

# A PROTOTYPE FOR THE FEDERATED WORLD DIRECTORY OF MATHEMATICIANS

JONATHAN M. BORWEIN, MASON MACKLEM AND JAEHYUN PAEK

ABSTRACT. In 1998, the International Math Union asked its Committee on Electronic Information Communication (CEIC) to consider an electronic World Directory of Mathematicians to replace the traditional hard-copy version. The CEIC concluded that intellectual property and privacy issues across various countries made such a directory infeasible for the 2002 edition. In 2004, the IMU endorsed moving ahead with a federated search protocol. We describe a prototype, the Federated World Directory of Mathematicians (FWDM), where a common interface searches and retrieves information online from national mathematical society directories, with no additional work for the user and no single combined directory. We also discuss some of the IP and copyright issues preventing a combined directory.

## 1. BACKGROUND

The International Mathematical Union (IMU) is a non-governmental and non-profit scientific organization that oversees the promotion and development of mathematics research throughout the world. The IMU has a wide range of responsibilities, including helping to improve mathematical education in developing countries, and sponsoring lectures and international meetings. These responsibilities are met partly through a number of IMU Commissions, including the Commission on Development and Exchange (CDE), the International Commission on Mathematical Instruction (ICMI) and the Commission on Electronic Information and Communication (CEIC).

The most public responsibility of the IMU is the organization of the International Congress of Mathematicians (ICM) every four years. This meeting includes presentations on the frontier of mathematical research, as well as the awarding of the Fields Medals and the Nevanlinna Prize. Timed to coincide with the ICM is the publication every four years of the World Directory of Mathematicians (WDM) by the IMU and the American Mathematical Society (AMS); the goal of the WDM is to list all active research mathematicians throughout the world.

Data collection for individual mathematicians for inclusion in the WDM is not performed directly by the IMU. The IMU does not have individual memberships; instead, its members consist of either national mathematical societies or national academies of science, with each member nation required to uphold standards of mathematical research. The IMU currently has 66 member nations, and delegates from each member nation form the IMU General Assembly, which meets at the ICM every four years. Data for the WDM is collected by each of the national societies, and is provided to the IMU for inclusion in the published list.

In 1998, the International Math Union asked the CEIC to consider an Electronic World Directory of Mathematicians to replace the traditional hard-copy version.

However, the CEIC concluded that a centralized electronic database would be subject to intellectual property and privacy laws on digital information, laws which vary across different countries; as such, the conclusion was that such a directory would be infeasible for the 2002 published edition.

In 2004, the IMU endorsed moving ahead with a federated search protocol. Federated searching connotes any system that provides a common user interface for searching and retrieving information across heterogeneous datasets over the Internet. Using a federated search model, a centralized database would be unnecessary, as this data could be gathered by combining the information retrieved separately from each member nation's membership list; this avoids the need for caching database entries, which raises a host of legal and privacy issues.

We describe a prototype, the Federated World Directory of Mathematicians (FWDM), where a common interface searches and retrieves information online from national mathematical society directories, with no additional work for the user and no single combined directory. We also discuss some of the IP and copyright issues preventing a combined directory.

## 2. PARTNER WEBSITES

A federated search engine provides a single user interface for input, which we will refer to as the *parent search engine*. The parent sends this input to a number of different *child search engines*, with each parameter formatted to make it compatible with the search parameters of each child search engine. The resulting output from each of the separate search engines is then returned to the user in a single combined output. In our prototype, the child search engines are those built by each member nation to search their respective membership lists, and the combined output consists of a list of links to the individual members' listings on their respective member nations' membership page. The prototype interface, with sample output, is shown in Figures 1 and 2.

The goal of developing a prototype for the FWDM is twofold: to explore the feasibility of constructing the combined membership list by combining information from the member's society databases; and to encourage standardized search engines for each across member nation's membership lists. As such, we have included only the following subset of the IMU's member nations for inclusion in the prototype:

- **American Mathematical Society's Combined Membership List (CML)**: The CML combines the membership lists of the American Mathematical Society (AMS), the Mathematical Association of America (MAA), the Society of Industrial and Applied Mathematics (SIAM), the American Mathematical Association of Two-Year Colleges (AMATYC), the Association for Women in Mathematics (AWM), and the Canadian Mathematical Society (CMS) - Société Mathématique du Canada (SMC). This list is updated daily.
- **Canadian Mathematical Society (CMS)**: Despite information on members being included in the results from the CML, the data included differs in some respects, such as the need to be fully bilingual, and as such both the CML and CMS search engines have been included.
- **Deutsche Mathematiker-Vereinigung (DMV)**: Membership page for the German National Mathematical Society.

- **National Committee for Mathematics** (NCM): List of Australian mathematicians collected in 2001 for inclusion in the WDM.
- **Société Mathématique de France** (SMF): Membership page for the French Mathematical Society.

A number of member nations also have search capabilities for their membership lists, but have not been included in the prototype for various reasons. We discuss these societies in a later section.

### 3. BASE SEARCH FIELDS FOR ALL MEMBER SOCIETIES

The prototype of the FWDM has two versions of the central interface: a *Simple Search interface*, which contains textboxes for fields common to all member society databases; and an *Advanced Search interface*, which contains input fields for information available in any member society database. The Simple Search interface currently contains fields for the researcher's First Name, Last Name, Employer/University, and Country; a screenshot of the interface is shown in Figure 1.

The Simple Search interface is the default shown when the FWDM is requested, as the working assumption is that the Advanced Search is useful primarily when a more narrow search, with more search criteria or less potential for ambiguity, is being performed. The intention with the Simple Search interface is that some additional fields will be added as more consistent databases are used by member societies, with the end goal being a standardized set of information collected by all societies. Presently, the base information required for a member society to be included are First and Last Name; this excluded a number of member societies from inclusion in the prototype, a situation that will be discussed in a later section.

### 4. MEMBER SOCIETY SEARCH ENGINES

The structure of the FWDM search engine is as follows:

- (1) Collect user-specified search parameters from the FWDM search interface.
- (2) Translate the FWDM variables to the corresponding variables for each member society's search engine.
- (3) Perform a separate search for each member society, by sending their respective search engine a URL containing the translated variables corresponding to the user's original FWDM input.
- (4) Collect the HTML output from each member society's search engine. Parse the HTML to identify and separate each individual listing returned, and create an array of returned names for each member society.
- (5) Once all member societies have returned their results, combine the arrays of individual listings into a single array, removing duplicates.

So for each member society, the key is to identify which fields can be searched directly via their search engine's URL, and determine how to isolate the HTML corresponding to each individual's name. Notice that this entire model is dependent on the specific HTML used by each society; as such, any changes or reformatting by a member society will remove their listings from the combined search results. Notice that this problem will be removed once each member society has a standardized appearance. In fact, one goal of the FWDM prototype is to encourage member

Dalhousie Distributed Research Institute and Virtual Environment

CECM | SFU CoLab | WestGrid | Faculty of Computer Science | DCRI | Experimental Mathematics | DocServer | IRMACS

D-Drive Home > FWDM > Query Form

Home

News

Seminars and Meetings

Research Projects

Technologies

Partners

FWDM

Contact Us

### Federated World Directory of Mathematicians Query Form

To Advanced Search

**Search**

First Name

Last Name

Employer/University

Country

Institution  Home  
(Checkbox options for CML only)

**Search**

Guide to Using the FWDM Query Form

#### Description

Federated searching is a system that provides a common user interface for searching and retrieving information across heterogeneous datasets over the Internet. This query form allows any user to find mathematician(s) in different mathematical societies' directories. The major difference between this version and the previous version is in the way results are returned. This program outputs three tables. The first table shows the user's query. The second table lists the results by name with each name having links to webpages that pertain to the corresponding name from each directory. The third table lists the number of results from each directory.

This version searches through the following directories: Combined Member's List (CML), Canadian Mathematical Society (CMS), French Mathematical Society (SMF), MathNet and Zuse Institute Berlin (ZIB)

p. 902.494.2501 · f. 902.492.1517 · Faculty of Computer Science · Dalhousie University · 6050 University Avenue · Halifax · NS · B3H 1W5 · Canada

FIGURE 1. FWDM Interface

societies to adopt a standardized appearance, and as such this shortcoming with the initial FWDM execution model has been embraced as a feature!

For each member society, we show sample output from their respective search engines, and discuss how information used by the FWDM is extracted from the output HTML. We also outline changes that can be made to improve both the search capabilities of the FWDM and the member society.

The screenshot shows the Dalhousie Distributed Research Institute and Virtual Environment (D-Drive) website. The header includes navigation links: CECM | SFU CoLab | WestGrid | Faculty of Computer Science | DCRI | Experimental Mathematics | DocServer | IRMACS. The breadcrumb trail is D-Drive Home > FWDM > Query Form. A left sidebar contains a menu with items: Home, News, Seminars and Meetings, Research Projects, Technologies, Partners, FWDM (highlighted), and Contact Us. The main content area displays a search form with fields for 'Your Query', 'First Name: borwein', 'Last Name:', and 'Country:'. Below the form is a table of results:

	Name	Society
1	Borwein, Dr. Peter B.	CMS
2	Borwein, Dr. Jonathan M.	CMS
3	Borwein, Dr. David	CMS

Below the results table is a section titled 'Directory Number of results' with the following data:

CML: 0 Results Found!
CMS: 3 Results Found!
SMF: 0 Results Found!
MathNet: 0 Results Found!
ZIB: 0 Results Found!

FIGURE 2. FWDM Interface

4.1. **CML.** On the search page for the American Mathematical Society's Combined Membership List, one can search for a member by filling out any of the following fields: last name, first name, position, state (U.S.), country, member organization, institution, institution city, institution state/province and institution country. Not all fields are text-areas: each member organization is selected using check-boxes, while position, state, country, institution state/province and institution country fields all are selected from drop-down menus. Notice that state/province drop-down lists include both Canadian provinces and American states.

Sample output from the CML search engine is shown in Figure 3. For each returned listing, there is information returned that is not available in the CML search interface, specifically: institution address, home and office phone numbers, fax number, email address, homepage URL, and research specialties (using AMS numerical subject classifications). The first and last name of each individual returned is isolated by parsing the HTML to find the number of results returned<sup>1</sup>, and create an array of this length; the HTML is then parsed to find each bold text-string, as the individual's name is the only information displayed in bold text.

4.1.1. *Future Work.* One of the search criteria available on the FWDM Advanced Search page is number of entries displayed, up to a maximum of 20. This value of 20 for the maximum is inherited from the search results for the CML, which groups multiple results together alphabetically in groups of 20; for example, if 43 results

<sup>1</sup>In the case where only one item is found, the number of items found is printed as `Items: 1-1 of 1`. As such, this case is entirely equivalent to the case where multiple items are found.



AMS | MAA | SIAM | AMATYC | AWM | CMS/SMC

## CML Search Results

[Search the CML](#) || [Update CML](#) || [Directory of Institutions](#) || [Help](#)

Matches for: Last Name(Borwein)



Items: 1-3 of 3

### **Borwein, David**

Prof Emer

University of Western Ontario  
Department of Mathematics  
London, ON N6A 5B7 Canada

Phone: Office: 519-661-2111 ext 86884, Fax: 519-661-3610, Home: 519-672-8057

Email: [dborwein@uwo.ca](mailto:dborwein@uwo.ca)

Memberships: AMS, MAA, CMS/SMC

Specialties: 26, 33, 40, 44, 46

### **Borwein, Jonathan M (Dr)**

Prof

Dalhousie University  
Faculty of Computer Science  
6050 University Avenue  
Halifax, NS B3H 1W5 Canada

Phone: Home: 902 412-1228, Office: 902 494-2501, Fax: 902-492-1517

Email: [jborwein@cs.dal.ca](mailto:jborwein@cs.dal.ca), [jborwein@cecm.sfu.ca](mailto:jborwein@cecm.sfu.ca)

Url: <http://www.cs.dal.ca/~jborwein>

Memberships: AMS, MAA, CMS/SMC

Specialties: 11, 41, 46, 49, 90

FIGURE 3. Sample output from the American Mathematical Society, multiple results from Last Name = Borwein

are found, they would be displayed on three pages, separated into results 1-20, 21-40 and 41-43. Since the CML search engine does not have a field for the number of search results to display, the separation into 20 listings places a hard bound on the number that can be displayed when combined. There is a workaround if more than 20 listings are desired: the index of the first result displayed is contained in the URL for the resulting CGI page, so if the FWDM engine detects that more than

20 results are found, it could send requests for the remaining search results via the URL, and retrieve them as with the initial 20 results.<sup>2</sup>

However, there are two problems with this method that would need to be resolved first to make this truly feasible:

- *Multiple executions*: Currently, the CML search engine re-executes the requested search to go from one set of 20 results to the next. For example, if you run a search with 43 results, then by following the link to the next page of results, the entire search is executed again to print the next 20 entries. A possible workaround would be to have a blanket CGI program that logs the current client's IP address, and stores the entire search results in a single string, thus maintaining state for each client. This CGI program would run on the CML server, and thus would not be part of the FWDM search tool.
- *Excessive search results*: The CML search tool prints the total number of results found as part of the search results; however, the user *can only access the first 100 results!* The CML search tool enables scrolling between pages of results, with 20 results per page, but only up to a maximum of 5 pages of results. This is even true for attempts to access the results via the URL: the search tool uses a URL-parameter called **counter** to specify the first entry on each page, but values of more than 100 return the page of results with entries 81-100. As such, a hard-bound of 100 printed results exists for the FWDM page, at least until this bound changes on the CML search tool.

In both cases, the solution would require changes on the CML search engine, and as such cannot be resolved entirely by the parent search engine.

4.2. **CMS**. The search engine for the Canadian Mathematical Society provides five fields for searches: Name, Employer/University, Interests, City, or Country. However, the search engine provides only one textbox to enter a search keyword, and the user selects the specific field to search by selecting a radio-button choice from the five possible fields. As such, only one field can be searched at a time, and thus the user cannot refine searches based on the results from a previous search.

Of all the child search engines included in the prototype, the CMS search engine is the only one with different formatting for the cases where single and multiple results are returned. When more than one result is found, a list of links to each individual member listing is returned, along with summaries containing Employer and Address fields for each member. When a single member is found, the search automatically redirects to the individual's personal page; the resulting webpage prints fields for Telephone, Fax, Home Page, Email, and Fields of Interest. A sample search on the CMS search engine is shown in Figures 4 and 5.

Two observations about the case when a single result is returned:

---

<sup>2</sup>One goal of the FWDM project is for the acceptance of a federated search engine to impact on the design of the component search engines. For example, in the Summer of 2004, the CMS search engine was designed to perform AND searches by default when multiple search criteria was entered, and to perform an additional OR search when no results are found. One result was that if a user entered *David Mumford*, for example, the search engine returned all *David's* within the database, which is obviously not what the user intended. Such bugs can be justified as features when the search engine exists independently, but can be avoided by adopting a more uniform search engine design across the various member societies.

- *Automatic generation*: The links returned when multiple listings are returned are simply calls to the search engine again with the text string for each member's entire name. This assumes no redundancy within the database, i.e. no common names to multiple members. In addition, this requires an additional call to the search engine to isolate a single member's listing from the multiple members returned initially, despite much of the member's information (Name, Employer, Address) already being returned from the first call.
- *Search fields vs. Information fields*: The structure of the CMS database cannot entirely be reconstructed from the information printed on individual member's listings. In the CMS main search page, the fields that can be searched are the member's Name, Employer, Interests, City, and Country. Despite this, the member listing returned when a single member is found contains additional fields, such as Address and Telephone, that are not searchable by the CMS search engine. In addition, the fields for two of the search fields, City and Country, are not printed separately on the member's individual listing. However, searching for Ontario in either the City or Country field returns no results; this indicates that these fields are not simply searches for substrings of the Address field for each member, and instead exist separately within the CMS database.

4.3. **DMV**. The search engine for the Deutsche Mathematiker-Vereinigung currently is integrated into the PERSONA MATHEMATICA, designed and driven by the Math-Net group of the Mathematical Institute / University of Cologne. This engine searches for more than 100 mathematical websites in Germany and Austria.

PERSONA MATHEMATICA has two search modes, *Standard* and *Advanced Search*. For the Standard Search option, four sources of information are available, according to the web-site's documentation:

- Math-Net metadata (used by default): The mathematics department for each university within the DMV potentially has a Math-Net compatible homepage. Each Math-Net page would contain Dublin Core-compatible metadata for every member of each department; this metadata is then gathered automatically by the PERSONA MATHEMATICA, and is included in the search results. This option has the advantage of including more structured information with standardized data-structures, and therefore the search criteria is easier to specify for the user. However, the quality of the search results using this option is entirely dependent on the quality of the metadata provided by the departments; in particular, departments without a Math-Net-compatible homepage would not be included in the search results.
- Department Member lists (used by default): This option performs separate searches from the department directories for each institution included within the DMV. As with the Math-Net metadata, this option depends on the quality of information provided within each directory; however, since this information is displayed on each department's homepage, presumably this information is more accurate than the Math-Net metadata, and thus this problem is likely negligible. A larger disadvantage is the fact that, since

The screenshot shows the website interface for the Canadian Mathematical Society (CMS). The top navigation bar includes a language selector for 'français', a home button, and links for 'about the CMS', 'media releases', 'search', and 'site map'. The main content area features a search bar with the text 'Borwein' and radio buttons for 'Name', 'Employer/University', 'Interests', 'City', and 'Country'. Below the search bar is a table with three columns: Name, Employer, and Address. The table lists three results for 'Borwein':

Name	Employer	Address
<a href="#">Borwein, Dr. David</a>	University of Western Ontario	Department of Mathematics University of Western Ontario Middlesex College, London Ontario, Canada N6A 5B7
<a href="#">Borwein, Dr. Peter B.</a>	Simon Fraser University	Department of Mathematics Simon Fraser University 8888 University Drive, Burnaby British Columbia, Canada V5A 1S6
<a href="#">Borwein, Dr. Jonathan M.</a>	Dalhousie University	Faculty of Computer Science Dalhousie University 6050 University Avenue, Halifax Nova Scotia, Canada B3H 1W5

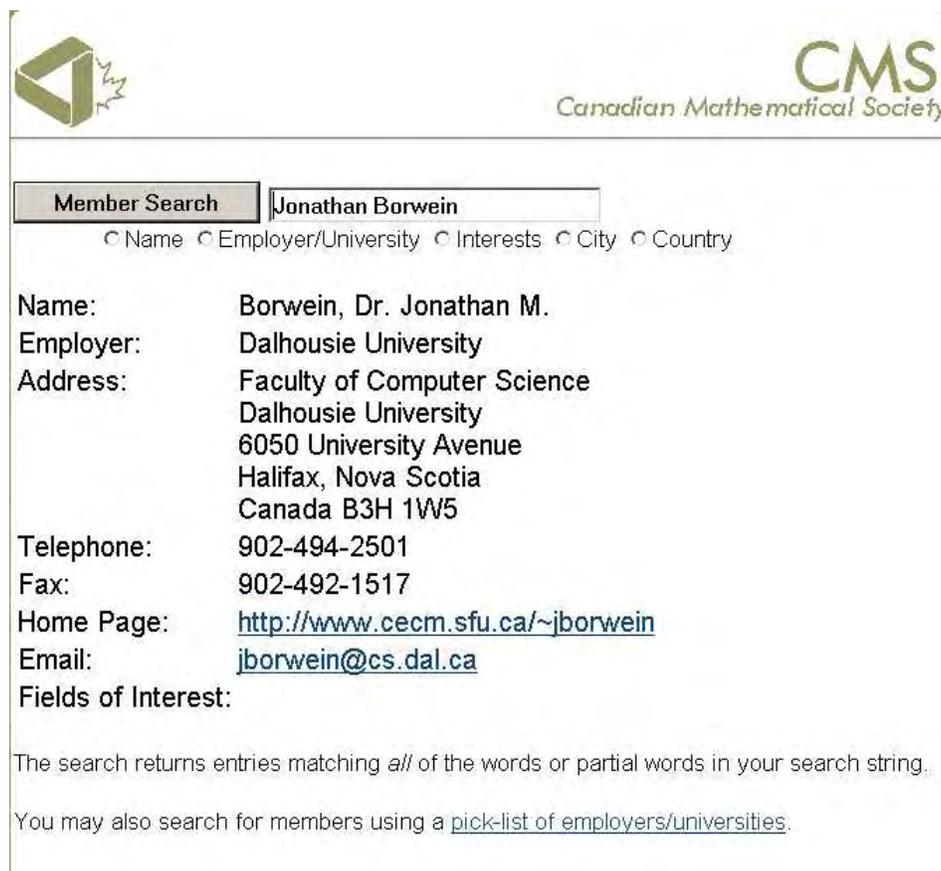
Below the table, there is explanatory text: 'The search looks first for entries matching *all* of the words or partial words in your search string. If that fails, it returns entries matching *any* of the words.' and a link: 'You may also search for members using a [pick-list of employers/universities](#).' The footer contains 'Copyright © CMS - SMC' and a 'comments?' link.

FIGURE 4. Sample output from Canadian Mathematical Society, with multiple results (input Last Name = Borwein)

each department is searched separately, including this option increases the time required to return the search results. This problem becomes even more severe within a federated search engine, as it slows down the response time for the entire search process.

- DMV database and Educational Math list (not used by default): Descriptions of the two non-default sources are not provided by the PERSONA MATHEMATICA website, and we have been unable to find a search that successfully returns information from either.

For the Advanced Search option, only metadata is used, and thus the other three sources of information are unavailable; the assumption seems to be that as the number of Math-Net departmental pages increases, the metadata will become the most accurate source of information, and the other three sources will become obsolete. The same problems with the metadata exist in the Advanced Search as existed with the Standard Search, namely that the quality of the search results is entirely dependent on the quality of the existing metadata. However, this is currently much more



The screenshot shows the CMS (Canadian Mathematical Society) member search interface. At the top left is a logo consisting of three green cubes. At the top right is the CMS logo with the text "Canadian Mathematical Society". Below the logo is a search bar with the text "Member Search" and a search input field containing "Jonathan Borwein". Below the search bar are radio buttons for "Name", "Employer/University", "Interests", "City", and "Country". The search results are displayed in a list format with the following details:

**Name:** Borwein, Dr. Jonathan M.  
**Employer:** Dalhousie University  
**Address:** Faculty of Computer Science  
 Dalhousie University  
 6050 University Avenue  
 Halifax, Nova Scotia  
 Canada B3H 1W5  
**Telephone:** 902-494-2501  
**Fax:** 902-492-1517  
**Home Page:** <http://www.cecm.sfu.ca/~jborwein>  
**Email:** [jborwein@cs.dal.ca](mailto:jborwein@cs.dal.ca)  
**Fields of Interest:**

Below the search results, there is a note: "The search returns entries matching *all* of the words or partial words in your search string." and a link: "You may also search for members using a [pick-list of employers/universities](#)."

FIGURE 5. Sample output from Canadian Mathematical Society, with single result; (input Last Name = Borwein, First Name = Jonathan)

of a problem than it will be in the longterm, and in the meantime more flexible searches are available using the Dublin-Core-compatible metadata.

The Advanced Search has three search fields: Names, Fields of Interest, and Keywords. This option also includes additional flexibility to exclude words and search for a collection of words. In addition, when no results are found, the Advanced Search performs additional searches using sound extension, by removing all vowels and performing a search on the resulting substrings; this process often solves problems resulting from the presence or absence of accents in the user's input. The Metadata search returns a member's name, email address, address, phone number, fax number, research interest, position/task, and a link to a website.

The FWDM uses the Standard Search, as the documentation on the PERSONA MATHEMATICA web-site recommends using this option for simple name searches. We initially included both the Math-Net metadata and the Department Pages in the FWDM combined listings. However, at the testing stage we decided that the Departmental results were sufficiently slow as to justify removing them from the combined results. Sample output using only the metadata is shown in Figure 6.

7 Math-Net Metadata result(s)

metadata:1	
<b>Name:</b>	<b>Achterberg, Tobias</b>
<b>Email:</b>	<b><a href="mailto:achterberg@zib.de">achterberg@zib.de</a></b>
<b>Address:</b>	<b>Takustr. 7, D-14195 Berlin-Dahlem, Germany</b>
<b>Phone:</b>	<b>++49 (0) 30 84185 301</b>
<b>Fax:</b>	<b>++49 (0) 30 84185 269</b>
<b>Research area:</b>	
<b>Position/Task:</b>	<b>Mitarbeiter</b>
<b>More...:</b>	<b><a href="http://www.zib.de/personal/personal.pl?language=EN&amp;name=achterberg">href="http://www.zib.de/personal/personal.pl?language=EN&amp;name=achterberg"</a></b>
metadata:2	
<b>Name:</b>	<b>Gaenzler, Tobias</b>
<b>Email:</b>	<b><a href="mailto:gaenzler@zib.de">gaenzler@zib.de</a></b>
<b>Address:</b>	<b>Takustr. 7, D-14195 Berlin-Dahlem, Germany</b>
<b>Phone:</b>	<b>++49 (0) 30 84185 366</b>
<b>Fax:</b>	<b>++49 (0) 30 84185 107</b>
<b>Research area:</b>	
<b>Position/Task:</b>	<b>Mitarbeiter</b>
<b>More...:</b>	<b><a href="http://www.zib.de/personal/personal.pl?language=EN&amp;name=gaenzler">href="http://www.zib.de/personal/personal.pl?language=EN&amp;name=gaenzler"</a></b>
metadata:3	
<b>Name:</b>	<b>Gayer, Tobias</b>
<b>Email:</b>	<b><a href="mailto:Tobias.Gayer@Math.Uni-Augsburg.DE">Tobias.Gayer@Math.Uni-Augsburg.DE</a></b>
<b>Address:</b>	<b>Universitaet Augsburg, Institut fuer Mathematik, D-86135 Augsburg</b>
<b>Phone:</b>	<b>++49 821 598-2244</b>
<b>Fax:</b>	<b>++49 821 598-2339</b>
<b>Research</b>	

FIGURE 6. Sample output from DMV search engine (input name = Tobias)

Unlike in the CML and CMS pages, the Name field is not formatted differently from the rest of the individual listing, as the entire listing is in bold. However, the listing is formatted as a table, and thus we are able to isolate each name by checking for the string *Name* in the first column of the table. Since all of the results are combined into a single output page, for DMV entries in the combined output on the FWDM page, we link the individual's name to the result found by searching on their entire name, thus returning only their individual listing.

4.3.1. *Future Work.* The primary next step for improving the search results from the DMV search engine within the FWDM is providing additional search options for accents. One of the advantages of the Advanced Search option is the sound extension searches, which are performed when no results are found. These searches remove all vowels from the user input and the source information, and search the

resulting substrings. One of the primary advantages of sound extension searches is avoiding problems with not finding results due to the presence of accented characters. Since the FWDM uses the Standard Search option for the DMV, we lose access to the sound extension searches. One obvious solution would be to use the Advanced Search option instead of the Standard Search, and we plan on doing this once the metadata available to the Advanced Search improves.

Another possible solution would be to perform multiple searches using the Standard Search, with accented and unaccented characters when no results are found. This allows us to simulate the sound extension searches using only the FWDM code, and thus without modifying the PERSONA MATHEMATICA code. This solution would require a hash table of characters that are commonly accented, or combinations of characters that are often used in place of accented characters; such a hash table would be easier to generate with access to the DMV database, but would be easy enough to create directly from the Math Directory lists, which are accessible locally to the FWDM code without raising any legal or privacy issues. A disadvantage of this solution is that it would require performing more searches for each FWDM search, and thus add more traffic to the DMV search engine than would otherwise be the case.

Another feature that we would like to add would be allowing the user the option of including the Math Department information source into the search results, using the PERSONA MATHEMATICA Standard Search option. This would not be the default setting, and would only be present in the FWDM Advanced Search form.

4.4. **SMF.** The searchable membership index for the Société Mathématique de France (SMF) has identical English and French versions; a sample output of the English version is shown in Figure 7. For either version, the user can choose between searching for an individual's name or their institution; this choice is made using a radio button to select which directory should be searched, with available selections for the Persons' Directory and the Institutions' Directory.

For the Persons' Directory, available fields consist of the individual member's First Name and Last Name; if only one name is entered, the search engine assumes that the name is the individual's first name. Available fields for the Institutions' Directory consist of the institution's name, city or country. The resulting output displays the member's First and Last Names, their email address, fax and phone numbers, and home and institution addresses. As with the CMS, additional information is displayed that is not searchable; for example, no results are found when the street address is searched in the Institutions' Directory. When multiple results are found, all information associated with each result is displayed on the same output-page.

Since the First and Last Names are both displayed in bold-type, the `<b><>` HTML tags for each name are used to isolate the member's full-name. Unlike the other search engines, which also display information for multiple results on the same output page, the SMF output page does not display the number of results found, complicating the process of creating an array to store the names of each member found; currently, the array is expanded automatically to include each result as the code parses through the HTML source for the output page, so this problem is not a major issue.

The "emath.fr" domain: [ACM](#) [Directory](#) [MaTeXo](#) [MathDoc](#) [Jobs](#) [SMAI](#) [SMF](#)


[News](#)
[Membership](#)
[Running of the Society](#)
[Education](#)
[Forum](#)  
[Site map](#)
[Directory](#)
[Publications](#)
[Employment services](#)
[Math on the web](#)  
[Site search](#)
[The authorities](#)
[Catalogue & orders](#)
[Miscellaneous](#)

## Members directory

The use of this directory is ruled by the french law "[Informatique et liberté](#)" (January 6, 1978).

It is prohibited to copy personal data contained in this directory to enrich an other data base particularly in view to draw a commercial or an advertising profit.

After the 34th article of the law, individuals mentioned in this directory have a right of access to alter, correct or suppress data which concerns them personally. To exercise your right you may contact the SMF for further information ([smf@dma.ens.fr](mailto:smf@dma.ens.fr)).

Some data of this directory (name and professional address of the SMF members) are at the disposal of the [French mathematical community's directory](#).

---

Persons' directory (enter name and surname)  
 Institutions' directory (enter name, city or country)

[ [Look for "bouscaren" in the French mathematical community's directory](#) ]

---

Name: **BOUSCAREN**  
 First name: **Elisabeth**  
 Email: [elibou@loquique.jussieu.fr](mailto:elibou@loquique.jussieu.fr)  
 Fax: +33 1 44 27 61 48  
 Phone: +33 1 44 27 37 68  
 Address: Université Paris VII  
 CNRS, UFR de Math, Case 7012  
 2 place Jussieu  
 75251 PARIS CEDEX 05  
 FRANCE  
 Home address: 95 avenue du General Leclerc  
 75014 PARIS

FIGURE 7. Sample output from SMF (input `bouscaren` using Persons' directory)

4.4.1. *Future Work.* The structure of the SMF search engine provides some problems with regards to finding results that would be found using the other search engines. For example, using the CML and CMS search engines, a user can enter a member's First Name only and get a list of all member's of that society with that first name. This option is unavailable using the SMF search engine, since when only a single name is entered rather than a fullname, the search engine assumes that the single name is the member's last name. Notice that the absence of this functionality is especially problematic when the user is unsure whether a member's last name contains accented characters. We will discuss other options for searching in the presence of accented characters in a later section.

Another absent functionality is the ability to search simultaneously using both Name and Institution information. For example, if a user gets multiple results when searching on a member's name, they might want to narrow the search down by searching for the member's institution from among the results returned from the

member's name. However, under the current search engine's design, this functionality is unavailable. Note that neither of these functionalities can be easily simulated by performing additional searches on the FWDM server; both issues need to be resolved within the SMF search engine software.

**4.5. NCM.** A number of national mathematical societies maintain directories, but not in a searchable form (or at least not with any internal structure or metadata). The Australian Mathematical Society does not have a search engine for its membership list. However, in 2000 a list was created of all working mathematicians in Australia, and the resulting list was posted online on a single HTML webpage, with the affiliations and addresses of each mathematician; no search engine front-end was provided.

For inclusion in the FWDM, we created a server-side search engine for this single HTML page, a sample of which is shown in Figure 8. Note that the name of each member is in bold type, distinguishing it from the rest of each individual's listing; we use the `<b></b>` HTML tags around the name as the delimiters for including the name in the FWDM output. When an individual on the NCM list is found by the FWDM, the output on the combined FWDM result page is a link to an anchor on the NCM HTML page directed to the beginning of the list of names starting with the letter of the given individual. This is clearly not ideal, being only an approximate location of the individual's information, but this is as close as possible given the limited structure contained within the HTML page.<sup>3</sup>

Given the absence of a search engine for this list, the design of a search front-end to be run on the FWDM server created some difficulties. For each search, we begin by reading the entire HTML source for the page listing all members. Since this HTML page has not changed since its initial publication, this step seems unnecessary, especially since it must be repeated for every FWDM search. However, the entire design of the FWDM was motivated by complying with IP and privacy laws, specifically by not transferring personal data across countries; as such, keeping a local copy of this directory would specifically undermine this entire process. However, this pushes the compliance issue forward a step, since it is unclear what the implications are with regards to these laws for copies existing in temporary memory for the purposes of isolating individual names, since this is still technically a local copy of this information. In the meantime, we are assuming that this will not be a major issue.

**4.6. Non-Conforming Search Engines.** One of the goals of the FWDM project is to encourage national mathematical societies without a searchable membership list to create such a search engine; the assumption is that the ability to have their membership lists combined within the Federated membership list, or at least the possibility of their members being absent from the federated list, would serve as encouragement to create a searchable membership list.

In this direction, a number of societies currently have a membership list available online, but for various reasons are not able to be included in the FWDM results.

---

<sup>3</sup>Note that, given the fact that each person's name is in bold, the entire list is alphabetized, and the anchors are directed to the beginning of each letter, there is some limited structure within the HTML list. As such, it is highly unlikely that the HTML list is the source of this information, and much more likely that the entire list was exported to HTML from a more-structured database. Given this, we are still unsure why a search engine front-end for the Australian Mathematical Society does not exist, and are hoping to find more information in this matter.

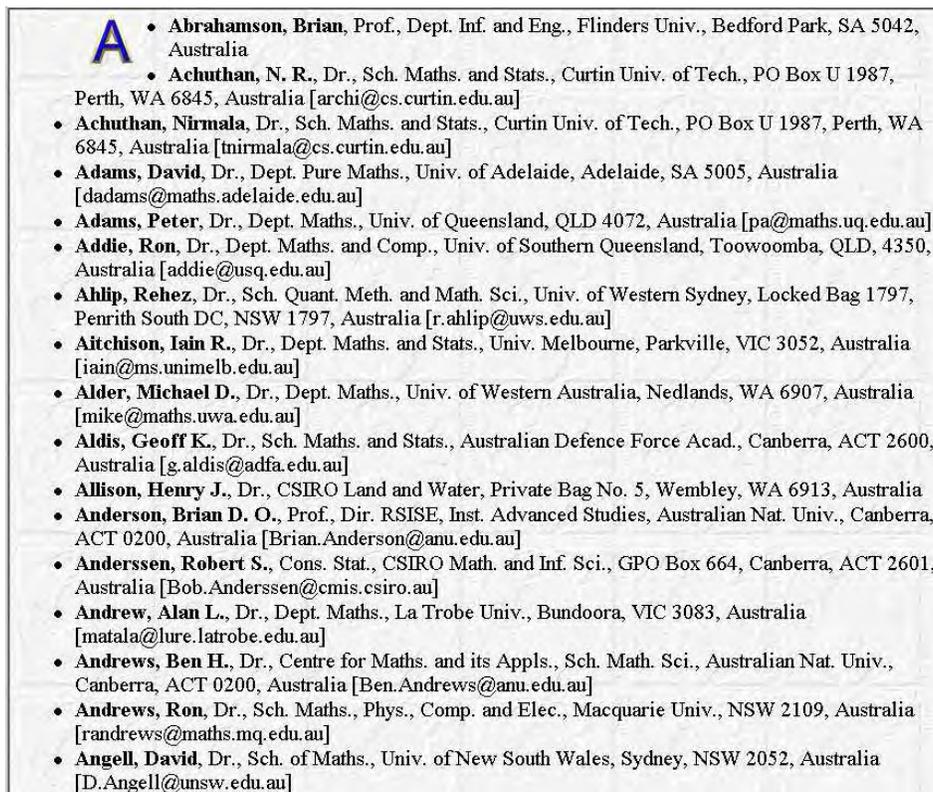


FIGURE 8. Sample of membership list from Australia's National Committee for Mathematicians

4.6.1. *Unin Matematica de Amrica Latina y el Caribe*. The Latin American and Caribbean Mathematical Union (UMALCA) currently has a searchable membership list, with a more extensive collection of information than many of the search engines currently present in the FWDM; for example, the searchable fields include the member's country of birth, and the institution where they were awarded their degree. Sample output from their search engine is shown in Figure 9.

However, despite the member's First and Last Names being present in the database, these two fields are not searchable, making it impossible to search using the Simple Search interface. Furthermore, the countries of birth and of current residence are both indexed by integer identifier rather than by text string, creating extensive overhead to search even using the Advanced Search interface.

4.6.2. *St. Petersburg Mathematical Society*. The St. Petersburg Mathematical Society currently has an online membership list which, like the NCM, consists only of a static HTML alphabetical list of members. However, unlike the NCM, the only information contained on this HTML list are the member's names, with some names linked to that member's personal homepage. Also, unlike the NCM, there are no anchors within the HTML document, making it impossible to link to even an approximate location of individual members within the list.



**UMALCA** **Directory of Latin American and Caribbean Mathematicians**

---

**Professional Name:** **Dr. Carlos Cabrelli**  
**Birth Country:** Argentina  
**Degree:** Doutor - 1984  
 UNIV. DE BUENOS AIRES - Argentina  
 Supervisor Dr. Miguel Herrera  
**Email:** [cabrelli@dm.uba.ar](mailto:cabrelli@dm.uba.ar)  
**Areas:** Abstract harmonic analysis  
 Approximations and expansions  
 Fourier analysis  
 Measure and integration

---

**Professional Name:** **Dr. Carlos Lizama**  
**Birth Country:** Chile  
**Degree:** Doutor - 1989  
 UNIVERSIDAD DE STUTTGART - Germany  
 Supervisor Prof. Dr. Laszlo Zsido  
**Email:** [clizama@usach.cl](mailto:clizama@usach.cl)  
**Areas:** Abstract harmonic analysis  
 Difference and functional equations  
 Fourier analysis  
 Functional analysis  
 Integral equations  
 Operator theory  
 Partial differential equations

FIGURE 9. Sample output from UMALCA (input `area = Abstract Harmonic Analysis`)

## 5. FUTURE WORK

Aside from including additional national search engines in the FWDM search results, some additional functionality remains to be added. Some of the search engines provide functionality to check for more than one variation of a given first or last name; for example, if the user enters `First Name = Jonathan`, then the search results would include all members with first names of `Jon` or `Jonathan`. However, since this functionality is present only in some of the search engines, then the results when one enters such a first name into the FWDM main page would be unexpected and inconsistent: in the earlier example, all `Jonathan`'s would be returned for every search engine, but only the search engines with the additional functionality would return `Jon`'s, and the merged entries displayed by the FWDM would be misleading. We can simulate this functionality within the FWDM code by performing separate searches using the name variations for all search engines without this additional functionality, using a list of names with variations that is currently under construction.

A similar technique can be used to avoid problems arising from the presence of accented characters in submitted names or locations. For names with characters that are frequently accented, multiple searches can be performed separately with and without accents; further, if accents are contained in user-input, we can execute

multiple searches using characters that are frequently used in place of the accented characters present. Note that this simulates, in a rudimentary form, the inverse of the sound extension search techniques included in the PERSONA MATHEMATICA software.

As multiple searches are added to individual component search engines, we can gain considerable time improvement by performing these search requests in parallel, by separating the requests across different threads or different processors. This raises the question of how best to present search results, either as a final combined list once all searches have been performed, or as a series of partial lists as each search engine returns its results.

Another problem that arises as multiple versions of the same search are performed for each search engine is that of duplicate removal. For example, for an individual that is a member of both the DMV and the AMS, if that individual's name appears in the DMV with an accent and in the CMS without an accent, then the current technique of duplicate removal would not detect them as the same name. This problem also arises with different forms of an individual's first name. Although this problem can be ignored as a problem with the source data that cannot be removed entirely by the FWDM, some of its effects can be lessened in cases detected by the same criteria as used to separate the multiple searches.

## 6. LINKS

- (1) International Mathematical Union (IMU)
- (2) Committee for Electronic Information Communication (CEIC)
- (3) American Mathematical Society (AMS), Combined Membership List (CML)
- (4) Canadian Mathematical Society (CMS), Membership list
- (5) Deutsche Mathematiker-Vereinigung (DMV), Membership list
- (6) Société Mathématique de France (SMF), Membership list
- (7) National Committee for Mathematics (NCM)
- (8) Unión Matemática de América Latina y el Caribe (UMALCA), Membership list
- (9) St. Petersburg Mathematical Society