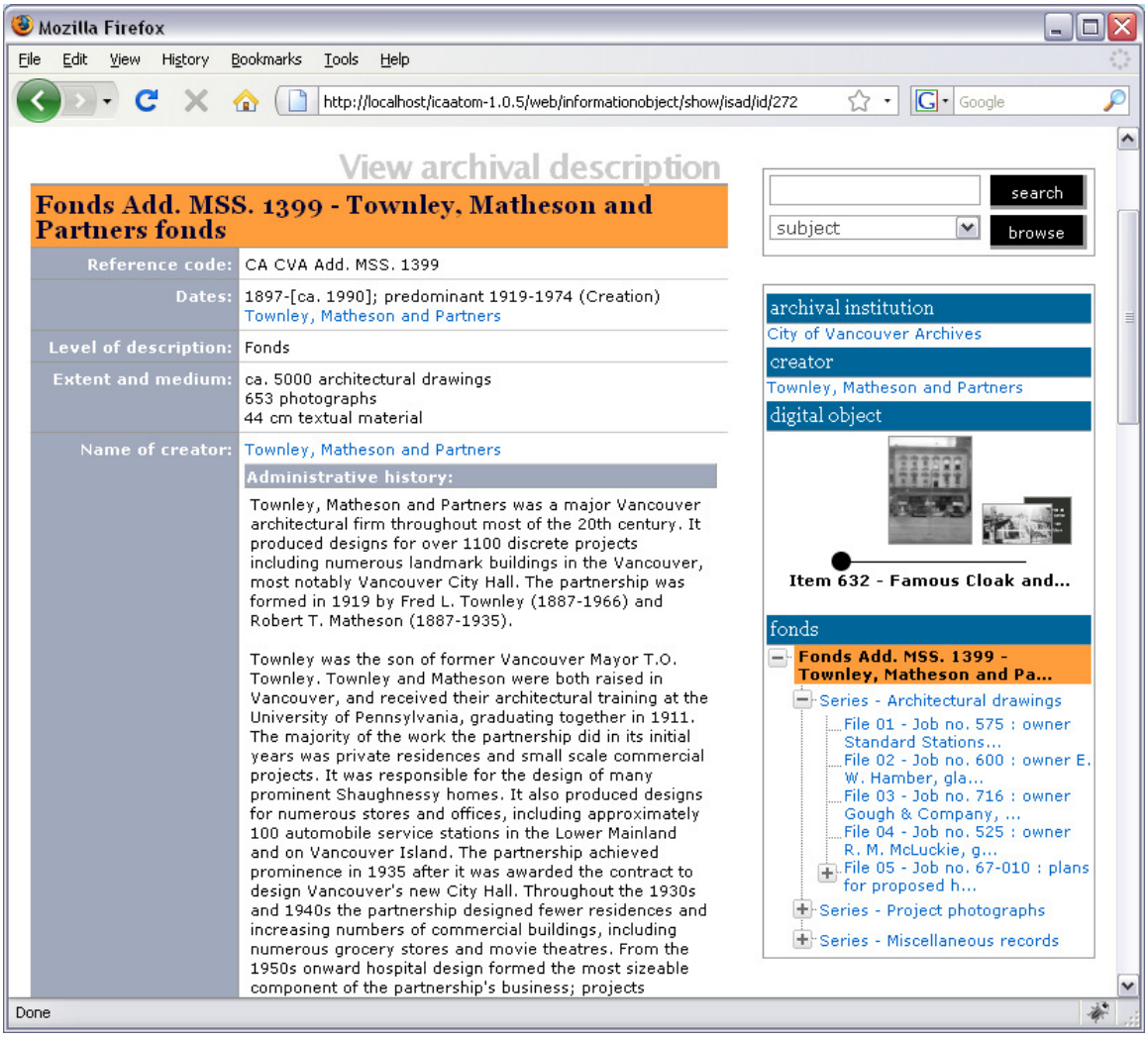
Users can navigate from the search results to the full archival descriptions which are shown in context to their multi-level description, as well as links to the creator's authority record. The user has the ability to browse by facets such as subject, place, names and media types. The user can also view any links to digital objects or browse all the digital objects for a particular aggregate level of description using a Coverflow viewer. ICA-AtoM creates access derivatives for uploaded digital objects, e.g. JPG images, Flash video. The application also provides a browser for multi-page image and text documents.

Users with log-in permission are able to add and edit archival descriptions, authority records or repository profiles. These are all compliant with the ICA's descriptive standards, namely:

- *International Standard Archival Description* (ISAD(G)) - 2nd edition, 1999.
- *International Standard Archival Authority Record (Corporate bodies, Persons, Families)* (ISAAR(CPF)) - 2nd edition, 2003.

- *International Standard For Describing Institutions with Archival Holdings* (ISDIAH) - 1st edition, March 2008.

The 1.1 release will also add support for the International Standard For Describing Functions (ISDF) - 1st edition, May 2007. The current release also contains data-entry templates for Dublin Core and the Canadian Rules for Archival Description



Names of creators and other actors can be linked from the archival description to an authority record via an Event entity that also records the dates and other information related to the event. Archival descriptions can be linked to access points and, if the user has permission, they can add additional terms to the controlled vocabularies.

All the data-entry sections correspond to the areas of description for each standard. As well, we have access point, digital object and physical object areas. The physical object area allows for links to the boxes and containers in which the analogue archival materials are stored.

All the terms that are available as access points and menu options throughout the application are maintained as controlled vocabulary taxonomies. By release 1.1 we will make these fully compliant with the ISO Thesauri standard relationships (e.g. Use, Use for, Broad Term, Narrow Term, See also)

All user interface elements (e.g. field labels) as well as database content (e.g. archival descriptions, authority records, static pages, etc.) can be translated into multiple languages. The current version of ICA-AtoM contains translations for Dutch, English, Farsi, French, Italian, Portuguese, Slovenian, and Spanish.

Users are able to export archival descriptions using the EAD XML format and they can also import EAD documents, including any multi-level description hierarchies and physical container elements. The 1.0.5 release includes the ability for ICA-AtoM to act as an OAI repository, making descriptions available to OAI harvesters. The 1.0.6 release will include the ability for ICA-AtoM to harvest and import OAI records from other repositories. We are co-developing the OAI feature with the Library and Archives of Canada which is interested in seeing this functionality in ICA-AtoM to enhance its capability to act as a multi-repository portal system which can receive data from contributors via direct data-entry, EAD XML import or OAI harvesting.

The ability to configure OAI, multi-repository and other settings such as interface languages or user accounts is provided through a basic Administrative interface. Administrators are also able to customize their site titles, static page content (e.g. homepage, contact page) and application menus. The 1.0.6 release will include a theming feature which will allow administrators to change the look and feel of the application in a single click as well as develop their own institutional themes.

As I will explain later in the technical architecture overview, ICA-AtoM is fully web-based software. This type of application is typically more complicated to install than a stand-

alone, desktop application. Therefore, we have included a web-based installer with the application to simplify this task. It performs a full system check to determine that the web server and support files comply with the application's minimal technical requirements. If not, it provide a report with explanations on how to re-configure the server environment accordingly. Thus far, about 80% of the installations we have performed are fully handled by the web installer with the remaining 20% requiring some manual configuration intervention.

I have just demonstrated the features that are available in the current 1.0.5 release of the software. 'Release early and release often' is one of the credos of the open-source and web application community. This is something we take to heart as well. I think a realistic expectation for any software application is that it should always be in a process of getting fixed, enhanced and improved. In late April we will release version 1.0.6 which is mostly a maintenance release but which will also include MODS templates and a theming/skinning module as two major new features.

The next milestone will then be release 1.1 which we will formally launch at the ICA CITRA meeting in Malta in November 2009. This will be the stable, production-ready version of the application. It will include a number of performance and workflow upgrades for the XML import and export. as well as full compliance for the relationship types required by the ICA ISAAR and ISO Thesauri standards. Release 1.1 will also provide support for the ICA's ISDF standard as well as EAC XML templates for importing and exporting authority records. One of the more important new features will be a full Access Control List (ACL) capability. This will enable system administrators to define complex permission rules for the system users (e.g. the ability to restrict edit privileges per taxonomy, or translation privileges per language, or edit privileges per repository).

The ICA-AtoM steering committee will hold another meeting in Malta and decide on the next steps for the software based on the community feedback and how the project's funding and governance model have been defined by that time. However, it is likely that the 1.2 release will address upgrades to the search module as well as adding an accessioning module.

Beta Testing

The beta 1.0 release of the software was successfully launched at the ICA Congress in July 2008 with a conference presentation, two end-user workshops and one administrator workshop. Over 1200 Demo CDs were distributed to delegates and 26 institutions were recruited to participate in a formal round of beta testing. This began in November 2008 with the 1.0.3 release of the software. A full list of the participants, including links to their sites, is available on the beta testing wiki page.[6] There is wide international representation within this group. Beta testing sites have been deployed in English, French, Spanish, Portuguese, Arabic, Farsi, Italian, Slovenian, Dutch and German.

The beta testing includes active dialogue between the users and developers which is all recorded on the public ICA-AtoM discussion list. [7] Bug fixes and new feature requests are being incorporated into the daily development schedule as part of this work. This first round of beta-testing will end later this month with a survey of the participants to ask for their feedback, impressions, and advice on the application's features. This survey and its replies will be posted to the ICA-AtoM user discussion list. We will launch a second round of beta-testing from May to August, based on the 1.0.6 release of the software. This will assist with finding and resolving any bugs prior to the 1.1 release and will give us further guidance on developing the application for the 1.2 release and beyond.

Much of the beta tester feedback thus far relates to the underlying data model and the implementation of standards. Examples include requests to add a template for the Canadian Rules for Archival Description and EAD XML import/export functionalities; both of these important features have since been added to the software. The Australian testers have requested that ICA-AtoM be adapted for use with the Australian series system. An early analysis demonstrated that the ICA-AtoM data model is flexible enough to accommodate this, including the flexible multi-level hierarchy and multi-provenance links that this requires.[8] Further analysis is now continuing to determine whether a separate template is required or whether small changes to the existing template will be sufficient to accommodate any extra metadata attributes. I should also note that a much

[6] http://www.ica-atom.org/docs/index.php?title=Beta_Testing
[7] http://groups.google.ca/group/ica-atom-users
[8] see http://groups.google.ca/group/ica-atom-users/browse_thread/thread/101426fb8444434f/

earlier analysis carried out in the summer of 2008 at the National Archives of Brazil revealed that ICA-AtoM was compliant with the Brazilian Nobrade standard. Since it is so closely modeled on the ISAD(G), the only change that was required was to add a 'conservation note'. This was done as early as the 1.0.1 release. Finally, the Direction des Archives de France conducted a thorough review of the 1.0.4 release for ICA-AtoM compliance with the full set of ICA standards. We are currently implementing the majority of the recommendations from the this report in the 1.0.6 release. All of the DAF report recommendations, including the ability to use the new ICA International Standard for Describing Functions, will be available in the 1.1 release in November 2009.

Some of the more technical feedback from our beta tester include requests for automated test scripts and a built-in data migration tool, both of which have now been implemented. However, the majority of feedback has been related to usability. Some of the testers have entered large numbers of descriptions, which allows them to become aware of usability considerations that are not necessarily apparent when the system contains only a small sampling of records. For example, the National Archives of Iran imported several tens of thousands of existing descriptions and authority records when they began testing ICA-AtoM. Linking to existing authority records is cumbersome when you have to browse a very, very large list. One way to deal with this is to add an AJAX auto-complete search feature. The developers at the National Archives of Iran coordinated the development of this feature with us and have released it under a GPL license so that we can integrate their work into the upcoming 1.6 or 1.1 release.

Some of the other usability requests have included improvements to adding new users, expandable fields for data entry, and alphabetical sorting in drop-down menus; these features have all been added to the software. Others that are not yet implemented, but are on the development roadmap, include bulk processing capabilities as well as the ability to work with default record templates to avoid having to add the same data repeatedly for large numbers of similar descriptions.

Many beta-tester questions relate to how to use existing functionalities. The most common of these are how to attach creator information to an archival description, whether to add a creator for all levels of description and whether to enter an archival institution for all levels of description. In ICA-AtoM, creator information such as

administrative history and dates of existence are maintained in separate authority records and are “added” to the archival description by linking to an existing ISAAR authority record or creating a new one. Both the creator and archival institution are inherited at the lower levels of description, which means that they should not be linked to those descriptions. This practice follows the principles of 'description of the general to the specific' and 'non-repetition of information', both of which are stated in the ISAD(G) and other archival standards. However, somewhat surprisingly, a number of archivists seem unfamiliar with these principles and therefore experience some confusion about how they are implemented in the software. We hope to add a validation feature in the 1.1 release software which gives the user some more explicit feedback on the screen if they are violating some of the data-entry rules or principles.

It seems pertinent at this point to add some other observations about beta-testing behaviour. In general, we have found that users prefer asking questions over reading the manual and forcing us to re-evaluate how much resources we devote to maintaining user documentation. Certainly, context-sensitive help should help fill in some of this gap. We are hoping to include this feature in the 1.2 release.

Moreover, the beta-testers are often reluctant to post questions to the users’ discussion list, preferring instead to rely on one-to-one e-mail communications with software support staff and developers. Possibly, archivists are afraid of asking the “wrong” question and making themselves and their institutions look bad. However, one-on-one communications are not sustainable in the long-term and result in information not being shared with the user community at large. It may take a little longer than expected to encourage archivists to embrace a culture of open collaboration and communication, in the meanwhile we will continue to try and lead by example. Wherever, possible we simply forward communication that does not include sensitive information to the discussion list and carry on the response thread there.

Other observations include the fact that the testers often have limited time to devote to the testing, although this was expected, given what we know about the resources available to most archival institutions today. Therefore, we have never asked for more than a few hours of beta-testers' time. Some have been able to provide a lot more than

that but we appreciate any contribution that is made and all of them are helping to improve the software.

Artefactual Systems is hosting most of the beta-testing participants sites (about 85%). This is partly due to the fact that those archival institutions that have internal technical support cannot spare it on a beta-testing project, they are too busy with day-to-day production responsibilities. However, it also a reflection of a larger technical support gap in the archival community. We anticipate that there will be a need to provide third-party hosting services for ICA-AtoM. Therefore, Artefactual Systems will be providing that for a minimal fee after the beta-testing period. Our first hosting client will be the Archives Association of British Columbia which is currently beta-testing the application and planning to migrate its provincial union list of archival descriptions to ICA-AtoM in a couple of months. We are able to provide hosting, backup and technical support at a cost of $125 a month which is a fraction of their current fees and less than the current hosting costs for all other commercial archives management applications. We will only charge a monthly fee per website and will not limit the number of users. This means the AABC can give out passwords to those institutions that do not have their own archival description software or do not have the capability to contribute to the provincial union list using EAD export or OAI harvesting. Each archival institution can thereby enter or update their archival descriptions directly in the portal, eliminating the current, cumbersome data migration tasks which are usually carried out by cutting and pasting from word processing documents.

I would also encourage the approach taken by the British Columbia Public Libraries Association where institutions centralize their technical support amongst themselves and share the costs. Here in Rio de Janiero, the Casa Rui Barbosa Foundation has partnered with the Catholic University. The University Lab is providing hosting support for their beta-testing site while their technical staff are taking the opportunity to study and enhance ICA-AtoM in relation to their own digital library system. This partnership could form the basis for an ongoing hosting arrangement that could be extended to other Brazilian institutions. Of course, that is only one suggestion. I don't want to put unnecessary expectations on either of these excellent beta-testing partners.

My final observation on the beta-testing period is that some testers have unrealistic expectations about the application. To be clear, the current version is a beta release not a stable production version. We are conducting a beta-testing phase to find bugs and get the feedback and recommendations from archivists and their institutions on how to improve the application in time for the 1.1 release.

Also, many users have trouble seeing past the vendor/client type relationship they are used to and appear to have trouble seeing themselves as being part of a broader community that has control over its own software. Yes, Artefactual Systems are the lead developers and the primary source of technical support right now. However, in this open-source eco-system, under the freedoms granted by the GPL license, we are replaceable and compatible with any number of other sources of technical support and software development.

I should note that there are some encouraging exceptions. The Library and Archives of Canada as well as the National Library and Archives of Iran has been very active in testing the full capabilities of the software and contributing to its development, customizing the software to suit their own needs where necessary. As well, developers here at the Catholic University in Rio are looking closely at how they can leverage some of their experience with designing multi-lingual interfaces for their digital library application to enhance the same features in ICA-AtoM.

At the same time, I'll continue to do the best I can to convince archivists that this is their software. This will require more education and acceptance of the open-source model and culture within our community. Therefore, I really appreciate the time I have been given by the AAB to explain the history and vision of the ICA-AtoM Project. In the meanwhile, the ICA and Artefactual Systems are dedicated to providing leadership and support but, in the end, this is your free kitten.

Project History

So now that you have a better idea of the ICA-AtoM's features and objectives, I'd like to take a step back and tell the story of how we arrived at this point. In 2003 the ICA Committee on Information Technology, under the direction of Peter Horsman from the

Netherlands Institute for Archival Education Research (Archiefschool), published the functional requirements for an Open Source Archival Resource Information System (OSARIS). It was the Committee's intention to use these requirements as the basis for developing an open-source archival description application; however, it did not have the funding to carry this out.

Then, in 2005, the UNESCO Information For All Programme (IFAP) granted €45,000 Euros to the ICA's Human Rights Task Force to create an online guide to archival sources related to human rights violations. UNESCO's IFAP actively promotes the use and creation of open-source software and that was one of the criteria for the online guide. Peter Horsman realized that this was a good opportunity to follow up on the OSARIS project. He pointed out that the software created for the online Human Rights database should be based on ICA descriptive standards. The UNESCO grant could be used to build the online database but also double as seed funding to create an open-source archival description application for use by ICA members and other memory institutions. The ICA Secretariat agreed and asked the Dutch Archiefschool to coordinate this work. At that time Peter Horsman and I were working together on the requirements analysis for an E-Depot at the City of Rotterdam archives. He knew I was keen to develop an open-source application for archives since we had discussed such a project several times in the past. Therefore, I eagerly agreed when he asked whether I would be interested in working on the UNESCO contract for the ICA.

During early 2006 I undertook an extensive technical analysis, which I'll revisit in my presentation tomorrow, and developed a first prototype release of the software. We held a project meeting at the ICA Secretariat in mid-2006 to review the prototype and to discuss the vision and objectives for the project, which centered around the need to get more archival descriptions and materials online for those organizations with archival holdings that did not already have the software to do so. We agreed that the ICA Congress in 2008 would mark the first milestone release for the software. It was also around this time that former ICA Secretary-General Joan van Albada chose the acronym for "Access to Memory" as the software name. Joan insisted that this and all other ICA products should use the "ICA" name as a prefix (e.g. ICA-ISAD) to raise awareness about the ICA and its activities. Therefore the software was officially named *ICA-AtoM*.

A second version of the ICA-AtoM prototype (alpha release 0.2) was completed in September 2006 and the results were presented to UNESCO. UNESCO was enthusiastic about the revised approach to the project, which would involve leveraging the online database work as a catalyst to launch an ICA open-source initiative. However, in spite of some early optimistic designs and timelines, the prototype stage had made it clear that to develop a fully-featured beta version of the software by the time of the ICA Congress in 2008 would require more time and money, and it was not entirely clear where any further funding would come from. I was working on a web archiving project for the World Bank Group Archives at that time and they agreed to contribute some of the time and fees, approximately €8,000 Euros, from that Artefactual Systems contract to continue design and development work on the next ICA-AtoM alpha release. Despite the uncertainty about project funding I was determined at this stage to ensure that the project would achieve its initial goals, and I carried on development between other consulting assignments, staying in close touch with Peter Horsman at the Dutch Archiefschool to brief him on my progress.

The faculty at the Archiefschool was very interested in the development of an open-source archival description application, sanctioned by the ICA and compliant with its standards, for use in classroom instruction. This helped them to avoid any conflicts that would arise if they chose to use one commercial application over another for classroom instruction. The Archiefschool was also interested in applying and extending ICA-AtoM as prototyping software in some of its research projects. Therefore, at a project meeting at the ICA Secretariat office to review alpha release 0.3 in early 2007, the Dutch Archiefschool committed to fund the next phase of software development in the amount of €100,000 Euros. Along with an additional €30,000 Euros made available by the Direction des Archives de France, this funding gave me the opportunity to work full-time on software development for the remainder of 2007 and to hire one additional programmer and one archivist to test the software and draft the first user manual. This funding helped to deliver alpha release 0.5, which I presented to the ICA membership at the CITRA meeting in Quebec City, Canada, in November 2007.

Prior to the CITRA meeting, Secretary-General Joan van Albada had asked me to prepare a budget to complete beta 1.0 release software development in the remaining six months leading up to the ICA Congress in July 2008. I presented a budget for

€167,500 Euros which included hiring two additional staff to complete all the remaining development, testing and documentation tasks. Once again, we carried on with the project work with blind faith that, if this project was in fact filling a real need for the archival community, the funding and resources would arrive in some form. After my CITRA presentation we had a roundtable discussion about next steps for the project and the variety of compromises or work-arounds I might implement to keep the project moving forward. At this point Abdullah Al Reyes, the Director of the United Arab Emirates Centre for Documentation and Research (CDR), and also vice-president for ICA Marketing, saw the potential that ICA-AtoM has to promote the ICA worldwide and committed the CDR to fund the 1.0 development budget, giving us the last push we needed to complete the first public beta release of the software in time for the ICA Congress in July 2008.

Project Governance

At this stage the project gained greater visibility and the various stakeholders agreed that a more stable governance model had to be put into place to ensure the long-term sustainability of the software and its community of users, developers and sponsors. The ICA-AtoM Steering Committee was formed to address this need. The current list of ICA-AtoM Steering Committee members is posted on the project website.[9] The Committee held its first meeting at the Dutch Archiefschool in Amsterdam in May 2008. There it drafted the following Project Purpose and Values statement.

The purpose of the ICA-AtoM Project is to provide free and open-source software:

1. that enables institutions to make their archival holdings available online, especially those who could not otherwise afford to do so
2. that manages archival descriptions in accord with ICA standards
3. that provides multi-lingual interfaces and content translation features
4. that supports multiple collection types
5. that is fully web-based, user-friendly and follows accessibility best practices
6. that is flexible and customizable
7. that is useful to both small and large institutions alike
8. that supports single or multi-repository implementations

[9] http://ica-atom.org/about.html#governance

The values of the ICA-AtoM Project are to conduct its affairs in the spirit of open-source collaboration, including:

1. openly sharing technical expertise
2. promoting professional best practices
3. nurturing an active community of users and developers
4. leveraging existing open web technology to deliver best-of-breed archives software
5. providing solutions to organizations with limited financial and technical resources
6. engaging active participation from organizations that have the necessary financial and technical resources
7. providing a common ground for cross-discipline collaboration with related communities
8. generating revenue to support ICA-AtoM and other ICA activities through a business model that benefits from widespread ICA-AtoM adoption

At the Amsterdam meeting in May 2008, the Committee also began discussion about the project's organizational structure and business model but these were not finalized at that time. Another meeting of the ICA-AtoM Steering Committee was held at the ICA Congress in Kuala Lumpur and I made a presentation on the project to the ICA Board. I reported that following extensive beta testing and additional core development, ICA-AtoM 1.2 should be the first production-ready, full release of the software. With the appropriate funding, release 1.2 could be made available by Fall 2009 in time for the next CITRA meeting in Malta.

However, the ICA-AtoM project costs are quite a bit higher than traditional ICA programs or projects and this was raising some concern within the ICA management. To be clear, the project has not taken away funds from existing program budgets. In fact, it has attracted new funding and resources into the ICA network. As well, the ICA-AtoM project expenses are actually quite reasonable compared to the budgets of similar projects, e.g. Dspace $2 million USD (2002), Fedora $4 million USD (2007). The main difference is that these other open-source projects have received large, one-time grants from major funding sources like the Andrew Mellon foundation. This is different from the more grassroots and serendipitous approach which characterized the early years of the ICA-AtoM project. That is not to say that the ICA-AtoM project is opposed to receiving grants.

This type of funding model (for example, an EU funding request) is something that has been suggested for further action at the previous two steering committee meetings.. However, the way that the project has unfolded thus far is to move forward with voluntary contributions from some key institutions that share the vision and objectives of the ICA-AtoM project.

After the 1.0 release this has included further funding from the United Arab Emirates Center for Documentation and Research, the Direction des Archives de France as well as some related ICA-AtoM implementation projects that Artefactual Systems is carrying out in Canada involving the Archives Association of British Columbia, the Library and Archives of Canada, the City of Vancouver Archives and Canadiana.org. These contributions will be used to deliver a stable, production 1.1 release in time for the next ICA CITRA meeting in November 2009. However, the more ambitious 1.2 release roadmap has been scaled back while the ICA continues to define the governance model of the ICA-AtoM project.

There is ongoing dialogue within the ICA Board, Secretariat and ICA-AtoM Steering Committee about the role of the ICA-AtoM Project. This project has challenged some of the traditional ways of doing things, in large part due to the speed and scope of the project developments. In many ways the governance is catching up to the technical developments: whereas I had defined a lot of the early direction and decision-making for the project, Artefactual Systems is now stepping back into a more formal, third-party contractor role while the Steering Committee assumes control over the central decision-making, promotion and communication. The Committee will likely be assisted shortly by a project coordinator based out of the ICA Secretariat office in Paris. My formal position as Software Release Manager is limited by the terms and length of any active contract between Artefactual Systems and the ICA. We are currently negotiating a contract for the release 1.1 development which will take place between May and October this year.

These are all positive changes in the growth and evolution of the project. They help to maximize the transparency and sustainability of the ICA-AtoM Project for years to come. I am pleased with Artefactual Systems' role as a third-party service provider and I hope that soon a number of other companies or consultants will emerge worldwide to provide ongoing ICA-AtoM support, training, development and hosting services for those archival

institutions that don't have these access to these kinds of resources within their own organization. This will encourage collaboration and competition equally, which is a signature quality of healthy open-source communities. Artefactual Systems will collaborate on the continuing development and enhancement of ICA-AtoM with other companies or institutions. At the same time, we will work hard to stay at the cutting edge of ICA-AtoM innovation and deployment to play a technical leadership role in such a community.

ICA-AtoM and the ICA Strategic Objectives

In the meanwhile, the ICA leadership is pleased to see more formal management structures put into place for the ICA-AtoM project and are encouraged by the potential of this project to meet all six of its recently defined 2008-2018 Strategic Objectives .

These objectives are:

1. Raising Awareness
2. Harnessing New Technologies
3. Building Capacity in the Archival Profession
4. Strengthening the ICA Network
5. Improving the Performance and Accountability of the ICA:
6. Building Partnerships

As free and open-source software that addresses a very practical need for archivists around the world, the ICA-AtoM application is already helping to address the ICA's first strategic objective to 'raise awareness'. The fact that I am standing here speaking to you today is direct evidence of that. Since the beta launch at the ICA Congress in July 2008, the ica-atom.org website has received 4,000 visits to the online demo version of ICA-AtoM. Much of the feedback we are receiving, including that from senior IFLA and UNESCO members, reflects pleasant surprise that ICA is pursuing such 'innovative' and 'cutting-edge' work.

We are helping to achieve the ICA's 'harnessing new technologies' strategic objective. ICA-AtoM is fully web-based software that is positioned to take advantage of new technologies which will allow archivists to provide timely and relevant technical solutions to the institutions and clients they serve. For example, the Archives Association of British

Columbia and the Canadian Council on Archives have evaluated a number of options for upgrading their archival description portals and have decided that ICA-AtoM provides the best architecture and features to meet their requirements. I will speak more tomorrow on the design and evolution of ICA-AtoM's technical architecture.

The ICA-AtoM project is also helping to achieve the ICA's third strategic objective to 'build capacity in the archival profession.' The ICA-AtoM release 1.1 will be fully-compliant with all ICA descriptive standards and will thereby serve as an excellent tool to promote and teach archival best practice. The Dutch Archiefschool and the ICA Section on Education have both beta-tested ICA-AtoM for use in classroom and professional workshop instruction.

The ICA-AtoM project is also helping to achieve the strategic objective of strengthening the ICA Network. By providing a free but high-quality software solution, the ICA-AtoM project is drawing interest throughout the ICA's international network, allowing archivists to work together and organize themselves regionally around a concrete project that is directly relevant to their day-to-day practice. This is currently happening most visibly in Canada, where I have spoken on behalf of the ICA and the ICA-AtoM project at several conferences and meetings. ICA-AtoM is the focus of a provincial and national portal upgrade and it was the predominant agenda item at the Canadian Council on Archives' Annual General Meeting in October 2008. Also, following this AAB conference we will hold a beta-testers meeting here in Rio de Janeiro for all those instructions that have already begun to implement the software.

The ICA-AtoM project is also helping to achieve the ICA's strategic objective to 'improve the performance and accountability of the ICA' by openly discussing the funding and governance issues as the project grows and evolves. As an open-source project, ICA-AtoM introduces a culture of open access and transparency to project management. Information about the day-to-day technical design, development and decisions is all available for monitoring online in the projects discussion lists, documentation wiki, code repository, issues list, and the logs of the online meeting room.

Finally, the ICA-AtoM is helping to achieve the sixth of the ICA's strategic objective, 'building partnerships.' From its outset, ICA-AtoM has been a collaborative project

involving partners inside and outside the archival community (e.g. the Canadiana.org project). It has also provided a common ground of collaboration with UNESCO, which is continuing with the Human Rights database project, beta testing ICA-AtoM for its own archives, and possibly working with its Memory of the World Subcommittee on Technology with Artefactual Systems and the City of Vancouver to integrate ICA-AtoM with other open-source tools to create a digital preservation system prototype. The ongoing success of the ICA-AtoM project will continue to depend on the ability to nurture and grow these types of partnerships.

Technical Architecture

So far, I have demonstrated the ICA-AtoM software and talked about the history and vision for the project. I would like to conclude my presentation now by taking a closer look at the analysis that led to software's technical architecture and data model. It is almost stating the obvious to say that most information seekers today turn to the Internet first to find resources related to their research needs. Although erroneous, there is a common belief that if information (or at least metadata about that information) is not on the Web, it does not exist, or it might as well not exist.

One of the primary motivations for the ICA-AtoM project was a concern that not enough archival materials, or at least archival descriptions, were available on the World Wide Web to represent the richness, variety and depth of the archival holdings that are available to researchers around the world. Therefore, the primary goal of the project is to provide free and open-source software that enables institutions to make their archival holdings available online, especially those which could not otherwise afford to do so. This means that the primary requirement for the ICA-AtoM software is that it be fully web-based and allow archival institutions to get their descriptions online where they will have the widest exposure to an international audience. Hence the name 'Access to Memory'.

So what does it take to get archival information on the web? The World Wide Web is based on the concept of resources, such as html pages, images, video, etc.. Each of these resources has its own unique resource identifier made available by web servers that respond to requests sent by client applications, namely web browsers such as Mozilla Firefox, Safari, Internet Explorer, etc. The simplest way to publish information on

the web is to manually create a static HTML page. Many archival institutions still maintain their own websites this way. As well, over the past decade many archival institutions have invested time in creating EAD XML finding aids, and they make these available as stand-alone XML files or convert them to HTML or Acrobat PDF files and post those as static files. However, these all have to be managed as individual resources and it is often difficult or not possible at all to provide integrated search, browse and presentation functionality for these resources.

A more efficient way to publish information to the web is with database-backed web applications where all the resources being made available on a website are being maintained and stored as an integrated system. Blogs, wikis, and web content management systems such as Drupal or Joomla! are all examples of this type of web application. In early 2006 I conducted a technical analysis of existing open-source software applications to determine whether we could simply extend these types of applications to meet the needs of the ICA-AtoM project. I evaluated Fedora, Dspace and Greenstone. All three are excellent open-source projects originating from the digital library or digital repository communities. However, at that time, each of these projects had fairly rigid internal technical architectures that made them difficult to adapt to archival description requirements. This was especially true of their data models, which would need to be adapted to allow for compliance with the ISAD(G) archival description standards and principles, such as flexible, multi-level description. However, an ever bigger issue is that these tools were Java framework applications which are difficult to package and deploy on cheap, shared hosting services. Part of the ICA-AtoM objective is to make an application that is easy to download, install and maintain on basic shared hosting services. These hosting providers tend to favour simpler LAMP stack applications.

LAMP is an acronym for Linux, Apache, MySQL and Python, Perl or PHP. It represents the basic and most popular open-source web application architecture that uses Linux for an operating system, Apache HTTP server as a web server and MySQL as the database server, in combination with a web application written in the Python, Perl or PHP scripting language.

Consequently, I turned my attention to some of the more popular open-source PHP web publishing platforms, namely Wordpress and Drupal. Both were already in wide use in 2006 and had large communities of users and active developers. However, I ran into the same issue again. It was difficult to adapt these applications to implement the data model required for archival description.[10] More troubling was that both of these applications grew very organically within a loosely organized open-source community. This was part of the reason for their popularity but it also meant that the underlying technical architecture for each of these applications were not very cleanly or clearly designed. The vision for ICA-AtoM was to implement a well-designed, scalable architecture that can form the basis for a long-standing archives management system which is adaptable to increasingly complex requirements for web services and digital preservation. It required a clearly articulated technical design and this ruled out Wordpress and Drupal at the time.

I should note also that in early 2006 there was no evidence yet of any other open-source archival description applications. I had heard about the Archivists' Toolkit project but it did not make its first version available until early 2007. Likewise, I did not hear about the Archon project until I attended its launch session at the Society of American Archivists conference in August 2006. However, I did stop and take a close look at both of these applications when they were released. At that time both used EAD as their core data model instead of ICA standards and neither offered multi-lingual capabilities or multi-repository capabilities, all of which are critical requirements for the ICA-AtoM application. Furthermore, the Archivists' Toolkit was always intended as an application to support the internal processing of archival materials, rather than providing access to archival holdings. It allows for the export of EAD finding aids but it does not provide a web-interface to its database of archival descriptions.

Both of these applications have recently added support for the American DACS standard which is very closely related to the ISAD(G). As well, Archon has added the ability to translate the application interface (although not the database content) and the Archivists' Toolkit can be installed in a multi-repository configuration. However, neither of these two tools address all of the same critical requirements for the ICA-AtoM Project. I also think it

[10] The Drupal CCK feature, which allows for the creation of custom content types in Drupal, was only in its design stage in early 2006.

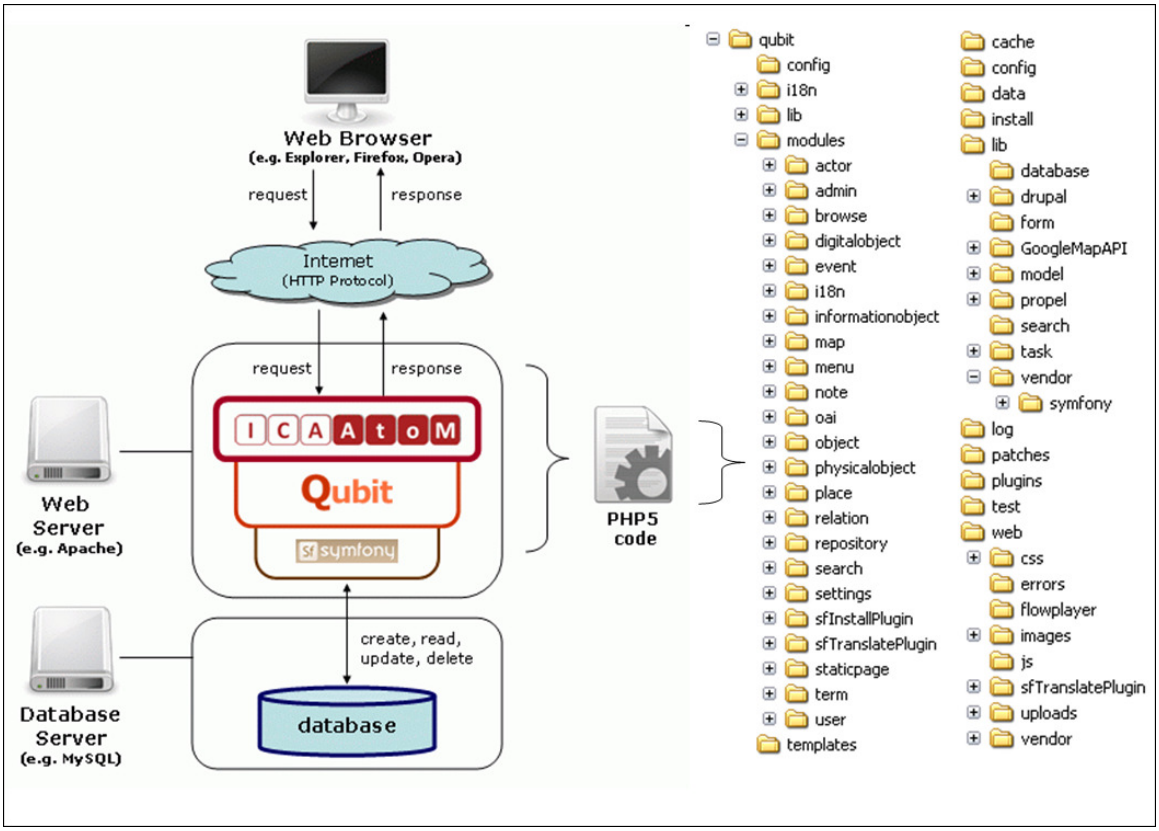
is fair to say that both of these projects are targeted largely to the needs of the American archival community and its sponsors. So merging the ICA-AtoM development with one of these projects did not make sense back in 2006. I believe that there are still enough differences in the technical architectures and community priorities to warrant separate projects today. I also believe this is a good thing for the archival community as a whole. Whereas we had been starved far too long for legitimate open-source archival management applications, archivists now have three excellent projects to choose from, allowing them to select the best tool and community to addresses their particular needs.

So, after completing my analysis of technical options in early 2006, it looked as though the only remaining option was to develop an application from scratch, a daunting and ambitious task given the very limited resources available to the project. However, at that time there was another trend emerging in web development, namely the popularity of Model-View-Controller (MVC) web frameworks for rapid development of web applications. I tested the most popular of these, namely Ruby-on-Rails, Django, and CakePHP before settling on Symfony.

I favoured Symfony because it was PHP-based, making it easier to deploy on a wide variety of shared hosts; it was also designed for developing enterprise applications and was based on fully object-oriented PHP5, allowing for the implementation of a number of best practice design patterns. There are many advantages to developing ICA-AtoM using the Symfony framework. By using the Symfony framework we avoid having to develop and document a whole number of common web application features, such as caching, user sessions, routing, forms templating, etc.. Instead we can concentrate on developing the business logic and templates required for an archival description application. This has saved us a large amount of development time and cost. We are also benefitting from being part of the larger Symfony open-source community so that we are not "going it alone". As well, it is much easier to add new developers to the project, since Symfony provides rich documentation and a consistent structure and logic for organizing the application code.

Symfony is a fairly typical implementation of the Model-View-Controller pattern. MVC is a software engineering architectural pattern which isolates data persistence (i.e. the model) from business logic (i.e. the controller) and from user interface components (i.e. the

view), resulting in an application in which it is easier to modify the visual appearance of the application without affecting the underlying business rules and vice-versa. In a Symfony application, client browsers interact with the application through the front controller which is the *index.php* file located in the publicly accessible */web* directory. Using the Symfony routing system rules, the front controller sends the client requests to the appropriate action code in one of the application's modules.



For example, this request maps to the informationobject module:

http://ica-atom.org/demo/web/index.php?informationobject/edit/id/100

It will load the edit action and pass 100 as the id parameter value. This will trigger some application logic that reads the informationobject with the id value of 100 from the database. The corresponding edit template is populated. The template is then decorated with the global layout template, which includes the header, footer and other generic layout elements. The output is sent to the user's browser where ICA-AtoM waits for the next request. Entering some data and clicking the save button will send an *update* request to the application, and entering a search query will send a new *read* request.

Each module such as informationobject, actor, or repository corresponds closely to an object in the application's data model. Most requests usually include some database create, read, update, or delete (CRUD) activity before passing the response back to a template in the view layer. The database CRUD actions are handled through an object relational mapping (ORM) layer which converts programmable application objects into rows or records in the underlying relational database, which is still the most effective and scalable data store for web-based applications. The MySQL database is used in development but ICA-AtoM uses the PEAR PDO database abstraction layer within the ORM and it is therefore also compatible with Postgres, SQLite, SQLServer, Oracle, etc..

We have also integrated Symfony with a number of other open-source libraries that provide additional features, such as JQuery and YUI for AJAX controls and effects. AJAX is a methodology that allows for web application controls to provide dynamic visual effects or get and post data to the web server without requiring the web page to reload. AJAX controls allow web applications to behave more like desktop applications and provide for a better usability experience.

By mid-2006 I had developed the first two alpha releases of the ICA-AtoM application using the Symfony framework. I had demonstrated it to the ICA Secretariat and we received approval from UNESCO to apply the 40,000 Euros they had made available for the Human Rights Archives web portal as seed money for this prototype. However, project financing and therefore development was limited through to mid-2007 when the Dutch Archiefschool committed funds to work towards the first beta release of the application.

Around the same time I had been in contact with developers of the Alouette Toolkit, another early-stage open-source development project based in Canada. They were also looking to develop a web-based application to help memory institutions in Canada that did not have any existing tools to make their digitized collections available online and to produce standards-compliant metadata that could be used in the Alouette website, a search and browse portal for digitized Canadian heritage content made available by libraries, museums, galleries and archives.

The underlying technical vision for the two projects was almost identical, as were the multi-lingual requirements since Canada is officially a bilingual country. The only slight differences were the data standards and the target communities for which these were being developed. The primary sponsors and target community for the Alouette Toolkit was Canadian research and special libraries with a focus on the Dublin Core and MODS standards. After some discussions and evaluation of the initial ICA-AtoM prototypes, the technical teams for both projects decided that, since both projects were starting from scratch, it was to the benefit of both to share the limited time, technical resources and finances available to each and to collaborate on the underlying technical platform. I should add that governance for the Alouette Toolkit has recently been taken over by Canadiana.org, a consortium of major Canadian memory institutions promoting digitization and access to early Canadian works. Canadiana.org will launch its first public release of the software in May 2009, renaming its version of the application to Digital Collection Builder.

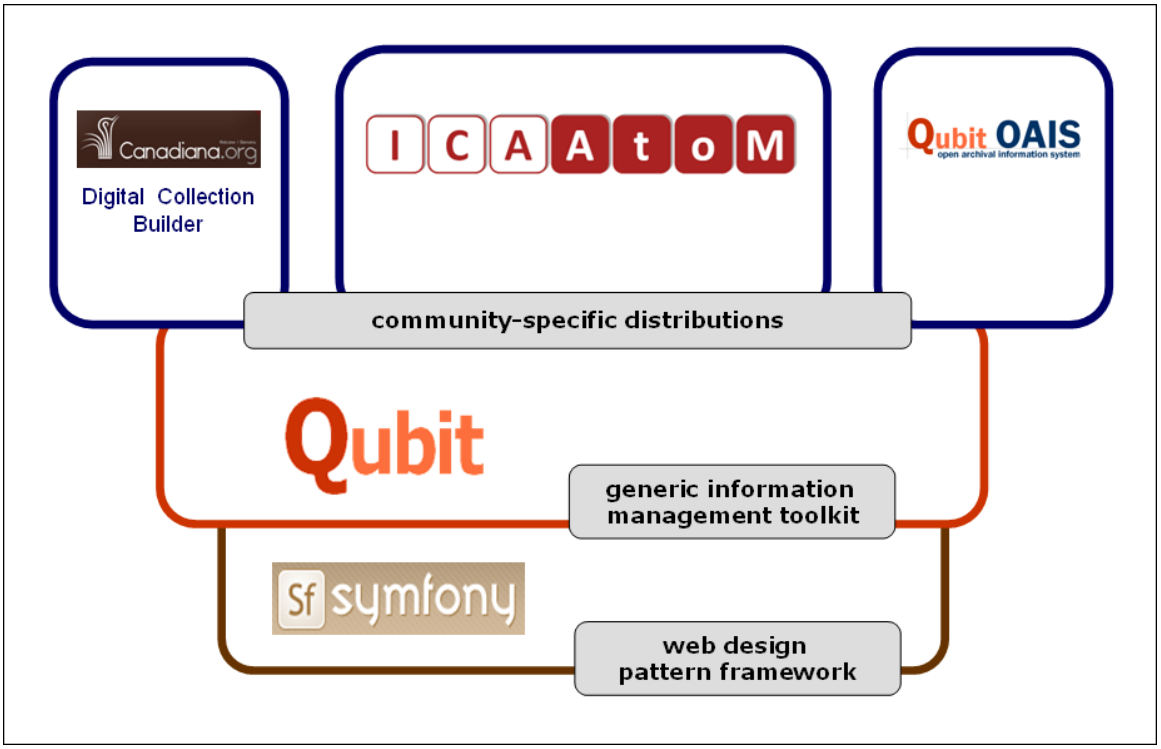
I had already designed the ICA-AtoM object model to be as flexible and generic as possible. Part of the rationale was that archival institutions would also need software to help catalogue their published reference material or any artefact collections for which they may be responsible. I felt that this capability might as well be provided by the same application. From a technical point of view, managing metadata about information objects, whether these are archival materials, published materials or artefact materials, can be made generic enough to accommodate all collection types while still respecting specific descriptive standards and practices. I also designed the data model to ultimately apply archival description to electronic records and digital collections as well as to map the ICA metadata to compatible standards. This flexibility made it possible for the Digital Collection Builder to adopt and adapt it for its own use.

The remaining issue was to appease the sponsors and target community of each respective project with an application that was branded and customized specifically to their audiences. To allow for this we created a generic underlying toolkit, which we eventually codenamed Qubit. A 'Qubit' is a quantum bit of information, the smallest conceptual unit of information know to science. It is an appropriate name for a system that is designed to manage untold Qubits of information and provide access and context

for them. Both the ICA-AtoM and Digital Collection Builder applications are managed as distributions of the Qubit toolkit.

The concept of *distributions* is fairly common in the open-source community. For example, there are a large number of Linux distributions (such as Red Hat and Ubuntu) each of which is based on the Linux kernel that is officially maintained by the original Linux developer Linus Torvalds. Torvalds decides what new code is incorporated into the Linux kernel on which the distributions are based. Another example is the Drupal open-source content management platform which has several distributions that are made available as installation profiles. These include distributions for education sites, civic organizations or the music industry.

Similarly to the Linux and Drupal projects, Qubit now has several distributions. *ICA-AtoM* is a Qubit distribution for the international archival community, *Digital Collection Builder* is a Qubit distribution for the Canadian Heritage Community, and *Qubit OAIS* is a more recent collaboration between Artefactual Systems, the City of Vancouver Archives, the InterPARES Project and possibly the UNESCO Memory of the World to create a Qubit distribution for institutions implementing OAIS-compliant digital preservation systems. Each Qubit distribution and Qubit itself is released under a GPL free and open-source license which gives each project and community the right to revise and re-distribute the software as it sees fit.



Each of these projects is contributing technical resources and financial sponsorship to develop the underlying Qubit architecture to meet the requirements of its particular sponsors and the needs of their community. Each one benefits in turn from contributions made to the underlying platform which become available to them in each subsequent release of the software. For example, much of the application core was developed via the ICA-AtoM project but the recent ICA-AtoM 1.0.5 beta release contains a custom menu module, multi-page digital object browser, and Dublin Core templates, all of which have been funded by the Digital Collection Builder project as contributions to Qubit.

So this is how we arrived at the current technical architecture for ICA-AtoM. As demonstrated yesterday, the application is fully web-based. All user interactions with the system take place through the user's web browser. Users access resources such as archival descriptions, digital objects, authority records, etc. on the web server; clicking a button or link triggers a PHP script that sends a command to the database and returns the output as HTML back to the user's browser.

ICA-AtoM is developed and supported on a fully-open source LAMP stack rather than proprietary software. Consequently, there is no cost to download any of the software

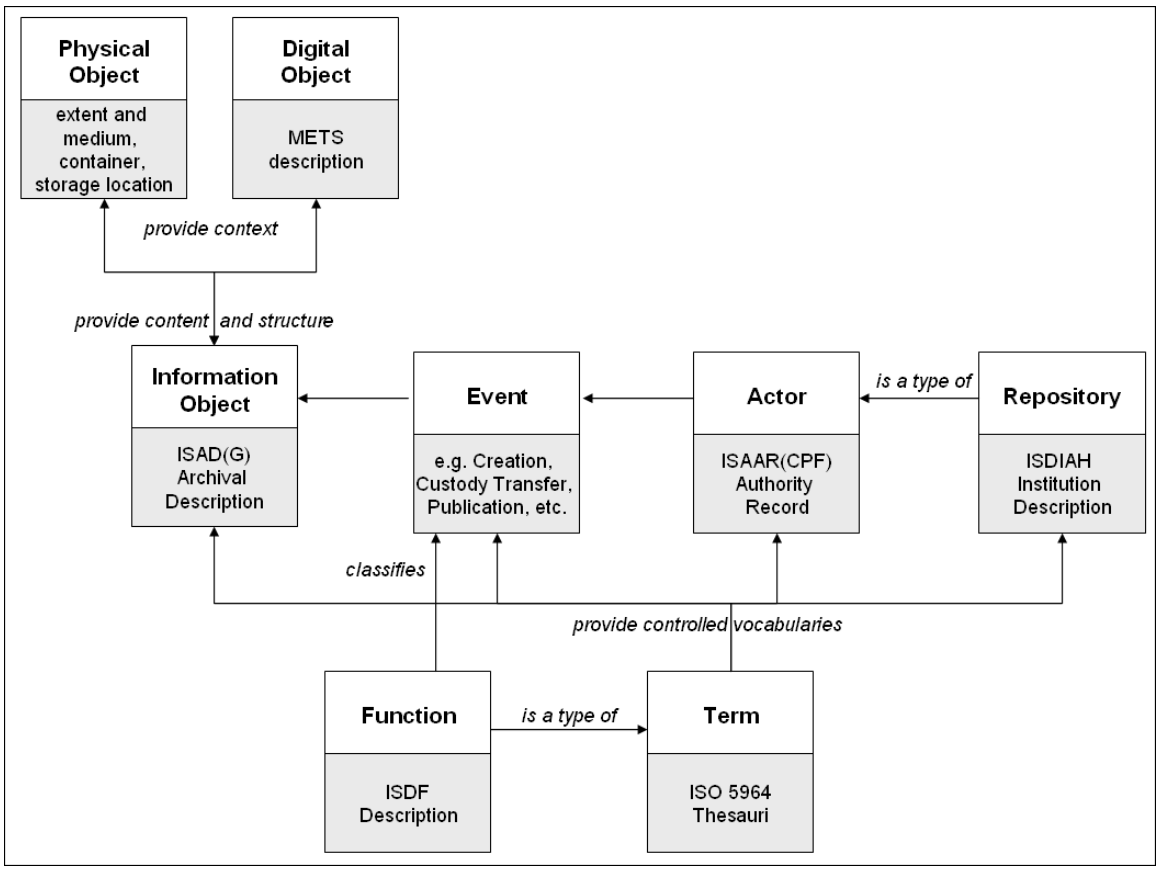
required to run the ICA-AtoM application, and users are free to make any enhancements or changes as they see fit.

ICA-AtoM can be used by a single institution for its own descriptions or it can be set up as a multi-repository "union list" accepting descriptions from any number of contributing institutions via direct data-entry, EAD import or OAI harvesting. Artefactual Systems is pilot testing this scenario and architecture right now in a collaboration with the Archives Association of British Columbia, the Canadian Council of Archives and the Library and Archives of Canada. There are many small archives in Canada that still do not have their own archival description software. Also, they don't really have the technical capacity to install and host a web application like ICA-AtoM. Representatives from these institutions will be given a repository-specific account so that they can enter and update archival descriptions directly in the central union portal. All they need is web access. Other institutions that have software with EAD export capability can use this method to contribute to the union list. Finally, our ideal scenario is to use OAI harvesting for the automatic export and import of archival descriptions throughout a hierarchy of provincial, thematic, or national networks. The 1.0.6 release will have the capability to harvest Dublin Core records, the standard metadata format used for OAI. However, simple Dublin Core is not rich enough to support archival description relationships, therefore, we are looking ahead to use EAD records within the OAI protocol.

Data Model

The ICA-AtoM data model is designed to be flexible enough for adaptation to other descriptive standards and it currently supports crosswalks from ISAD(G) to EAD, Dublin Core, the Canadian Rules for Archival Description, and MODS. The ICA-AtoM data model consists of a number of core entities, namely, archival descriptions, authority records, archival institutions, terms and events.

Archival materials are a type of information object that we preserve for future access. By thinking of archival materials in a more generic way, as information objects, it makes the underlying data model more flexible and allows it to be re-used for other types of information objects, such as published material or cultural artefacts.



Archival descriptions provide contextual information about archival materials; they are representations of the archival materials. Archival descriptions are arranged into hierarchical levels (e.g., fonds, series, files, items). This is done in ICA-AtoM using a flexible and sophisticated nested set hierarchy model. Each archival description includes the data elements found in the ICA's ISAD(G) standard. Each archival description may be linked to a physical object or a digital object describing location, physical characteristics, structure and technical information about the archival material or record itself. Physical objects and digital objects posses content and structure. They need the context provided by representation information to become information objects. The full name for the information object in the ICA-AtoM data model could be 'information object representation' but that would be redundant. In the end, all of the objects in the data model contain representation information for the objects they surrogate, so the 'representation' suffix is implied.

Authority records provide descriptions of the actors that interact with the archival materials represented by archival descriptions. The primary actor in an ISAD(G)

description is a creator. ICA-AtoM authority records include the data elements found in the ICA's ISAAR(CPF) standard.

Archival institution entities provide descriptions of repositories that preserve and provide access to archival materials. An archival institution is a type of actor. Like all actors, an archival institution has its own ISAAR authority record. But its characteristics as a repository (e.g. its opening hours, research services, contact information) are described separately in an ISDIAH institution record. In other words, an archival institution record inherits all the applicable fields from the Actor's authority record (e.g. Authorized Form of Name, History, etc.) and extends the repository-specific fields in an ISDIAH record.

Term entities organized into separate taxonomies provide controlled vocabularies used throughout the system (e.g. as subject access points or as drop-down menu options for levels of description or authority types, for example). Release 1.1 will be upgraded to make all the term taxonomies compliant with the structure and behaviour of the ISO Thesauri standard, which calls for the implementation of equivalence (Use, Use for), hierarchical (Broad term, Narrow term) and associative relationships (See also).

Most definitions of records in archival science emphasize the importance of 'acts', 'transactions', or 'business processes'. The Event entity formalizes this concept in the underlying conceptual model rather than being lost in a weak link directly between an archival description and an authority record. Event relationships link ISAAR authority records (descriptions of actors) and ISAD(G) records (descriptions of archival materials).

Through events, a single actor can have zero to many relationships with zero to many archival units and each relationship can be treated as a separate entity, delimited by normalized start and end dates. The ICA-AtoM Events are very similar to the 'Business' entity in the Australian SPIRT metadata model[11] but also the Event object found in the PREMIS digital preservation metadata standard, the former capturing the context of the record's relationships in its business, juridical and procedural context and the latter capturing the physical context of the record's relationships to its content and structure.

[11] http://www.sims.monash.edu.au/research/rcrg/research/spirt/deliver/conrelmod.html

The Event entity is designed to capture all events related to the archival materials, not just creation. This could include custodial history changes, conservation treatments, digital object reformatting, access logs, description revision history, etc. The Event entities provide the true context of what has happened to the archival materials over their lifetime. This is especially useful in managing digital archival materials and adapting ISAD to the description of digital archival materials.

Functions described using the new ICA International Standard for Describing Functions (ISDF) can be used to further classify Events. For example,

- the Event entity indicates when a record was created,
- a link from the Event entity to the ISAAR authority record indicates which Actor created it,
- a link from the Event entity to the ISAD record describes the document that was created
- and a link from the Event entity to the ISDF record classifies what type of act the Event is in relation to the organization's functions (e.g. the Event is an 'instance' of the function described in the ISDF record).
- Also, a link to the Place term taxonomy can classify the geographical location of the event.

All of these Event relations can be used to build rich search and browse interfaces to the archival materials. I believe this design is a practical and powerful way to implement the otherwise theoterical definition of a record as 'process-bound information' (as the Dutch school calls it) or the dispositive and probative records defined by *diplomatics*.

However, the recent ICA-AtoM beta testing phase did raise some concerns about how the Event concept related to the existing ICA standards and how archivists would normally expect to enter creation dates and links to the authority records for creators. Therefore, based on the feedback from beta-testers and the DAF report, it has been decided to completely hide the Event entity from the end-user in release 1.1 and to limit the Event types to creation events only in the ISAD(G) templates. Unfortunately, this will mean that creation dates will not be entered in direct relation to creators and it limits the potential of the Event entity concept. However, it will remain in the underlying data model and likely be re-introduced as a link initiated from the Actor ISAAR record back to the

archival description. It may also be re-introduced in some other form as the ICA Committee on Descriptive Standards begins its own work on a conceptual model for archival description. Also, the Event entity is still accessible from the Dublin Core template, which calls for Creator, Publisher and Contributor actors, and the Canadian Rules for Archival Description template which incorporates a number of different types of dates and actor relationships.

Closing remarks

Of course, I am fascinated by technology and archival systems and I don't really tire from talking about these topics. However, I also recognize that not every archivist shares this interest as strongly. Therefore, I want to thank the AAB once again for the honour to speak to you at length about the ICA-AtoM project and technology. I hope that some of you will decide to check out and join the ICA-AtoM community or that I've at least discussed some things that help you to evaluate or analyze the technical architectures within your own institutions. *Obrigado, tchau e ate a próxima.*