

**THE REJECTION OF D-SPACE:
SELECTING THESES DATABASE SOFTWARE
AT THE UNIVERSITY OF CALGARY ARCHIVES**

Lisa Atkinson, University of Calgary, Canada
lisa.atkinson@ucalgary.ca

1. INTRODUCTION

This paper presents the results of a recent University of Calgary Archives project to investigate the replacement of our existing theses and dissertations database. In collaboration with the University Library's Information Technologies department the Archives undertook a pilot project aimed at replacing the existing descriptive database and expanding on its capabilities by implementing D-Space as an institutional repository for electronic theses and dissertations. Ultimately D-Space was rejected as a viable option for housing our theses database, although the software is still used by the University for its Institutional Repository which contains other forms of scholarly output by the university's academic community. The Archives has since selected another software program for its theses database which I will provide details of later in this paper.

This project to replace our theses database was part of a larger initiative to establish an electronic theses program at the University of Calgary. Due to time constraints my comments here will focus only on our findings in regards to D-Space but I would be happy to discuss the e-theses program later if anyone is interested.

2. THE CONTEXT

I wanted to start off this paper by putting my comments about our experiences with D-Space at the University of Calgary into context particularly for those of you from outside Canada who may never have heard of Calgary, let alone know anything about the city or its university.

Calgary is located in western Canada within the province of Alberta, about 300 kilometres north of the Canada-United States border, and about 100 kilometres east of Banff National Park in the Rocky Mountains. The University of Calgary is a relatively young institution – it was established as an autonomous, degree granting institution in 1966 though its roots date back to 1905. It is a medical-doctoral university offering professional programs in engineering, law, medicine, nursing, social work, environmental design and, beginning in 2007, veterinary medicine. It has 16 faculties, 53 departments, and over 30 research institutes and centres. In 2005-2006 the University had approximately 28,000 full time equivalent students, including 5,300 graduate students. Some of these graduate students are engaged in programs which are course based and do not include a thesis element.

Currently, between 550 and 600 theses and dissertations are completed annually, and in 2005 the theses collection contained approximately 13,000 titles. The University Archives is responsible for preserving and making available this collection of original theses which is recognized on campus as representing a significant portion of the research output of the institution. The University Library holds microfiche copies of the majority of the theses held in the Archives, and it also has a large collection of bound theses which researchers can take out on loan.

3. CAVEATS

In coming to this conference to talk about electronic theses generally and about D-Space more specifically, I am somewhat uncomfortably aware that I am likely speaking to an already converted

audience: that many of you are convinced of the benefits of both e-theses and of this software. Certainly, in preparing for this talk it became clear to me that there are numerous universities across the world that have chosen to use D-Space to make their electronic theses available to researchers. I noted however that D-Space does not have a total stranglehold on the market, but that there are universities that are using or proposing to use something other than D-Space for their electronic theses databases.¹ So perhaps there are those of you who have found like we did that D-Space is lacking some elements essential to your specific requirements, and perhaps others who are still in the process of weighing the available options for their e-theses repository software.

My comments about D-Space come with a couple of caveats. Firstly, the IT department of the Library, in hearing that I was proposing to give this talk on our experiences, informed me that the version of D-Space we used during our pilot project is not the most recent version now available. It is possible that the current version of D-Space offers some of the functions I found it lacked when we undertook our research on it. I don't think that is the case: I didn't find anything online when I was preparing this paper that was more fully functional than the version we were using, but you should nonetheless be aware of that consideration.²

Secondly, to clarify, this project was interdisciplinary in nature, involving specialists from information technologies and from the Archives. The Library's IT department was responsible for configuring D-Space to the Archives' specifications and for providing us with technical support and information on the capabilities of D-Space. The Archives, as the creator and user of the database, was responsible for determining how the data in D-Space should be configured. Once data had been migrated, the Archives was also responsible for determining whether the system's technical capabilities measured up to the requirements we have in performing our day to day activities *vis a vis* the theses collection. I therefore can answer for the aspects of the project the Archives was responsible for, but not those that fell into the domain of IT.

4. WHY D-SPACE?

So how did the University of Calgary Archives come to contemplate using D-Space to make its theses available to researchers?

Since about 1998 the University Archives had had an online theses database, which contained descriptions of the theses in our holdings. The database was available on the internet and researchers worldwide were able to search it to find theses that were relevant to their work. No digital copies of the theses were attached to the descriptions, but researchers had a number of options available to them to access content which were explicitly stated on the theses website.

Some months ago, this public database ceased functioning. It was soon clear to the Archives that our clients wanted access to this resource and there was pressure to replace it quickly. To complicate the issue, the theses database was only one module of a suite of software programs used in the Archives to control its large collection of permanent holdings, and to manage active and semi-active records on

¹ Content DM, for example, is being proposed by the University of North Carolina at Chapel Hill, and other institutions such as the Universities of Saskatchewan and Waterloo, the Australasian digital theses program and the universities that take part in that program are using still other products.

² During the discussion following the presentation of this paper, a conference delegate indicated that the most recent version of D-Space has addressed only some of the display issues that I identify in this paper. The majority of my criticisms of the software were supported and confirmed by other delegates who are currently using D-Space in their institutions.

campus through the Archives' records management program. This software had been in use since 1996 and was also in need of replacement, though not urgently. Due to the complexity of our database requirements we recognized that selecting new software would be quite time consuming so we were open to other options for making the theses records available online.

At about the same time the idea of developing an Institutional Repository at the University of Calgary was gaining support amongst some upper level administrators. The Information Technology department of the University Library – which is in the same administrative division as the University Archives – had selected D-Space as the Institutional Repository software for the campus and was actively seeking participation in this initiative. It was suggested that migrating the theses data from its defunct database into D-Space would be an easy solution to the Archives' problem of replacing the public theses database quickly. At the same time importing 13,000 records into D-Space would help promote the institutional repository initiative by showing the university community what D-Space could do, and would potentially help to encourage the development of an electronic theses program on campus as well.

5. TESTING RECORDS IN D-SPACE

5.1 Migrating Records

Discussions amongst the Library's IT Department and the University Archives therefore quickly got underway to determine how to go about migrating records from the existing theses database into D-Space. D-Space required some tailoring by IT staff in order to add fields that we wished to retain from our former system, to rename existing fields to more closely reflect their content and to incorporate some data entry modifications that the Archives requested. Eventually about 450 theses records were migrated to D-Space to serve as a test collection on which we would examine the searching and reporting functionalities of the system. Our testing uncovered shortcomings in a number of areas that eventually led to our rejection of D-Space as viable database software for our theses collection.

5.2 Searching Issues

The most significant issues we found were related to the very limited search capabilities of D-Space. D-Space offers a basic search through which researchers can choose either keyword searching, or browsing by title, author or date. In testing the theses data we first attempted to search by date in an effort to pull up all the theses completed in a specific year. We require this information when reporting annually on the number of theses added to the collection for insurance purposes. However we found that the D-Space software does not allow a 'search' of a specific date to be made, but rather offers only a date 'browse'. When date browsing, we found that D-Space pulls up all theses in the database and highlights those that are relevant to the year requested. (See Figure 1)

An accurate count of theses for a particular year is therefore not given – rather, the count is of the entire database. In order to produce an accurate figure for our reporting purposes it would therefore be necessary for us to conduct a manual count of the items in the system for a specific year.

In an effort to circumvent the 'browse' issue, we conducted a basic search for the date "1960" in hopes of pulling up all the theses completed in that year. Unfortunately this technique resulted in numerous hits that were not relevant to our intended inquiry as shown in Figure 2.

D-Space does not clearly indicate to researchers what fields are being searched through its basic 'search' mechanism but it would appear to be conducting a keyword search of multiple fields. When we investigated the search results more closely we found that a number of items had '1960' in the subject fields, for example "Camus, Albert – 1913-1960", as shown in the detailed record for this 1978 thesis (See Figure 3).

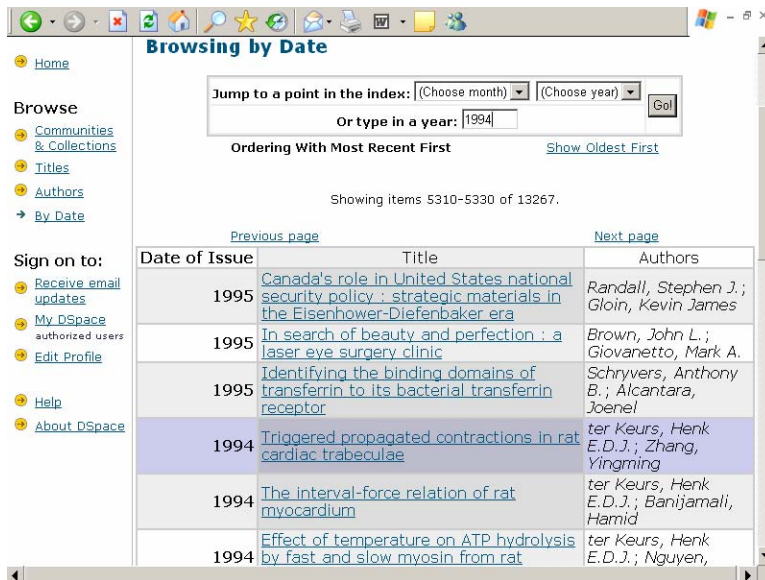


Figure 1: Browsing by Date.

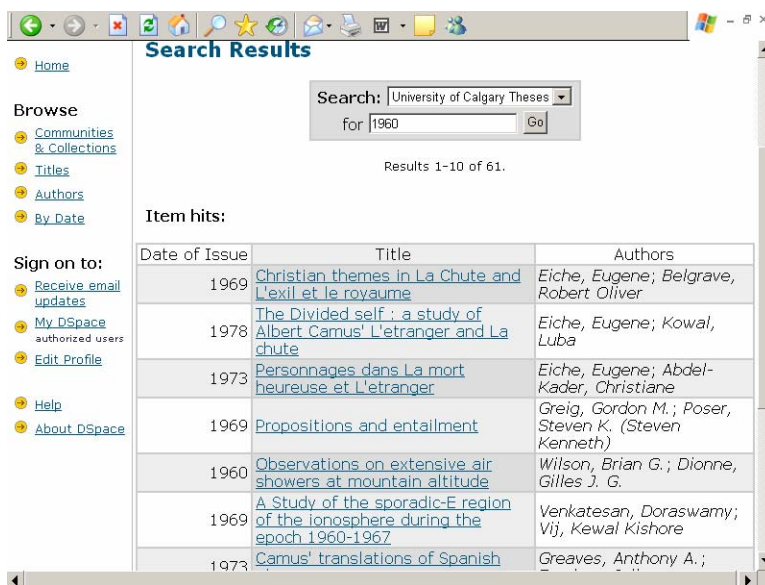


Figure 2: Basic Search results for "1960".

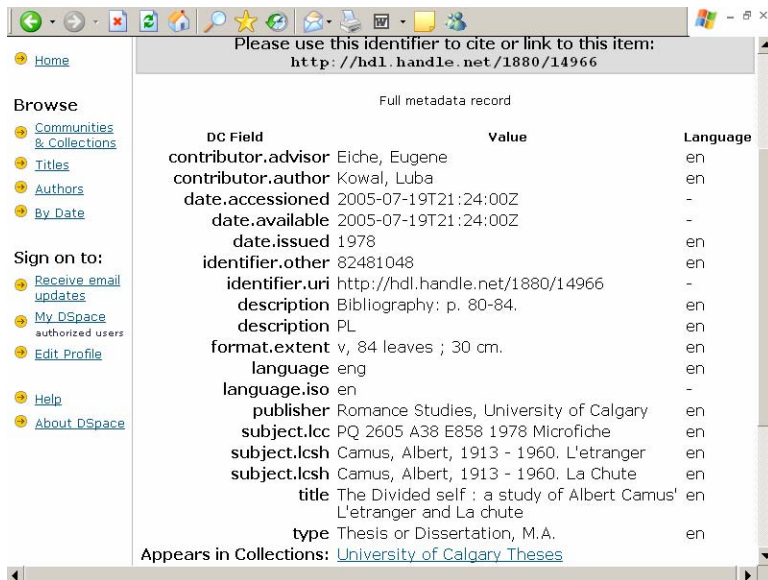


Figure 3: Detailed record for “1960” search result.

Again, we were unable to receive an accurate count of the number of theses that exist in the system for a specific year, and a manual count would be required in order to obtain the information we require for insurance reporting purposes.

D-Space also offers an advanced search which allows up to three fields to be searched at one time. This advanced search has nine fields to choose from, which are pre-set for the whole of D-Space and cannot be configured to suit individual collections within it. We found that several of these pre-set fields are not relevant to the theses records in our database, so we were left with only 4 fields to choose from for an advanced search, including ‘keyword’, ‘author’, ‘title’ and ‘subject’. Fields that are critical to effective searching of the theses collection were not offered, including ‘date’, ‘supervisor’, ‘department’ and ‘degree’. Consequently, although the advanced search does increase the viability of searching in D-Space by allowing more than one search term to be used, we found that it is not sophisticated enough to allow us to conduct searches which will satisfy our researchers’ needs and our own reporting requirements. I have a couple of examples to illustrate our difficulties.

Firstly, the University Archives reports annually to Petroleum Abstracts on theses pertaining to the upstream oil industry that have been added to our collection during the previous year. In order to do this we need to be able to search on two fields at a minimum and preferably three: a date (i.e. 2005) and a department (i.e. Geology and Geophysics, or History, or Law), a date and a keyword (i.e. “oil”, “petroleum” etc), or a date, a department and a keyword. Date and Department fields are not available in the advanced search tool. Using the keyword field to search for the desired date and department yields results that are not relevant to our intended inquiry, since the software hits on the specified keywords in fields other than date and department.

A second example: researchers frequently come to the Archives wanting to look at the theses that a certain department and/or supervisor has produced for specific years and, within that group of theses, for specific degrees and subjects. Such searches may require as many as four or five different fields to be searched at one time: i.e. Degree, Department, Year, Supervisor, Subject. D-Space will not allow this level of sophistication in searching – we cannot search on more than 3 fields and we cannot search on many of the fields we need to be able to in order to receive relevant hits. Four of these five fields are not included as options in the advanced search tool. We also found that D-Space would not allow us to search for a type of degree – either as a field specific search or as a keyword. Our attempts at searching

for “M.Sc.” and “Ph.D.” were unsuccessful (see figure 4) even though items with these degrees exist in the test collection (see figure 5).

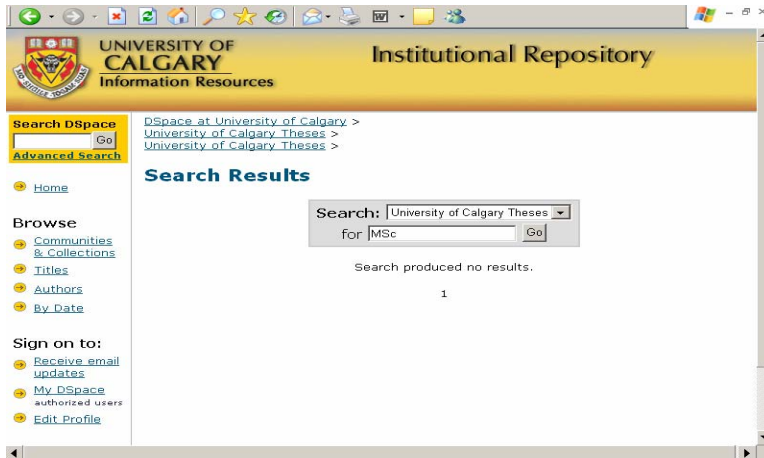


Figure 4: Searching for degree type.

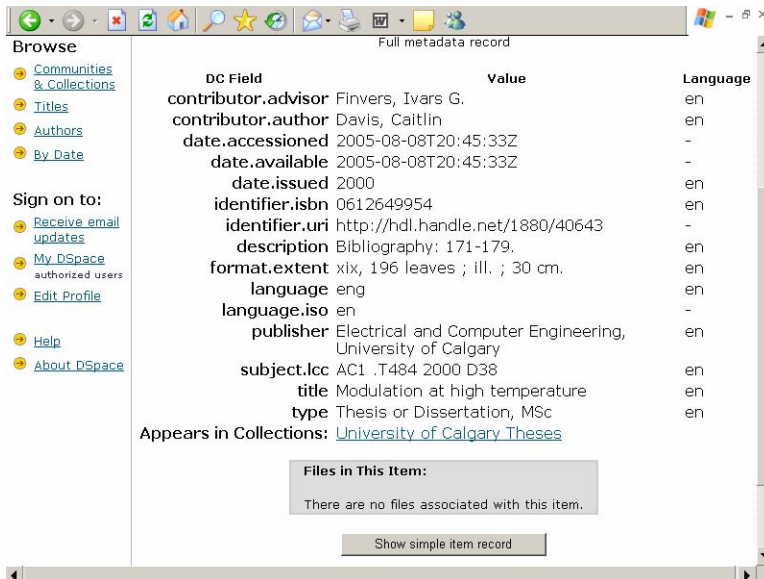


Figure 5: MSc thesis full metadata record.

5.3 Reporting Issues

Other problems we found with D-Space functionality related to the reporting capabilities of the software. In attempting to print the details of items in D-Space we found that the system does not allow us to conduct a search, choose a number of items from the hit list and then print a report containing details of the chosen hits. Instead, individual entries need to be clicked on to show the desired level of report -- either the short or long version -- and then printed off one at a time: a somewhat cumbersome process at best. Our researchers often want printed reports of relevant theses that they can take away with them: we have been able to provide this service in the past and would ideally want to retain that capability.

A second shortcoming in the reporting capabilities that we found is that D-Space will not allow us to choose selected entries from a search results list, save them and then e-mail them in a report to clients as an electronic attachment. We found that it is possible to save individual entries and attach those, but again this is a somewhat arduous process when dealing with numerous items at one time. Our office regularly

uses electronic reports when communicating with Petroleum Abstracts about potential additions to their records of our holdings, and also when responding to reference requests from clients. It is therefore very useful for us to be able to save relevant hits from a search in a single report – we were disappointed that D-Space will not allow this functionality.

We also found that it is not possible to move from the detailed report of one item on the hit list directly to the detailed report of the next item. Instead, it is necessary to drill down from the hit list to a report of each item, then retreat to the hit list and drill down again for details of the next item. This tends to be rather slow, particularly if one needs to consult the detailed report which is two screens beyond the hit list.

Our final complaint regarding D-Space’s reporting capabilities is that the software does not contain the option to sort the hits in a search results list. Our office has been able to sort search results by any field in our former system: i.e. Date, Supervisor, Author, Title etc. This is a useful function which allows us to group together records for researchers based on their interests and priorities. The lack of this function will prove arduous for researchers who wish to arrange the data to suit their requirements, particularly in cases where there are long search result lists.

5.4 Administrative and Display Issues

Our staff was also unhappy with the fact that D-Space does not allow both administrative and public views of the database. Our database contains information that is required for administrative use in-house, but which we do not wish to show in the public view – i.e. physical location, barcodes and notes on processing issues that are awaiting resolution. Ideally therefore, we require separate public and in-house views in order to administer the collection effectively while presenting clients only with the information they require to access the records they need. Our choice with D-Space would appear to be between two imperfects: either show data that the public does not need and perhaps ought not to have access to, or maintain a separate system containing the data we do not wish to make public.

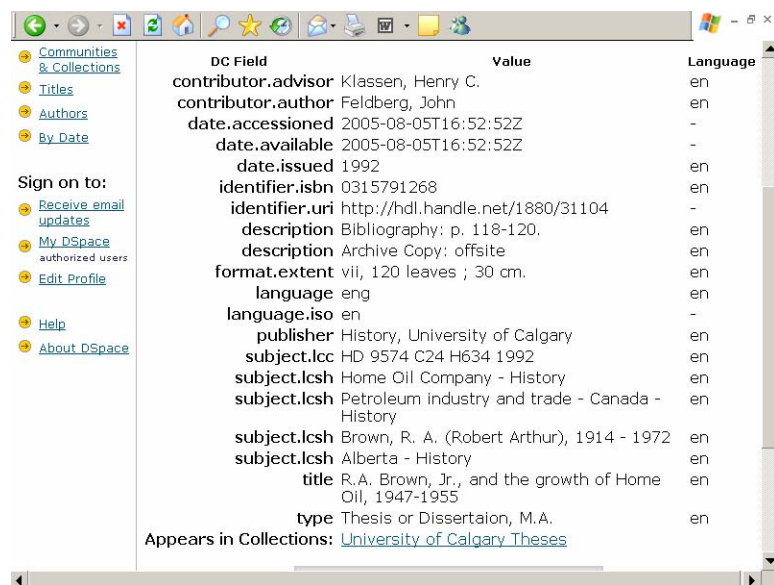


Figure 6: Full metadata record: unattractive, confusing and counterintuitive?

In general our staff felt that the display of information in D-Space is confusing and unattractive. This is particularly true of the full metadata report where there is a lot of information that users do not need to be

concerned with and fields that are oddly named and confusing to the layman (see figure 6). The data is also arranged in a manner that is almost counter-intuitive. In the full metadata record the title and degree type appear at the bottom of the page while less pertinent information about the item such as when the data was added to the system appears at the top of the report.

6. FINAL THOUGHTS ON D-SPACE

Finally, we were concerned that many of the problems we found with the functionality of the software might be expected to increase correspondingly as the volume of entries in D-Space increased. We were uncomfortably aware that the test collection contained only 450 theses. The theses database that was being considered for migration to D-Space had more than 13,000 titles. What impact would migrating the entire database and then adding 600 new titles annually have on the functionality of the program and our ability to effectively use it to manage our theses collection?

In reflecting on our findings, it strikes me that we were possibly being unfair in our analysis of the D-Space software – we were perhaps asking it to perform a role larger than it was designed to fill. However, even if we discount our issues with the lack of an administrative view and the limited reporting functions, D-Space still scores very poorly in its searching capabilities. Of all aspects to consider when selecting a database, the ability to effectively retrieve the information contained within it is surely the most important.

When the University Archives was offered D-Space as a “quick fix” to our desire to make our online theses database available quickly, we seized it too eagerly. With hindsight, we should perhaps have spent some time in careful consideration of how the theses collection could best be managed. If the move towards electronic theses is a permanent one, how will we who are charged with preserving these items effectively handle a large collection that exists only in paper form on the one hand, and an electronic theses collection which may or may not exist also in paper format on the other – both of which we wish to make publicly available? Above all there is clearly a need to ensure that there is no duplication of effort. Theses are already extremely time-consuming to process. Running parallel systems for accessing the paper and electronic theses is not a viable option.

In replacing its database programs, the University Archives selected a program that we feel will effectively manage both our paper and electronic theses.

7. THESES DATABASE SOFTWARE SELECTED AT UNIVERSITY ARCHIVES

As I indicated earlier, our former theses database formed part of a suite of archival databases which contained 3 modules: Theses, Archives and Records Management. The Archives has selected WebGencat for our new database system.³ The Library module of this program has been adopted for our theses database. It has not been fully tailored yet, but I will show you some screenshots from it anyway to give you a sense of how it looks and functions.

WebGencat allows both administrative (or librarian) and public displays, so we will be able to show only those fields we wish clients to see in the public display, keeping administrative information confidential for use in-house only. Figure 7 shows the various options that are available for searching in both the administrative and public views.

³ WebGencat is the creation of Eloquent Systems Inc. located in Vancouver, Canada (<http://www.eloquent-systems.com/>).



Figure 7: Librarian and Public view options

WebGencat will allow searches to be conducted on multiple fields at one time. Figure 8 shows the public template search screen and Figure 9 shows the public advanced search option.

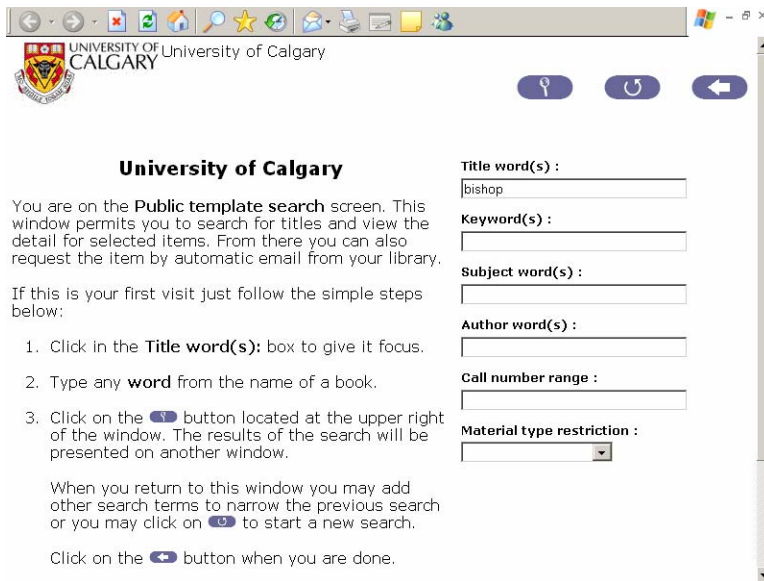


Figure 8: Public template search screen (untailored).

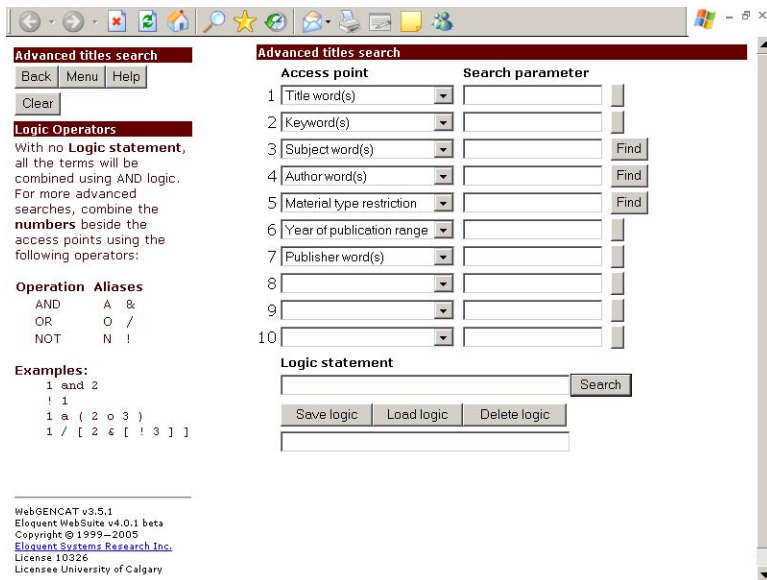


Figure 9: Public advanced search screen (untailored)

Figure 10 shows a list of hits in the public template search on the left side of the screen with details of a selected hit on the right. Multiple “hits” can be selected for printing in a single report or for saving in various formats, including pdf, to send electronically to clients.

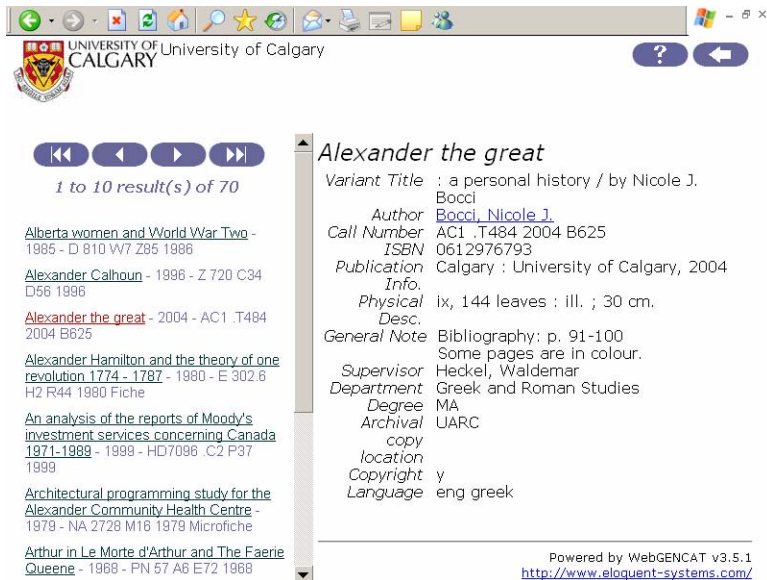


Figure 10: Search results – public template search

Figure 11 shows selected hits on the right and reporting options on the left – this is shown in the administrative view, but the public also has most of these options in the public view. Electronic documents can be linked to an entry so we will be equipped to offer users an online search tool with electronic copies of these attached.

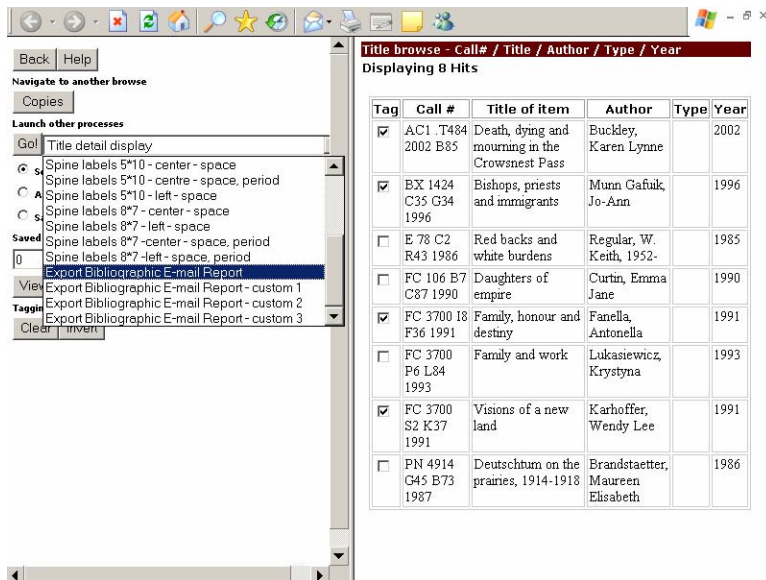


Figure 11: Administrative view: multiple hits selected for a report. Reporting options are on the left of the screen.

WebGencat is a more sophisticated database than D-Space and it more closely meets the needs of the University Archives in all aspects of its functionality. It offers us more control over the data in the database in all areas of administration, reference and reporting. We also feel that the display is more appealing and user-friendly than we found D-Space to be.

We hope to have the software fully tailored in the near future and our public database back online.