

R&D @ Laval University Library - Archimede and ETD's

Verbatim of a presentation given by Guy Teasdale at Iatul 2005 on June 2, 2005, 15:00. To be read alongside the PowerPoint presentation

Good afternoon. My name is Guy Teasdale. I am director of support and development services at Laval University Library. I understand that I have 15 minutes so I will just outline a few of the Research and Development projects taking place at Laval concentrating more specifically on our Institutional Repository System

Mr. Rida Benjelloun, head of our Digital Initiatives Research and Development Office is here with me if you have some tougher technical questions to ask later.

CLIC

Before we begin let me talk about the peculiarities of our organizational chart. A few years ago, responsibility for the Library's development projects was given to the Department of Support and Development Services **CLIC** so that the Librarians in the Science Library and the Humanities Library could concentrate on their four main user services tasks: Information Communication, Collection Development, Library Instruction and Cataloguing and Indexing of original documents. To be more precise, I should mention that most of our librarians are also participating in development through involvement in the different Committees

Anyway this administrative structure allows us to undertake more demanding projects like one that I'll explain later.

CLIC

We have been involved in many collaborative projects with many partners across Canada. Let's mention two of the largest Canadian digitization projects:

In the first, Early Canadiana Online, we were pioneers contributing to the pilot project. This project is the result of a partnership between the Canadian Institute for Historical Microreproductions (CIHM), the University of Toronto Libraries, the Université Laval Library and the National Library of Canada

The research and development aspects were in:

- Formatting: we used SGML ;
- Metadata: we used Text Encoding Initiative elements
- Search and retrieval tools: we applied raw OCR on digitized microfiches
- Web Interfaces: these were completely bilingual

This project is now fully managed by CIHM and has become the most important large scale digitization project in Canada (currently they have almost 2 million pages from almost 13 000 books)

The second project, Our Roots is also the largest and most complete digitized collection of local and regional histories from across Canada. It is a fully bilingual site containing public domain and current copyright cleared materials. In this project we applied metadata standards and best library practices in digitization and digital preservation. Our Roots is an initiative funded by Canadian Heritage and a partnership between :
2 Lead Institutions, the University of Calgary Press and the Université Laval Library, plus many Digitization nodes in Universities across Canada

Now for the main part of my presentation I will concentrate on some aspects of our e-publishing initiatives which comprise our Institutional repository system:

CLIC

Erudit, our ETD project and Archimède

CLIC

While investing in large digitization projects mainly targeted on materials of the past, we became also concerned by the incremental production of *born digital* materials on campus. We noticed a proliferation of Web sites for different research communities each with its strengths and weaknesses. As we wanted to organize efficiently the intellectual production of our University, we got involved in e-publishing

CLIC

While developing our institutional repository system, our primary objective was to give access to all this information by providing users with a federated search tool capable of finding content in all Université Laval's research communities. Users would no longer have to browse through a great number of Web sites to gather the information they need. But doing so we tried not to mix apples and oranges, we maintain distinct sites for different purposes. For now we consider that our Institutional repository comprised of these three components. (Later we may add another one for learning object materials and one for administrative e-documents).

The objective was not to gather a massive amount of materials together at the same place but to have some materials of the same type, easily searchable, separately or together if necessary. With the interoperability technology we are using we could pretty easily assemble a big set of electronic documents and make it searchable but this could end up like this:

CLIC – image of the BLOB

Which we would like to avoid. Lets talk about these three components.

1)

Érudit is a Portal hosting and giving access to more than 10 000 articles from 50 scholarly journals and other research publications from Québec Universities. It is a not-for-profit consortium created by Université de Montréal (IT services), Université Laval Library and Université du Québec à Montréal (Research Office). Érudit is funded by the three partners institutions, the Québec granting council FQRSC and the Québec government.

Its aim is to facilitate long-term preservation and access. We have a strong commitment to openness and interoperability and to do so we are using open formats : XML as an acceptable archiving format that also facilitates the exchange of metadata.

Partnership and interoperability projects have just been signed with France's Centre national de la recherche scientifique which is using the Érudit DTD and our production workflow. (I should mention that this DTD is also available in English through project Synergy)

Concerning our partnership, the French were concerned, as we were, by the risk of seeing their cultural heritage taken in charge by foreign publishers or commercial vendors.

There is no doubt that Érudit is of very high Importance for scholarly publication in French.

Electronic Theses and Dissertations (ETD)

Our ETD project is a joint realization of the Library and the School of graduate studies of Université Laval. It enables electronic submission by students using MS Word, LaTeX or OpenOffice (soon to be implemented) after they followed an Online Training module on WebCT on the correct use of style sheets. This permits a semi-automated workflow allowing automatic transfer from MsWord, LaTeX, OpenOffice to XML format (our archiving format) We are using the DocBook DTD

- Then with an XML file we can transfer from XML to XHTML and PDF as dissemination formats
- The ETD metadata are in XML, allowing an automatic conversion to MARC records. We are also conforming to OAI protocol in order to actively participate in the Library and Archives Canada portal for e-thesis and dissertations.
- Laval is one of the few Universities in Canada using XML as an ETD archiving format (I think that the University of New Brunswick is also using XML)
- I am pleased to add that we will organize the next Networked Digital Library of Theses and Dissertations (NDLTD) conference in June 2006.
- Partly in preparation for this we are planning to migrate our ETD into Archimede this year in order to benefit from its multiple possibilities and also to get to a critical mass of documents as a proof of concept. Speaking about Archimede

Archimède

Archimède is Université Laval's solution to host publications and data of Laval research communities to improve scholarly communication and to disseminate ~~the~~ research results.

- It was conceived and developed at Université Laval Library by Rida Benjelloun and Nicolas Bélisle, from a range of Java technologies and full Open Source applications. It is freely available through a GPL Licence.
- It is a decentralized system arranged around auto-regulated research communities which have collections containing one or more files each associated with 3 to 15 metadata elements.
- You may find a detailed description in the Soros Foundation « Guide to institutional repository software »
- There is also a more technical article to appear in Library *Hi Tech* and in a book about Institutional Repositories also to be published.

We are presently undertaking a second phase of development at Laval. I should mention that being Open Source, Archimède has already benefited from the open source developer community. For example the indexation and research capabilities of Archimède were well received by the Open Source community. This led us to make a separate package for LIUS (Lucene Index Update and Search) and distribute it as a distinct tool. It only takes a few lines of code to add it to any Java application to expand its document indexation capabilities. A developer has offered to add indexing of MP3 and MPEG files.

DIAPO 10

WHY DEVELOP OUR OWN SYSTEM?

Prior to starting the development of Archimède 2 years ago, we made an analysis of available *Open Source* software solutions and decided to create our own customized application. This was not because other repository systems were not good, quite the contrary since we modelled ours on them, but because none met all our needs.

Université Laval being a francophone institution, our first criterion was that the system must meet the needs of our French speaking users. The software we were looking for had to have an easy way to integrate additional languages, such as French, into its interface without having to rewrite the code of the interface after each upgrade of the code. However, none of the solutions that we looked at offered this possibility.

The second important criterion for our Library was to be able to index metadata **and** full text, from many document formats. Our surveys revealed that most authors provided very minimal metadata about their documents. Frequently, metadata contained only a "title" and an "author". This led us to ask the following question: « If the title is not relevant, how could one find the document? ». Most of the tools analyzed did not index metadata and full-text, and those presenting such features were limited in the document formats supported. With Archimède we have taken into account the maximum

number of formats possible, ensuring a maximum flexibility of our repository system. We also took into account the fact that academics don't like to input metadata. So we designed a metadata export/import facility.

As our third criterion, we were looking for a system that could run on Windows as well as on Linux. At the time, the Université Laval Library only had Windows servers (and expertise) and we wanted a product that could adapt easily to the technological infrastructure in place, without buying a dedicated server for the software. All the solutions that we looked at during our analysis were designed for Linux or UNIX, so it was obvious that we had to take the cross-platform approach while developing our application.

Aside these criteria, we had to take into account of a tight development budget. To make it all possible, we chose to work with well known and approved *Open source* software, and used *automatic code generation* tools for everything related to the database and the persistence layer.

CLIC

Archimède offers a secure module based on privileges. This allows restricting access to resources and functionalities by user login. Thus, we have 5 user types with different levels of privilege:

- Visitor: Each visitor to the institutional repository can browse all the public deposits of the research communities. The visitor status authorizes browsing and permits simple search or advanced search.
- Community user: The user of a research community can browse all his specific community deposits and the public deposits of other research communities.
- Community member: The member of a research community can browse all his specific community deposits and the public deposits of other research communities. He may also add or remove documents in his community collection.
- Community administrator: The administrator can browse ALL his community deposits and the public deposits of other research communities. He can create or delete collections in his own community, upload documents or delete any deposit created by his community members. Only the community administrator can create the collections where the deposits are uploaded.
- System administrator: The system administrator has all the rights and privileges – creation, modification, and deletion. He also has full access to the system management functionalities such as importation, exportation, re-indexing, etc.

Archimède's goal is to host all Université Laval's autoregulated research communities. The users of these communities are invited to upload their publications with the appropriate description (metadata) by means of a very convivial and secure interface.

This secured module allows the uploading of several documents at the same time, with forms to enter metadata. The files are hosted in virtual folders that restrict access to authorized users only.

A navigation system and a powerful search and indexing tool that ensure precise access to content.

In addition to the document management functionalities, Archimède offers mechanisms for ensuring the dissemination of content. This is made possible through the navigation and search features of Archimède and through compatibility with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH 2), thanks to the use of the Dublin Core metadata element set. Archimède also offers a « selective dissemination of information » service (SDI) that sends new content in their fields of interest to registered users. All this information is protected by an access control system based on privileges, allowing the limiting of access to resources and functionalities of the system.

CLIC

Here is a view of the technological architecture of Archimède which will be described in detail in an article to be published in Library HiTech. Those interested in more information can ask M. Benjelloun

CLIC

CONCLUSION

In this brief review you can see a common pattern in our digital initiatives:

- Commitment to partnership and cooperation in and outside the library community
- « Work in isolation is non productive »
- Commitment to large-scale and/or open source projects because Local only initiatives are not easily sustainable in the long term
- Commitment to standards and best practices
- Commitment to long-term preservation and accessibility (through standards)
- Commitment to open access and interoperability
- Commitment to research and development

Thank you for your attention