

# My Gateway at the University of Washington Libraries

William Jordan

*Much of the content of the University of Washington Libraries' Information Gateway is generated from an SQL database extracted from the libraries' OPAC. The database, known locally as the Digital Registry, drives multiple services on the Information Gateway. My Gateway is the personalized component of the Information Gateway and provides a way for users to create their own views of the Registry and to integrate their own URLs into those views. This article describes the My Gateway service, its architecture, and procedures for content selection and maintenance for the Registry database. Future development plans and lessons learned are also discussed.*

**M**y Gateway is the personalized component of the University of Washington Libraries' Information Gateway and provides a way for users to maintain categorized lists of URLs that they select (see figure 1).<sup>1</sup> Users can create their own categories or "subscribe" to public categories created by library staff.<sup>2</sup> Users may add items to or delete items from categories they create (see figure 2). Public categories can only be subscribed to *in toto*, although this may change in the next version of the interface. The URLs can be drawn from the libraries' URL database or can be the users' own URLs. If the URL is in the database and the URL changes, that change is automatically reflected in users' URL lists. A "show items added since last login" feature lets users review resources that are new to the database.

Librarians may use My Gateway to assemble lists of URLs that are related in ways that are not easily represented in a classification scheme or by subject terms. These lists may then be "published" as static Web pages using My Gateway's HTML generator, or they may be made available for users to include on their own My Gateway pages.<sup>3</sup> Currently, thirty-five staff users publish seventy-five pages to the Information Gateway using this tool (see figure 3).

The only user data stored in My Gateway is the user's fourteen-digit bar code; all other data resides in the patron database in the libraries' Innovative Interfaces (III) integrated library system. When a user logs into My

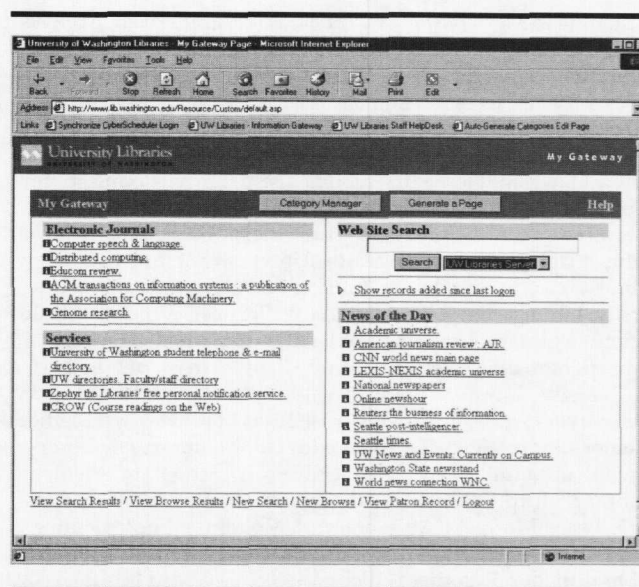


Figure 1. Screenshot of My Gateway Page

Gateway, he or she is authenticated against the III system and additional patron data is fetched from the patron database. Once authenticated, patron data is available to other request services on the site. When My Gateway was initially released, the "view my record," renewal, and request functions could all be accomplished with a single My Gateway login.

Two rounds of usability testing were done on My Gateway, the first focusing on terminology and the second on functionality. We have done very little promotion of My Gateway outside of listing it on the Information Gateway itself and providing one My Gateway class each quarter. Some librarians promote it more heavily to their individual user groups, however. Over seven thousand My Gateway accounts have been created, approximately eighteen hundred of which have been accessed at least once during the last academic quarter.

Because of the way the database is structured, it is difficult to get good demographics for My Gateway users, but it is possible to look at some overall usage patterns. Relatively little customization of categories is done. Fewer than 15 percent of users create additional categories beyond the three defaults, preferring instead to edit the contents of existing categories. The 15 percent who do create additional categories have created over five hundred of them in aggregate, but with no meaningful clusters emerging. At the item level, users have added nearly one thousand distinct URLs to their My Gateway pages that are not represented in the libraries' URL database. The majority of these URLs have only a single user associated with them. Of the resources represented in the

**William Jordan** (bjordan@u.washington.edu) is Head, Distributed Computing Systems, University of Washington Libraries, Seattle.

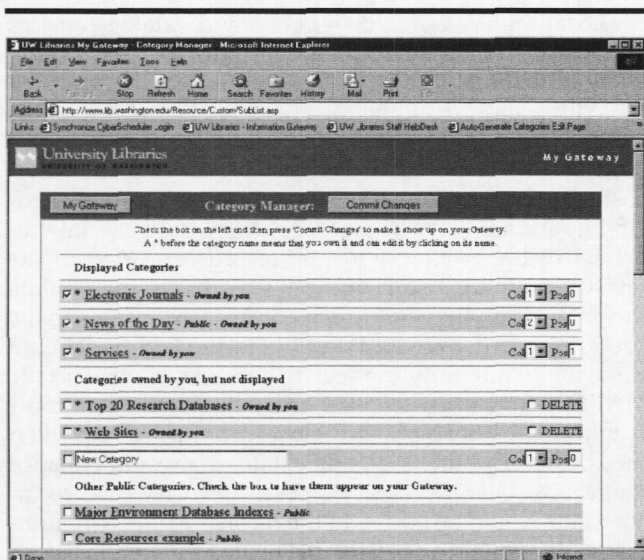


Figure 2. Screenshot of Category Manager

URL database. My Gateway users show a marked preference for databases over other types of resources; twenty-nine of the top thirty records included on My Gateway pages represent databases or catalogs.

## Project Background

### Web Services Reengineering

Throughout the 1990s the libraries delivered a number of databases to the UW campus community using a locally developed X-Windows interface (Willow) and its VT-100 analog (Wilco).<sup>4</sup> Programming and systems administration for the Willow and BRS servers were provided by Computing and Communications (the campus IT shop); the libraries contributed database profiling, some design and usability expertise, and user support in addition to purchasing the database content. We purchased tapes of MEDLINE, ERIC, INSPEC, and approximately ten other abstracting and indexing databases and loaded them under BRS/Search. Another eight to ten databases from OCLC, Silverplatter, and RLG were available via Willow's Z39.50 module.

Although we had purchased the Innovative Interfaces integrated library system in 1993 and were using it for back-end functions, we did not purchase the OPAC module. The libraries' public catalog was another BRS database that was searched with Willow. Records

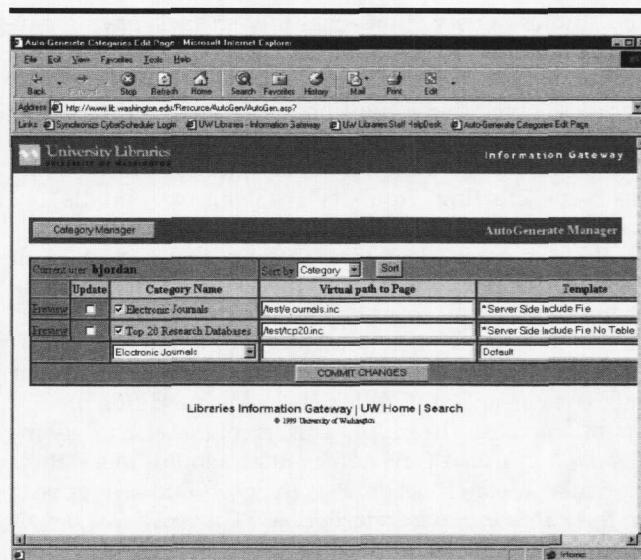


Figure 3. Screenshot of Autogen HTML Generator Page

were exported from III and loaded into BRS, nightly for material on reserves, weekly for "regular" material.

In late 1997 the libraries began reengineering its Web presence with the intention of moving to the Web as the primary database and online service delivery platform. The scope of the project was large and included migrating all our locally mounted index databases from Willow/BRS to the Web interfaces of third-party vendors (Silverplatter, ISI, NLM, and others) and installing III's WebPAC catalog. The libraries' Web site, which reflected the administrative organization of the libraries, would be redesigned with a task-oriented approach. Finally, the 250+ X-Windows terminals that served as public terminals throughout the libraries did not provide Web access and so needed to be replaced with PCs. The director appointed seven middle and senior managers to serve as a cross-functional steering committee to manage the overall project, and a Web prototyping group was launched to begin blocking out the new Web site.

### Design Considerations

A five-member prototyping team began work in mid-December 1997, meeting on a weekly basis. To arrive at a conceptual framework, the group employed several techniques; extensive discussions on principles, organization, content, and terminology occupied several weeks. Several members of the group used the conceptual elements to create Web sites that could then be tested and explored with the user in mind. Beyond the meetings, the team researched and read articles discussing user needs

and studies of user approaches to Web sites, paying particular attention to issues of navigation. Actual end-user questions were gathered at public service desks in several libraries' units; these questions were analyzed to determine the terminology used to frame inquiries. Commonalties in these reference questions formed the basis for creating "toolkits" or "starting points" of resources targeted at particular user groups.

What became clear was that no single method of organization would work well for all users. The response of the prototyping team was to present the same information resources in a number of different ways. For example, the PsycINFO database might appear in a list of databases arranged by title and in several different subject-oriented lists. In addition, users might wish to assemble *ad hoc* lists of resources that would meet their particular research needs; PsycINFO should also appear on these lists if the user desired. My Gateway was developed to meet this latter need.

While redundancy should make it easier for a user to find what he or she was looking for, it also makes site maintenance more difficult since any URL changes must be made on all pages that contain the resource. The prototyping team's response to the increased maintenance load was to create a database of URLs and associated metadata that could then be used to drive template-based HTML writers. This database became known locally as the Digital Registry and is discussed in more detail below.

Resource discovery is only one part of a user's information-seeking behavior. Equally important is actually delivering the information content to the user, whether that content is available electronically, in print as part of the libraries' holdings, or obtained through interlibrary loan. The team's vision was that content delivery ought to be integrated into the process as much as possible; *My Gateway's* integration with other request services was an attempt to do that. The team's prototype, final report, and implementation recommendations were delivered to the project steering group in early May 1998, and work to rebuild the site began in earnest at that time.

### Budget and Timeline

The overall project was broken into phases:

- Phase I, due September 1998: initial Web redesign and X-Windows terminal replacement; Digital Registry in place, Web site content mostly in place;
- Phase II, due January 1999: initial release of WebPAC, Web-based alternatives to most database services to be in place and operational; continue revising/enhancing Web site and services; and
- Phase III, due June 30, 1999: remaining database services to be available with Web interfaces; decommission Willow and BRS.

Local and Web-based database and catalog services would run in parallel from January through June 1999 to allow time for users to adapt to the new environment.

With a short deadline for Phase I deliverables, it was clear that existing processes and workflows needed to be used wherever possible. For the Digital Registry database, the decision to use existing workflows had particular importance. The Digital Registry would become an extract of the catalog, with the records exported to the registry only because it was not possible to manipulate the data within the catalog. Customized extensions to the check-in record were created to hold the additional data elements required by the registry, and a registry-specific MARC record exporter was ordered. Acquisitions staff would create brief bibliographic records (as they do when they order print material) and cataloging would enhance the records later. LC class numbers and USMARC country codes would be added to the records to support automatically generating lists of resources in classified (subject) and geographic arrangements. Technical services staff would add records for all the commercial resources that the libraries had licensed. In addition, to "seed" the registry, cataloging staff would begin creating records for noncommercial resources that were listed on the selector's existing subject Web pages.

The rapid development techniques needed to bring the project in on time ("guerilla development," one public services manager complained bitterly) ran counter to an organizational culture that stresses extensive communication and lengthy consultation between units. Multiple project threads needed to move forward in parallel even when, ideally, some would have been done sequentially. For example, some public services staff were concerned that the deliverables would not be in place far enough in advance of public release for staff training and documentation, and this certainly proved to be the case. Many of the staff training materials were prepared using prototypes or mockups and did not always reflect the released version. Discussions and consultation with units within the libraries often had to be conducted on a theoretical basis, with no real product to which units could react. Predictably, this led to misunderstandings and differing interpretations of what had actually been decided.

There were some areas of the project that were somewhat controversial or that did not have sufficiently broad buy-in from staff. For example, some selectors objected to the decision to develop a mapping of LC class numbers to subject areas and to automatically build lists of resources by subject, even though a mechanism was provided by which selectors could override the automatically generated subject pages if they wished. Only about 30 percent of our subject areas had handcrafted pages, and developers saw this automatic page generation as a way to fill in the gaps. We tended to approach the automatically generated

---

pages with a relatively high tolerance for error and failed to recognize the proprietary feelings that some selectors had about subject pages. The problem wasn't that we were generating content for *their* pages, but that we were generating subject-oriented content at all. There was a related "horns effect" among this group that resulted in skepticism about any services that relied on the Digital Registry. The registry had, for them, become inextricably identified with automatic page generation. In retrospect, it would have been better to uncouple problem areas like this from the initial release and devote more time to getting staff to buy-in to the concept. Even if we didn't end up getting complete agreement, the additional time at least would have allowed the issues to be discussed more fully and without the pressure of an impending release deadline.

## Architecture

### Toolset: IIS/ASP, Java/COM, VBScript Glue

The choice of a toolkit for the Web site portion of the project was left to library systems staff. Given the short development cycle, we opted to use tools with which we already had some experience; there simply wasn't time to identify and become fluent in an "ideal" toolkit. We needed a database engine, scripts to query the database and format the results into HTML, and a Web server to deliver the results. Since we were already running Microsoft's Internet Information Server and had a passing familiarity with Active Server Pages, we decided to use Microsoft's SQL server as the database system and use IIS and ASP (with VBScript) to query it and format the results. Where we needed more flexibility than ASP gave us, we would use Java to create classes that could either be controlled from ASP as Component Object Model (COM) objects or run in the server's Java virtual machine as applications.

### UW "Digital Registry" (SQL Database)

The SQL database created to support this project is known locally as the Digital Registry (DR) and currently contains about ten thousand records. Users can and do search the registry directly, but the main use of the DR is to support other services. In addition to My Gateway, the DR is used to generate various resource lists for the Information Gateway, including lists of databases, e-journals, and other materials arranged by title and by subject. We currently generate subject lists for more than 270 subject nodes. In some subject areas, virtually all of the content is generated from the database, while in other areas

the subject librarian has handcrafted some or all of the pages. DR-generated pages are hit 65,000 to 70,000 times each week on average and account for about half the number of hits on the subject pages. The registry is also used to generate the proxy automatic configuration files that support our proxy server. Finally, a locally developed URL checker iterates over a section of the database each night to verify that the URLs are valid.

The DR database schema itself is mainly unremarkable. There are tables to hold common bibliographic elements like author and title, plus others for subject and LC class numbers. The bibliographic elements are linked to an "instantiation" table that holds the URL, access control information (whether the resource is IP restricted, freely available, or password restricted) and other instance-specific data. The same bibliographic data can be linked to multiple instantiations of a resource in cases where we have more than one access method.

A classification-to-subject-string table maps class number ranges to locally meaningful subject strings; both LC class numbers and MARC country codes are accommodated.<sup>5</sup> Cataloging staff add additional class numbers or country codes to local fields in the records to reflect secondary aspects of the item where appropriate. The class-to-subject mappings are then used to control the generation of resource lists in a subject arrangement. For example, any items with LC class numbers E184.5 through E185 and E441 through E453 will appear on the "African American History" resource list. Items with USMARC country codes beginning "e-yu" will appear on the "Yugoslavia" resource list.

In addition to the resource-oriented tables in the DR, several additional tables support My Gateway functions. A "category" table holds the name of each category, plus the owner's bar code and a flag indicating whether the category is public or private. A "subscription" table holds the record numbers of resources to which a user has subscribed, the category to which the user has assigned them, and an optional user-supplied note. A "user" table associates user bar codes with their categories; a subsidiary user table holds a last login date and an access level that defines what privileges the user has. The default privilege level allows the user to manage his or her own page, a "staff" level allows the user to create public categories and publish categories to the Information Gateway, and an "administrator" level adds the ability to edit and create new templates for the HTML publishing feature.

### III PATAPI (Patron File API)

As mentioned above, My Gateway is integrated with the III patron database. The design goal was to eliminate the need to ask authenticated users to supply information we

already had on file (name, address, departmental affiliation, etc.) when they wished to use a request service. Integration is accomplished through an http-based read-only API to the patron file (PATAPI). PATAPI accepts a user identifier (bar code in our case) and returns a tagged HTML page representing the patron record. It will also accept a user identifier plus a Personal Identification Number (PIN) and will return a "valid" or "invalid" status based on whether the PIN matches what is in the database. Access to the PATAPI is IP-limited; for privacy reasons, we only allow our Web servers to access it.

Since we wanted to make patron data available to other request services on the site, it was important to only expose the data once the user had authenticated. We designed a Patron class, written in Java, that in turn communicates with PATAPI. Other request services can then ask the Patron class for data, which is provided only if the user has authenticated. The Patron class enforces our business rules (e.g., to require authentication before providing data), avoids duplicating "utility" functions (e.g., connections to PATAPI, parsing the data, error handling) and keeps code maintenance in one place. The advantage of this latter characteristic was apparent during our Y2K updates. When the format for the patron account expiration date changed, only the code in the Patron class had to be updated; had each request service queried PATAPI directly, each service would have needed to be updated.

## Content Development and Maintenance

### Selection and Deselection of Content

Selection and deselection of content to appear in the Digital Registry is the responsibility of approximately seventy subject selectors throughout the libraries. In order to initially populate the database in 1998, catalog records were created based on links in selector pages. At that time there were no guidelines in place specifically for the selection of Internet resources.

In 1999 the Information Resources Council appointed a task force to develop a collection development policy for Internet resources. The final draft of their *Selection Guidelines for Internet Resources* was completed in October 1999.<sup>6</sup> The guidelines confirm that general collection development procedures should guide the selection of materials in all formats, including Internet resources. They also address issues of authority, content, reliability, design, and access that are specific to Internet resources, particularly "free" ones. Selectors are reminded of the indirect costs of their selection decisions, since even free resources have a cost. As the *Guidelines* point out, "selec-

tion of resources simply because they are 'free' should be weighed against the time necessary to describe and provide access to the site, provide registration or scripting, or other resources required to make the site available."

Selectors are now being asked to review sites in their subject areas to see if those selected before the guidelines were in place fall within current collection development policy for selection of Internet resources.

### Maintenance

URL maintenance is straightforward. A URL checker iterates over a portion of the DR database each night. It sends an http HEAD request for each URL, increments a "needs attention" counter each time a 4xx or 5xx status code is returned, and decrements the counter each time a successful response is received. If the counter reaches a predetermined value (currently set at three), the URL is added to an error report. Technical services staff manually investigate all URLs on the error reports and either correct them or remove the resource. Since the DR is an extract of records from the catalog, URL maintenance is done once, in the catalog, and the changes are reflected in the Digital Registry. We are able to check each URL about twice each week right now.

Code maintenance has been more problematic, not because of bugs (although there are some) but because of the "prototype in production" nature of the project. Student programmers wrote most of the ASP scripts during our rapid prototyping stage, and we intended to rewrite the scripts before going into production. The timetable did not allow for much rewriting, however, and this has resulted in a body of poorly documented and inelegant code. Changes and enhancements have been delayed because of the need to clean up entire sections of script before making functional changes. The loss of a crucial staff member in late 1999 has further delayed the maintenance project. The Java code and SQL code, largely written by regular staff members with an eye to production, has proved much easier to maintain.

## What Works, What Doesn't, and Lessons Learned

The "view my record," renewal, and request functions all went through My Gateway at the initial release. Since we were not yet running the III WebPAC in a publicly accessible mode, these functions were all proxied by My Gateway which then submitted the requests to WebPAC. Although users appreciated the convenience of a single authentication step and reacted well to our filling out already-known information for them on our request forms, the model

proved too brittle once we brought WebPAC into production. In order to provide this functionality, we parsed the forms and rewrote any URLs to point back to My Gateway. Unfortunately, changes to the format of the form would often break the service, and we ultimately had to abandon this approach. As more of our request functions were moved into WebPAC, fewer were available via a My Gateway login, which diluted the value of the service. The usage rate for My Gateway has declined over the past several quarters (both in the number of new accounts created and in the number of repeat users); anecdotal evidence suggests that at least some of that decline is due to the loss of integration with request services.

Currently, My Gateway serves more as a roving bookmark service than as a general personalized service delivery mechanism. As a bookmark service, there are some obvious deficiencies. Users have no automated way to transfer URL lists between their browsers and My Gateway, or between My Gateway and other "My" services that they might use. Users do appreciate being able to add URLs from the registry without retyping them, as well as the automated URL maintenance that we do for them. The "view items added since last login" feature is seen as a useful current awareness tool by many, although several users have suggested adding the ability to further limit the results by keyword or by subject would make it more useful.

Campus computing services external to the libraries use a different authentication/authorization model than My Gateway, so single-login integration with nonlibrary services isn't yet possible. Recently, Computing and Communications (the central IT organization) released a cookie-based Web authentication mechanism for campus. Our users have made it clear that they prefer a single login environment, so we are currently investigating ways to bring the libraries' authentication/authorization scheme and the campus authentication service together. We switched authorization for electronic reserves to the campus "pubcookie" system in spring 2000, and that change was received very positively by our users.

## Future Plans

In June 2000 Computing and Communications released MyUW, a personalized campus portal for the University of Washington.<sup>7</sup> The initial release targeted students and included online registration, tuition balances, class schedules, etc., but additional portals for faculty, staff, and alumni are being actively developed. The libraries are looking at how we can use MyUW as a delivery vehicle for a range of targeted, personalized services, includ-

ing those currently provided by My Gateway. It is clear that we can best serve our users by delivering content to MyUW and by integrating library services with other campus services rather than by developing a separate portal. Of course, integrating library services becomes much easier when the various library systems are able to use the campus authentication, authorization, and directory services. So far, library vendors have not placed a priority on interoperating with enterprise middleware; one hopes this will change as the installed base of enterprise middleware services grows and libraries begin to ask for enterprise interoperability in their RFPs.

We have prototyped a number of interface changes to My Gateway that usability testing confirms would make the product easier to use. Most of these changes affect the way categories and items are managed. The prototype also includes a "news feed" feature that would let subject selectors send short messages to users in their departments, highlighting new services or resources that might be of interest. Our current plan is to release the new, and final, My Gateway interface sometime in the fall of 2000, but without the news feature. Any additional personalized features or services will be built for the MyUW environment.

## References and Notes

1. My Gateway and Digital Registry design and programming were done by library systems staffers William Jordan, Alex Wade, and Michael Reynolds. Additional data elements definitions, checking record design, and database schema analysis for the Digital Registry were done by technical services staffers Steven Shadle, Kathleen Forsythe, and Tricia Hart.
2. University of Washington Libraries, "My Gateway" sample page, Nov. 1998. Accessed Aug. 3, 2000, [www.lib.washington.edu/about/registry/common/mygateway.asp](http://www.lib.washington.edu/about/registry/common/mygateway.asp).
3. University of Washington Libraries, "Autogenerate Categories" sample page, Nov. 1998. Accessed Aug. 3, 2000, [www.lib.washington.edu/about/registry/common/autogen.asp](http://www.lib.washington.edu/about/registry/common/autogen.asp).
4. D. S. Ketchell et al., "Willow: A Uniform Search Interface," *Journal of the American Medical Informatics Association* 3, no. 1 (Jan./Feb. 1996): 27-37.
5. University of Washington Libraries, "Class Ranges," Nov. 1998. Accessed Aug. 3, 2000, [www.lib.washington.edu/about/registry/common/classranges.html](http://www.lib.washington.edu/about/registry/common/classranges.html).
6. University of Washington Libraries, "Selection Guidelines for Internet Resources," Nov. 1998. Accessed Aug. 3, 2000, [www.lib.washington.edu/about/registry/common/internetselfguide.htm](http://www.lib.washington.edu/about/registry/common/internetselfguide.htm).
7. University of Washington, "MyUW," 2000. Accessed Aug. 3, 2000, <http://myuw.washington.edu>.