

Discoverability

Phase 2

Final Report

University of Minnesota Libraries

Submitted to the Project Sponsors: 27 September 2010

Cody Hanson (co-chair)
Heather Hessel (co-chair)
Deborah Boudewyns
Janet Fransen
Lara Friedman-Shedlov
Stephen Hearn
Nancy Herther
Nicole Theis-Mahon
Darlene Morris (Duluth campus)
Stacie Traill
Amy West

Table of Contents

Executive Summary	3
Introduction	8
Vision	9
User Communities	12
Peer Institutions	16
Internal Data Sources.....	20
External Aggregators.....	23
Case Studies	27
Best Practices.....	33
Conclusion.....	35

Appendices 1-7 available as separate documents:

- Appendix 1: Project Charter
- Appendix 2: User Community Personas
- Appendix 3: Peer Institution Materials
- Appendix 4: Internal Data Sources Spreadsheet
- Appendix 5: Internal Data Sources Profiles
- Appendix 6: External Data Aggregator Profiles
- Appendix 7: External Data Sources Spreadsheet

Executive Summary

The Discoverability Phase 2 group was charged in spring 2010 to generate a vision for the University Libraries' discovery environment. In addition, the group was asked to build on the work of Phase 1 (see the Phase 1 report here: <http://purl.umn.edu/48258>), addressing some of the practical implications of decentralized discovery by recommending strategies for making local collections discoverable in external systems, and for integrating remotely-managed data into the local discovery environment.

Specific objectives and deliverables outlined in the charge are presented below, with a brief summary of the group's research, analysis, and findings for each task.

Two of the group's overarching objectives were informed by the work done to fulfill more specific deliverables: proposing a high-level vision for the Libraries' discovery environment, and recommending best practices for data exposure.

Vision

***Objective:** Present a high-level vision for our new Discovery environment, synthesizing current research on user needs and collection imperatives/opportunities.*

Seven key concepts emerged from the group's research and analysis:

1. **Scopes.** The range of user needs identified in the User Communities section of this report shows that University Libraries user communities require discovery interfaces and services that can be scoped based on those needs. An acceptable discovery system must integrate data from disparate sources, have a selectable range of material availability, combine or filter results based on criteria such as format, publication date, source authority, subject, or author, and intelligently handle a range of granularity in metadata.

The Libraries' environment must enable dynamic creation of discovery scopes, based on discipline, role, research task, course, library instruction, or individual user needs. Because a user's current need can't be determined based solely on their community affiliation, users must be able to customize their own scopes. Further, the discovery environment must be able to generate scopes determined by usage patterns and by metadata or format characteristics.

2. **Integrated metadata layer.** To support the creation of these scopes, the Libraries will need a system or service layer that integrates metadata from internal, external, owned, licensed, and freely-available data sources selected by library staff. This layer must support a full range of metadata granularity, allowing article- or item-level metadata to coexist with database- collection- or aggregation-level metadata. and this metadata layer must be pre-indexed.

3. **Expose internal data to external systems.** The Libraries should generate well-described, standardized metadata for local collections and data sources that can be exported, harvested, or made available for crawling by external systems.

4. Capitalizing on external data aggregations to expand discovery and delivery options.

External aggregators offer data of potential value to our user communities and are key for enhancing the local discovery environment. The Libraries' discovery environment should include metadata from selected external sources,

5. Object manipulation and personal curation. Searchers should be able to discover, access, and interact with a digital resource as a seamless process. Types of interaction can include adding personalized data, e.g., annotations; creating and sharing ad-hoc collections; repurposing content in new ways (as rights allow); or enhancing access points for other searchers.

6. Data-driven decisions. The Libraries should utilize systems which provide consistent data sufficient for measurement and analysis of user behavior and interactions throughout all aspects of our discovery environment.

7. Ongoing review and evolution of discovery environment. The Libraries should develop a dynamic process for conducting ongoing reviews of the environment as a means to accommodate technological changes and the evolution of the discovery and delivery expectations.

Best Practices

Objective: Codify best practices for exposing our data.

The best practices toward exposing the Libraries' metadata proposed below are organized under three headings: metadata, systems, and community. Distinct kinds of policy and process decisions are needed in each of these areas. Discoverability requires the expenditure of resources; trade-offs and prioritization are essential. Decisions about discoverability services are practical only to the extent that library resources can sustain these efforts.

Metadata

- Descriptive metadata should be delivered with all necessary contextual information, and critical descriptive elements should always convey useful information.
- Access points should include a uniform data string for retrieval, such as a resolvable URI or a text string, preferably with reference to an established controlled vocabulary.
- The process of making local metadata available to external systems should include a consideration of the elements that would facilitate optimal relevancy ranking by those systems.
- Information about rights associated with an object should be included in any exported data. Optimally, this data would allow external systems to derive rights status in recognition that status is a dynamic attribute.
- Both the content and format of metadata should be kept up to date in the home repository, and updated metadata should be made available or pushed to external sites.
- Metadata should adhere to shared standards, in order to facilitate and simplify the data transformations necessary for successful metadata exchange and interoperability.

Systems

- Local customization of platforms and systems that hinders the ability to facilitate exchange of data should be limited to promote ease of interoperability.

- The process of exposing data to external systems should include a consideration of service continuity where discovery is coupled with expectations for delivery.
- Exposed metadata should include any relevant information (such as a need for user authentication) that would assist with the process of delivery, and should enable external systems to offer choices based on availability.

Community

- Community participation, such as collaboration with service providers to maximize local exposure and contribution of innovative models for data, software and process to the larger community, should be encouraged. Such activities help to simplify the work of integrating community-level changes into the local discovery environment.
- Accurate and accessible documentation of library policies, practices, and data and service models should be made available to assist both end users and collaboration and development partners in making good use of the libraries' resources.

Subgroups

Four subgroups were formed to address the more specific themes in the objectives and deliverables. These were: identifying and describing user communities, describing characteristics of external data aggregations, learning about significant discovery work at peer institutions, and performing an inventory and analysis of internal data sources. Each objective is stated below, and followed by a brief description of the work of its subgroup.

***Objective:** Identify relevant characteristics in the user communities we serve in regards to Discovery and Delivery, conducting additional research as needed.*

The **User Communities** subgroup developed sixteen personas to reflect the diversity of the Libraries' user communities through a review of the Fall 2009 survey of users of MNCAT OPACs, a review of recent literature (2005 or later) exploring information seeking behaviors for particular user communities, and consultation with Libraries staff. The personas are detailed in Appendix 2. They are also available as a website: (<https://sites.google.com/a/umn.edu/library-user-communities/>).

***Objective:** Investigate and summarize how selected, notable peer institutions have redefined and redesigned their discovery layers.*

The **Peer Institutions** subgroup consulted with sponsors, senior staff members, and professional contacts at other institutions. The group gathered published or web-posted information from a broad range of these institutions, and selected four institutions to contact directly: California Digital Library (CDL) for the University of California; University of Wisconsin--Madison; University of Chicago; and the University of Kansas.

***Objective:** Identify the Libraries' primary data sources and determine which of our resources can and should be accessible to external systems and emerging interfaces. Identify Libraries collections that cannot be made adequately discoverable through external systems.*

The **Internal Data Sources** subgroup identified a range of characteristics to use to describe each data source, and then developed an extensive inventory of locally produced primary data sources (Appendix 4) from several of the U of MN campuses.

Objective: Investigate and make recommendations for how the Libraries can best utilize large data aggregations in offering discovery services to our users. Identify external systems to be targeted in our efforts to expose our metadata and holdings and discuss how to accomplish these objectives:

- Consider the potential impact of availability through a given resource as well as the effort required on our part

- Review the process of fulfillment for all of the key external search interfaces.

Make recommendations for optimizing fulfillment, if needed.

The **External Aggregators** subgroup assembled a list of high profile and representative services which aggregate metadata and data objects (see Appendix 6). The group categorized these aggregators by broad type, and identified thirteen factors for consideration when the Libraries decide which aggregators to contribute to and feature for users.

Recommendations

The recommendations from each of the Discoverability Phase 2 sub-groups are intended to guide the next steps for the Libraries. Many of the recommendations suggest the need for further investigation in order to determine compatibility, value and impact, as well as cross-analysis with information gathered on other recommendations. Some of these complexities and cross-analysis are directly referred to in the recommending statements. In addition to the Vision detailed above, the final recommendations from the phase 2 Discoverability group are noted, here, in abbreviated form:

- 1) If it is still viable and valuable, standardize data from internal data stores to facilitate better sharing of the content.
- 2) Consider a full analysis of each data store for its potential to contribute exclusive data to external aggregations, offset by cost and other factors.
- 3) Generate well-described, standardized metadata for local collections and data sources that can be exported, harvested, or made available for crawling by external systems.
- 4) Select systems or datasets based on a cross-analysis of the type of aggregation and criteria for participation.
- 5) Carry forward with the objective to examine specific external systems and aggregators.
- 6) Remain responsive to user needs and make data-driven decisions by favoring systems that provide consistent data sufficient for measurement and analysis of user behavior and interactions throughout the discovery environment.
- 7) Develop a dynamic process for conducting ongoing reviews of the environment as a means to accommodate technological changes and the evolution of the discovery and delivery expectations. These reviews must encompass user needs, data mapping, relevance rankings, and source quality.
- 8) Evaluate the opportunity to use Primo for exposing local holdings to public search engines such as Google, Yahoo, and Bing; and that will also expose holdings records to the university website search engine.
- 9) Balance the potential gains from adoption of a discoverability service against the practical requirements of implementing and sustaining the service.

Our users need to be able to discover, access, and interact with relevant data from internal, external, owned, licensed, and freely-available data sources. To invigorate the vision for a

new discovery environment, the Libraries will need to consider options and strategies for designing and coordinating the discovery environment that address the key concepts that emerged from the group's research and analysis.

Introduction

The Discoverability Phase 2 group, like Phase 1 before it, was charged by the University of Minnesota Libraries' Web Services Steering Committee. The group's full charge is available in Appendix 1.

Phase 2 was intended to generate a vision for the University Libraries' discovery environment. In addition, the group was asked to build on the work of Phase 1 (see the Phase 1 report here: <http://purl.umn.edu/48258>), addressing some of the practical implications of decentralized discovery by recommending strategies for making local collections discoverable in external systems, and for integrating remotely-managed data into our local discovery environment.

It is important to note that the group was explicitly instructed not to consider specific systems. This report does not look at available products or services, but instead focuses on a vision and recommendations that we hope will be instructive in future purchase and implementation inquiries.

The specific objectives and deliverables as outlined in the charge are as follows. Where appropriate we have also reprinted these objectives in the report alongside relevant sections.

Objectives

- 1) Present a high-level vision for our new Discovery environment, synthesizing current research on user needs and collection imperatives/opportunities.
- 2) Identify relevant characteristics in the user communities we serve in regards to Discovery and Delivery, conducting additional research as needed.
- 3) Investigate and summarize how selected, notable peer institutions have redefined and redesigned their discovery layers.
- 4) Investigate and make recommendations for how the Libraries can best utilize large data aggregations in offering discovery services to our users.
- 5) Identify the Libraries' primary data sources and determine which of our resources can and should be accessible to external systems and emerging interfaces.
 - Identify Libraries collections that cannot be made adequately discoverable through external systems.
 - Discuss where trade-offs occur in regards to library branding.
- 6) Identify external systems to be targeted in our efforts to expose our metadata and holdings and discuss how to accomplish these objectives
 - Consider the potential impact of availability through a given resource as well as the effort required on our part.
 - Review the process of fulfillment for all of the key external search interfaces. Make recommendations for optimizing fulfillment, if needed.
- 7) Codify best practices for exposing our data.

Methodology

The Phase 2 group kicked off in late April 2010 with a review of the outcomes and recommendations from Phase 1, and a discussion of the objectives and expectations for Phase

2. The team agreed that a four-pronged approach made sense as an initial starting point. Four subgroups were created to refine scope and gather material on specific deliverables:

- 1) User communities (Janet Fransen, lead)
 - a. Research evidence in information-seeking behavior literature
 - b. Analyze data available on information-seeking behavior, e.g., MNCAT survey
 - c. Solicit feedback and create opportunities for engagement with library staff with expertise in user communities, e.g., liaisons, reference desk staff
- 2) Research on external data aggregations (Stephen Hearn, lead)
 - a. Research and classify prominent aggregations; identify critical characteristics
 - b. Draft a decision-making framework for discussion
- 3) Interviews and study of discovery work at peer institutions (Nancy Herther, lead)
 - a. Draft a set of questions for inquiry
 - b. Compile a list of institutions to contact and conduct interviews/correspondence
- 4) Inventory and analysis of internal data sources (Amy West, lead)
 - a. Identify and contact knowledgeable library staff
 - b. Identify critical characteristics for each source

As each subgroup made progress on discussions around their assigned topic and objective, the team members sought input and feedback from the full project team and adjusted their strategy as needed. In July, the co-chairs checked in with project sponsors to ensure that the project was on target and to solicit comments. By late summer, the full group synthesized the work from each section, and began scoping the required Vision Statement and Best Practices for Exposing Metadata. In late September, the project team met with sponsors to submit and review the final report.

The group acknowledges and greatly appreciates feedback and contributions by other University Libraries staff, particularly the project sponsors, Janice Jaguszewski and Shane Nackerud; liaisons in the Academic Programs and Health Sciences Libraries divisions; Jeff Peterson; Betsy Friesen; Jim Stemper; Katherine Hedin.

The group also gratefully acknowledges the work of Lorcan Dempsey, who as part of the University Libraries' strategic planning process provided us with a vocabulary for describing concepts such as the inside-out and outside-in aspects of contemporary discovery. Dempsey's slides are online here: <https://wiki.lib.umn.edu/Staff/UniversityLibrariesSpeakerSeries#dempsey>

Vision

Objective: Present a high-level vision for our new Discovery environment, synthesizing current research on user needs and collection imperatives/opportunities.

Defining Key Terms/Concepts

The following concepts are the working definitions of several terms used throughout this report as understood by the members of the Discoverability Phase 2 group:

Environment. References to the discovery environment are not meant to refer to a specific system or tool, but rather the suite of tools and services that we knit together to enable our users to discover and access relevant information resources.

While the University Libraries' current environment includes one system that is marketed as a "discovery interface", Ex Libris's Primo, that system is by no means the sum total of our public-facing discovery environment. In addition to Primo, our environment includes the

Aleph OPAC, SFX link resolver, a number of web sites, various homegrown databases and indexes, hundreds of licensed resources, and more.

Silos. The Libraries' systems and interfaces are often described as information silos, which has assumed a negative connotation due to the idea that searchers will miss valuable resources unless they know of and search all the relevant information silos. Many factors, including those historical, technical, legal, political, and administrative, explain why resources have or haven't been cataloged how they came to be contained in discrete systems. Certain discovery and information retrieval tasks may be well-matched to existing silos, for example, to fully benefit from a certain level of data granularity. However, silos are also inevitably barriers to discovery. Until very recently, the only method of transcending silos containing content managed remotely has been through federated search technology, such as Ex Libris's MetaLib application.

Scopes. Creating scopes refers to the action of filtering or grouping the set of resources available in our discovery environment, presenting resources and/or tools in logical groupings for searching or browsing, based on the tasks associated with various user communities. In effect, scopes are user-centered information silos.

User communities. Library managed metadata and information resources should be discoverable by a wide and diverse user population, including affiliates, non-affiliates, users of locally-managed interfaces (e.g., website, catalog), users of remote systems (e.g., WorldCat, Google Scholar), communities defined by discipline, geography, task, institutional role (e.g. faculty, staff, undergraduate, graduate student). In some cases access may be restricted by license agreements and other contingencies; but decisions about discovery should not be limited to serving only a particular segment of the potential users.

Vision for Discoverability

Scopes. The group's research on user communities indicates a range of discovery tasks and techniques so diverse that user needs are unlikely to be met with monolithic discovery interfaces. Each University of Minnesota user community needs to connect to a range of items owned or otherwise controlled by University Libraries. Any search needs to be properly scoped for the user, based on the user's community, past usage, and stated preferences. In an ideal discovery environment, the search target contains metadata for any resources the user might want, but scoping prevents information overload. Scoping should be transparent to the user based on their community and past behavior, but easily adjustable by the user when necessary.

University Libraries' user communities require discovery interfaces and services that can be scoped to:

- **Integrate data** from disparate sources. A searcher should not miss out on valuable resources because he does not realize those resources aren't included in the Libraries catalog.
- Encompass a **selectable range of material availability**. If the searcher's need is immediate, she may want to limit a search to only online materials. But in most situations, the searcher is willing accept a range of delivery alternatives: visit a library shelf, request that a book be delivered from one campus library to another within a

couple of days, or even request that a rare item be delivered from a faraway library within a few weeks. Choosing the level of material availability should be the searcher's prerogative.

- Combine or filter results based on criteria such as **format, publication date, source authority, subject, or author** as indicated by metadata. An integrated data environment will necessarily return large result sets; the searcher must have tools available to limit the number of results returned in predictable ways.
- Intelligently handle **a range of granularity in metadata**. In an integrated data environment, objects from different sources will be represented by different levels of descriptive metadata. The discovery interface must account for this range as part of its selection and ranking algorithms.

The Libraries' environment must enable dynamic creation of discovery scopes, allowing library staff to curate scopes based on discipline, role, research task, course, library instruction, or individual user needs. Likewise, users must be able to customize their own scopes. Further, our discovery environment must be able to generate scopes determined by usage patterns and by metadata or format characteristics.

A particular scope is initially determined by the choice of which data sources to include. The scope is limited by filters applied to the metadata and by availability of the material given the current context of the user--both factors that might be thought of as search criteria. The number of results returned may still be large enough to preclude reviewing every entry, so the discovery environment must also apply user- and context-sensitive relevancy ranking algorithms. With all the pieces in place, the searcher sees the results most likely to fit their needs at the top of the list.

Integrated metadata layer. In order to support the creation of these scopes, the Libraries will need a system or service layer that integrates metadata from internal, external, owned, licensed, and freely-available data sources selected by library staff. This layer must support a full range of metadata granularity, allowing article- or item-level metadata to coexist with database- collection- or aggregation-level metadata. In order to provide an acceptable level of performance within the limits of hardware and the host servers' environments, this integrated metadata layer must be pre-indexed.

Expose internal data to external systems. Users' discovery activities don't take place solely within interfaces managed by the Libraries. The Libraries should generate well-described, standardized metadata for local collections and data sources that can be exported, harvested, or made available for crawling by external systems. While local metadata creation may be limited by our organizational capacity, and exposure of metadata may be limited by license agreements, our default position should be to create and maintain readily interoperable metadata.

Capitalizing on external data aggregations to expand discovery and delivery options.

The Libraries' discovery environment should include metadata from selected external sources, using the suggested criteria framework detailed in the External Aggregators section. External aggregators offer data of potential value to our user communities. As noted in the peer institutions section the integration of external data sources is key for enhancing the local discovery environment.

Object manipulation and personal curation. Searchers should be able to discover and access a digital resource, and be able to interact with it as a seamless process. A “resource” as described here may be at varying levels of granularity, including citation, journal, database, or “scope” as described above. Types of interaction can include adding personalized data, e.g., annotations; creating and sharing ad-hoc collections; repurposing content in new ways (as rights allow); or enhancing access points for other searchers. These tools for interaction can also be used by all library staff to add value to an object and create new avenues of access, enriching discoverability.

Data-driven decisions. To remain responsive to user needs and make data-driven decisions the Libraries should utilize systems which provide consistent data sufficient for measurement and analysis of user behavior and interactions throughout all aspects of our discovery environment.

Ongoing review and evolution of discovery environment. The Libraries should develop a dynamic process for conducting ongoing reviews of the environment as a means to accommodate technological changes and the evolution of the discovery and delivery expectations. These reviews must encompass user needs, data mapping, relevance rankings, and source quality.

User Communities

Objective: Identify relevant characteristics in the user communities we serve in regards to Discovery and Delivery, conducting additional research as needed.

The Libraries’ data environment exists to serve our users. Any Libraries’ group discussion seeking a decision about what will best serve “our users” is sure to be filled with statements like “that won’t work for my users,” or even “my users will hate that.” The communities of users at the University are simply so diverse that no single search box or feature is appropriate for all.

This problem is not new or unique to libraries, of course: Most software and web designers, as well as consumer product designers in general, hope that whatever they create will be used by a wide range of people for many different purposes. Reaching that goal requires understanding who the users are, but one can’t keep the needs of thousands of people in one’s head and expect to produce a meaningful result. In his 1999 book *The Inmates are Running the Asylum*, software designer Alan Cooper suggests that the solution is to design for a small set of very specific and fictional users.¹

This report takes Cooper’s advice and uses personas to represent a wide range of University Libraries users. The sixteen personas, detailed in Appendix 2, represent a cross-section of University Libraries users. Not every actual user is perfectly aligned with a single persona, and not every possible group is represented. But when taken as a set, nearly all users’ needs are represented. In addition to the report appendix, the personas are also available as a website (<https://sites.google.com/a/umn.edu/library-user-communities/>). This format not only

¹ Cooper, A. (1999). *The inmates are running the asylum*. Indianapolis IN: Sams.

promotes easy access to the personas but also provides flexibility as user needs change or if additional personas are needed later.

Process

In the Fall of 2009, the University Libraries conducted a survey of users of our MNCAT OPACs. The survey showed distinct differences in satisfaction with the existing catalog interface and approaches to discovery among undergraduates, graduate students, and faculty. Figure 1, for example, illustrates the differing responses when users were asked how often they searched the catalog for an item by title and how often they searched for any relevant resource. Differences among users affiliated with different colleges were also apparent.

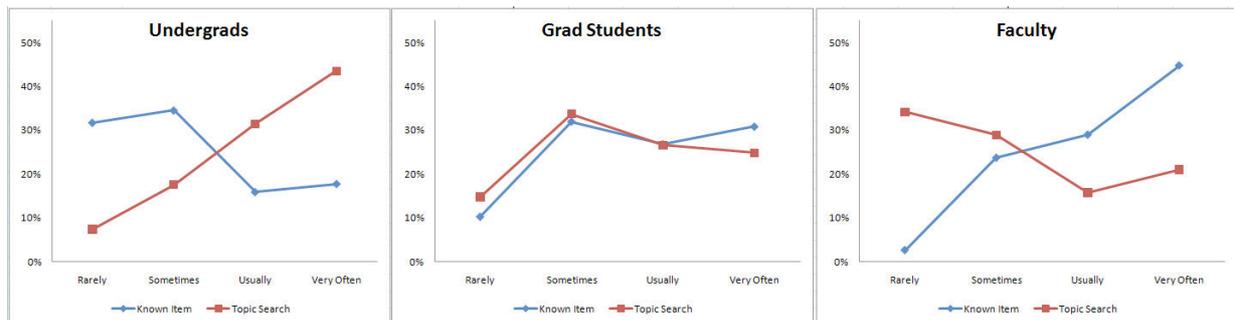


Figure 1. Searching for a known item vs. any relevant resource (MNCAT Plus respondents)

Survey respondents were self-selecting and therefore not necessarily representative of the user communities at large. Nevertheless, the survey results seemed to confirm the anecdotes regarding the diversity of user needs and behaviors that surface when librarians who work in different disciplines discuss their user communities.

Based on these survey results, the group created an initial inventory of user communities, including non-affiliate users, undergraduates, graduate students, faculty, and users working in many disciplines. In order to identify relevant characteristics of the user communities with regard to discovery and delivery, we sought out and reviewed recent literature (2005 or later) exploring information seeking behaviors for particular user communities. Based on findings from the literature, or lack thereof, the group fleshed out, combined, or discarded various community descriptions. In addition, where appropriate the group consulted Libraries staff with relevant experience. The final cross section of user communities at the University of Minnesota includes profiles and personas for the following:

Affiliates	Non-Affiliates	Health Sciences	Other Disciplines
Undergraduates	High School Students	Clinicians	Arts and Humanities
Graduate Students	Engineering	Professional Students – Health Sciences	Visual Arts
Faculty	Professionals	Researchers/Fellows – Health Sciences	Engineering
	Commercial Users		Law
	Archives Users		Physical/Life Sciences
			Social Sciences

User communities dropped due to lack of relevant literature or insufficiently unique characteristics include alumni, non-affiliate citizens of the state, social science/education professionals, and librarians at other institutions.

After careful consideration, the group chose not to include library staff as a separate user community. User communities in this context are groups with shared information seeking behaviors; library staff spend more time with the Libraries' discovery systems than any other group, but their purposes and behaviors are generally aligned with one or more of the user community identified here. The discovery environment envisioned here would fulfill the needs of each user community, which satisfies the major concerns of reference staff and liaisons. This discovery environment does not preclude excellent internal systems that address staff needs and preferences.

Findings

Affiliate Users by Type

Perhaps no user community at the University has been analyzed more than undergraduates. Not only does this community have the most members of any we described, but also this is the group most subject to change over time. After all, this group has a nearly complete turnover over the course of four years. The undergraduate community is likely to search library resources because they are told to do so and, compared with other communities, the least likely to understand when and why they need to search in multiple places to find what they need for a single assignment. As an undergraduate matures from freshman/sophomore to junior/senior, her information needs and the places she searches are likely to change and become more aligned with her discipline.

In the language of information retrieval, members of the undergraduate community are likely to want a high precision, low recall result: They don't want to wade through hundreds of results, and they want the first few they see to fit their current need. By contrast, graduate students, especially those working on dissertations, want to see every single possibly relevant resource (low precision, high recall). Their many responsibilities and lack of time make them a logical market for a single search interface. Yet in the current environment they would be the community least well served by a one-stop search.

Graduate students are very much defined by their disciplines, but are still learning how to use library resources effectively. Depending on their undergraduate experience, many still struggle with basics, such as reading and interpreting a citation, and they lack a systematic approach to their searches. Faculty members have mastered these basics: They understand the relationships among types of literature of their own discipline: when to use a peer-reviewed journal article vs. a conference paper, how a standard comes to be, the role of archival materials in research. Many speak fondly of days spent in the library stacks, but there is a growing reliance on electronic materials across disciplines. Members of this community are not bothered by visiting different sites to find the materials they need. They are more concerned with the current awareness aspects of discovery: keeping up to date on new items that fit search criteria, are added to a collection, or are published in a journal in their field.

The abundance of literature available regarding information seeking behaviors of users in the Health Sciences allowed us to uncover striking differences between clinicians and other types of users. Clinicians' discovery and delivery needs are defined by their setting and task at hand: They need high precision and vetted information delivered immediately. They are often conducting their searches either at the bedside, in the lab, or out in the field and may not have access to the physical library. Unlike graduates, clinicians do not have time to wade through

hundreds of results, and unlike undergraduates they need to make sure that their information is evidence-based and best suited to the current question they need to answer. Clinicians are most concerned with evidence-based and practice-based application of information. They rely on specific databases for quick information retrieval and journal articles for references.

Discipline-Based User Communities

The disciplines described in the Discoverability cross section differ in many ways. Some include:

- The role of currency. Scholars in the sciences and engineering are primarily concerned with recent work. Scholars in many other fields place a high value on older literature.
- The importance they place on serendipity. Scholars in arts and humanities value browsing and enjoy discovering material not by searching as much as finding by chance. Scholars in other fields prefer discovering what they need within heavily filtered result sets.
- Type of literature. Scholars in many engineering disciplines consider conference papers very important; scholars in other sciences consider the filter of peer-review paramount. Humanities scholars find great value in monographs.

These and other differences seem to point toward several distinct kinds of discovery environments to serve all these needs. Careful analysis is necessary to define an environment where these differences are respected but the environment itself and its upkeep are affordable, sustainable, and scalable.

Non-affiliate Communities

The non-affiliate user communities we profiled are very different, but their discovery and delivery needs seem to fall into one of two groups:

- 1.) Users who know exactly what they need and are merely using University Libraries for fulfillment (delivery).
- 2.) Users who aren't sure what they need but have figured out or been told that University Libraries might be able to help them. These users present themselves to a librarian or an archivist for mediated search help.

There are a few non-affiliate users who are content to use our public computers to search for what they need without help from staff, but their needs and preferences can likely be represented by discipline-based user communities.

Summary

User communities at the University of Minnesota each need to connect to a range of items owned or otherwise controlled by University Libraries. The range needed by one user, however, is likely to be completely different from the range needed by another, even if they use the same search string. In other words, any search needs to be properly scoped for the user. In an ideal discovery environment, the search target would contain metadata for any resources the user might want, but scoping would prevent information overload. Scoping would be transparent to the user based on their community and past behavior, but easily adjustable by the user when necessary.

The group acknowledges that even with an ideal discovery environment available at the Libraries website, users will often discover a resource they want somewhere else. In order to provide seamless delivery, the Libraries must continue to find ways to “advertise” its ownership of an item in any environment our communities might use.

Peer Institutions

Objective: Investigate and summarize how selected, notable peer institutions have redefined and redesigned their discovery layers.

Methodology

The peer institutions subgroup consulted with sponsors, senior staff members, and professional contacts at institutions with which we had prior knowledge or experience to compile a list of peer institutions. The group gathered published or web-posted information from a broad range of these institutions, and selected four institutions to contact directly. Those institutions, and a list of questions intended to provoke discussion related to the project are as follows:

Institutions

- California Digital Library (CDL) for the University of California
- University of Wisconsin – Madison
- University of Chicago
- University of Kansas

Questions for discussion

- Do you have any reports or research on your institution's work in the area of discovery and delivery to which you could point us?
- Has your institution developed guidelines or principles for your discovery layer which you could share with us?
- What tools, systems (and strategies) have you implemented to promote discovery across collections (to move away from "silos")? If you are using specific tools or systems (e.g. Summon, Primo, Blacklight, etc.)? Can you briefly describe the aspects of your collection, environment, and/or user needs that led you to choose this approach?
- Are there external data sources that you feel are important to integrate into your local discovery environment (e.g. HathiTrust records)? If so, what are those sources and how are you planning to integrate them?
- Have you identified any user groups that you feel require a specialized discovery environment? Do you segment your discovery environment for any specific user groups?
- Where (i.e. to what services) is your library exporting data or making it available to be harvested? e.g. WorldCat, REPEC, Google Scholar, Digital Dissertations, OAI sources
- Are there specific data aggregators to which you feel are critical to send your data, i.e. aggregators that you prioritize when you share your data (e.g. OCLC WorldCat, Archive Grid, EBSCO)?
- If we have further questions, can we contact you or can you direct us to someone else to talk to?

Two of the institutions chose a telephone interview option and the other two contacts chose to answer by email. The full text of the email introduction and a bibliography of reports and planning documents consulted are available in Appendix 3.

Results

These interviews and investigations confirmed that the issues related to discovery with which the University Libraries are wrestling are hardly unique among our peers. The past several years have seen active work by some leading institutions such as the California Digital Library. Other institutions are just beginning to tackle the challenges of evolving their discovery environment and are looking to the University of Minnesota Libraries as a leader in this area; the University of Kansas specifically cited Discoverability Phase 1 as a touchstone for their current explorations. Though the University of Minnesota has the benefit of the work of our forward-looking peers, it's clear that there is much yet to contribute to this discussion.

Reports or research

A number of other institutions have prepared reports that intersect with the scope of the University of Minnesota project team. A bibliography with links to referenced reports and related materials has been compiled in an Appendix. Most influential to the project team were the Bibliographic Task Force report from CDL: Rethinking How We Provide Bibliographic Services for the University of California and Report from the CUWL [Council of University of Wisconsin Libraries] Resource Discovery Exploratory Task Force.

Guidelines or principles

Each of the institutions contacted or researched by the subgroup was at some point in the process of investigating or implementing an updated discovery interface. The University of Kansas was in the early stages of investigation, but was working on a very compressed timeline, and expected to issue recommendations this Fall. The California Digital Library was in the process of revisiting some of their pioneering work described in their 2005 Rethinking How We Provide Bibliographic Services report, and specifically asking the following questions:

- Should there be a single point for all University of California (UC) content?
- Is there a one-size-fits-all consideration? How can/should individualized needs be supported?
- What defines UC content? What is owned? What is licensed?
- We live in a hybrid world, some content still paper-based; yet moving to digital. How to best manage the differing formats?
- Should we share all metadata?
- What about new, emerging areas, such as data sets, which are viewed as another kind of content.
- Metadata from museums and archives need visibility. Should they be included in the single delivery point system?
- What are appropriate steps in the consideration of archiving web content?
- Where discovery ends and delivery begin?

The University of Wisconsin had generated a set of recommendations that hew closely to the work of Discoverability Phases 1 and 2, though they expand to include staffing and course integration. These prescient recommendations are reproduced below in their entirety:

- Decouple the interface from the ILS so that it is sleek, lean, and enabled for rapid change.
- Maintain complete control over the discovery interface, data, and index. Nothing should be unchangeable.
- Emphasize simplicity in the interface. As Lorcan Dempsey noted: “‘simple search’ but supported by smart results and rich browse” (single search box, single sign on, clean layout).
- Include sophisticated search and result functionalities (faceted browsing and/or topical clustering, natural language, obvious relevancy ranking, searching within results, clarity via FRBRization).
- Seamlessly integrate and deliver UW collections and resources at the campus and at the system level (library catalogs, library web sites, digital collections, museums, archives).
- Adapt to user behaviors and expectations (personalization, recommendations, "did you mean?" functionality, internationalization).
- Encourage personalization and customization of the discovery environment in MyUW and course management systems, including Learn@UW and Moodle.
- Deliver library search functionality, links and services where our users work and play, including off-campus resources (Amazon, iGoogle, Facebook, WorldCat).
- Compare well in design and user experience to popular Internet destinations. Resource discovery in the libraries must become Fast, Smart, and Engaging to compete in the current and future information marketplace.
- Be staffed for excellence and continuous change (developers, graphic and interaction designers, and public services staff). This includes collaboration and leadership within the Open Source community.

Tools, systems, or strategies for discovery across collections

Though Discoverability Phase 2 was specifically instructed to exclude consideration of specific tools and systems, it seemed appropriate to survey our peers to begin to determine the breadth of the potential options for future consideration. In particular, we were curious as to how our peers were accomplishing the cross-silo discovery that we hope to be able to achieve in our discovery environment.

The University of Wisconsin implemented the open-source Project Blacklight to create their Forward database (<http://forward.library.wisconsin.edu/>). Forward provides unified discovery across the traditional catalog as well as the University’s digital collections and institutional repository. Forward exhibits a set of simple scopes, one for each UW campus, which serve as an interesting model.

The University of Chicago provides a traditional, broadcast-style federated search across licensed databases and indexes using Ex Libris’s MetaLib, the same tool we’ve used at the University Libraries. In addition, they’ve implemented the AquaBrowser discovery layer (<http://lens.lib.uchicago.edu/>), which provides search across the catalog as well as the library’s special collection finding aids and several digital collections.

The University of Kansas has been using 360 Search from Serials Solutions (<http://catalog.lib.ku.edu/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First>), and is investigating their Summon product, which provides pre-indexed metasearch across licensed databases and indexes, as well as across the local catalog and collections. Summon has had several recent high-profile implementations, including at the University of Michigan.

External data sources integrated into the local environment

The highest-profile recent innovations in discovery include tools like the above-mentioned Summon, as well as Ex Libris's Primo Central, both of which integrate records from local catalogs and collections with those from external licensed and open-access sources. The group felt it appropriate to investigate to what extent our peers were taking steps toward this type of integration in their discovery layers.

California Digital Library has been a part of the UC system's implementation of WorldCat Local, a system which can integrate local collections with those of libraries around the world, allowing patrons to discover and request materials from other institutions. Most commonly mentioned by our peers were several resources that the University Libraries already integrate into our MNCAT interface, namely HathiTrust, OCLCL/WorldCat, and Google Books. Like the University Libraries, most of our peers use these sources to enhance existing catalog records, and do not add records for items not already in their collections.

Specialized environments for specific user groups

As mentioned above, our research found that the needs of our users are diverse, and in some cases incompatible with the traditional library model of the single catalog interface serving the entire community. The subgroup inquired with our peers to determine if any had progressed down this path toward specialized interfaces for specific user groups. Few mentioned efforts to segment their discovery environments.

Wisconsin noted that they “envision a role for librarians and library staff as development partners as we build a resource discovery tool that emphasizes our users’ perspective.” Chicago shared a report on persona development from their Assessment office. CDL is working to create “a variety of segments and specialized environments.” Their eScholarship Repository (<http://www.escholarship.org>) is linked to their OPAC, Melvyl. Calisphere, which includes multimedia, text, audio, pictures, etc., is “a free website that offers educators, students and the public access to more than 200,000 primary sources...reveal[ing] the diverse history and culture of California and its role in national and world history.” (www.calisphere.universityofcalifornia.edu/) CDL identified the major issues as “prioritizing and scoping” and suggested that there were no clear cut answers. However, they noted that WorldCat Local is allowing the organization an opportunity to experiment with all types of sources and metadata as well as different types of content, with a goal of seamless integration or even just basic discovery.

Export and harvest of local data

As noted in Discoverability Phase 1, the range of viable tools available to our users for conducting research has expanded, and is no longer limited to those offered directly through

the library. The subgroup was eager to learn what steps our peers were taking to ensure their collections would be discoverable in these external systems.

Without exception, these institutions mentioned efforts to maintain holdings information in WorldCat and Google Scholar. In addition, California Digital Library mentioned a number of efforts above and beyond these, including both search engine optimization and OAI publishing, including eScholarship (<http://escholarship.org>), ITHistory (<http://www.ithistory.org/>), LiveSearch (<http://livesearch.spaces.live.com/>), and exalead (<http://www.exalead.com/software/>). The University of Wisconsin recently “set up a harvest page to support the Archive Grid proprietary spider,” and also targets OAIster. Archive Grid was also mentioned by Kansas and CDL. CDL also mentioned EBSCO's discovery service, and WEST – the Western Regional Storage Trust (a Mellon-funded project creating “a shared print repository service, focused initially on retrospective journal archives.”)

Summary

From these discussions with selected institutions, the group found that many of our concerns and challenges are shared by our peers. Though everyone the group spoke to was interested in streamlining discovery and in bridging gaps between information silos in user-centric ways, a number of considerations may lead different institutions to different solutions. These considerations include:

- The nature of the institution’s user community
- The economics of both building and maintaining systems, including the need to incorporate sustainable design in a highly customizable environment..
- Nature of the institution (research versus four-year undergrad focus; whether there is a large online component to instruction; broad-based curriculum versus focused programming; private vs. publicly funded, etc.)
- Role that the institutions see for Google, HathiTrust and other initiatives in their planning
- The breadth of materials and formats in the institution’s collections
- And finally, the need for buy-in among library staff as well as more broadly within the institution.

Internal Data Sources

***Objective:** Identify the Libraries’ primary data sources and determine which of our resources can and should be accessible to external systems and emerging interfaces. Identify Libraries collections that cannot be made adequately discoverable through external systems.*

One of the foundational recommendations from the Phase 1 Discoverability study was to share unique metadata and holdings from local systems to external systems in order to make University of Minnesota Libraries’ resources more discoverable by searchers. To that end, the Internal Data Sources subgroup developed an extensive inventory of locally-produced primary data sources (Appendix 4) from several of the U of MN campuses. Examples included resources such as:

- MapHappy, in which MNCAT records are enhanced with geospatial data and pushed to Google Maps
- East Asian Film Digitization and Subtitling Projects - which include two separate projects described individually in the Internal Data Sources spreadsheet - in which we digitize and subtitle films in our collections and push them to the College of Liberal Arts' Digital Content Library.
- UMD's Digital Commons, hosted in a UMD dSpace instance, which uses Dublin Core metadata to describe items.

Data Attributes

The group identified a range of characteristics to use to describe each data source. These characteristics included:

- general description of the collection
- whether the source is currently maintained and by whom
- what kind of metadata structure is used; use of controlled vocabularies
- whether the described object is also available online
- whether the data source is already exposed and/or pushed to external systems

The subgroup distinguished between data sources that have some form of online presence (either a digital surrogate or descriptive metadata) from those sources that have not yet been described. For those data sources that are online and could theoretically be made (more) discoverable now, the group created brief profiles, also available in Appendix 5. Examples of these sources include the metadata for materials from the Libraries which are being digitized via the Google Book Scanning Project.

An example of an undescribed data source is the UMD Oral History Interviews. The interviews have been conducted and recorded, but haven't yet been placed online or had metadata applied to them. In addition, the Libraries Technical Services department maintains a list of uncataloged collections within the University Libraries at <https://wiki.lib.umn.edu/TS/CurrentProjects#uncat>.

In general, internal data sources either use one of a few standard metadata schema - MARC, EAD, VRA and Dublin Core - or some form of non-standard metadata. Standard catalog data is stored in MARC format, while instances of dSpace, the University Digital Conservancy and AgEcon, use Dublin Core. Everything in the DXLS instance uses EAD. Many of the digitized visual resources use the Visual Resources Association standard.

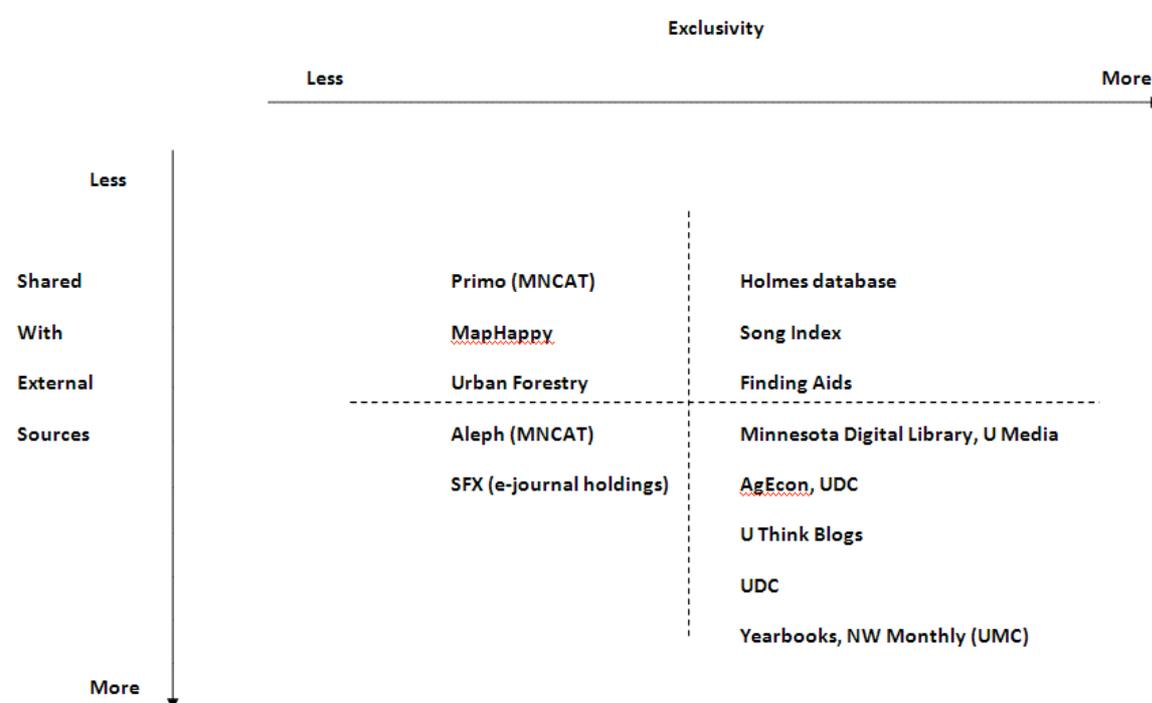
Many of the ad hoc internal data stores developed over time to meet different kinds of needs and did not follow any standard descriptive structure or guidelines. Even in cases where standardized metadata wasn't used initially, the group encouraged consideration of standardizing this data to facilitate sharing if the content is still viable and valuable, and the project suggests a positive return on investment.

The group also encourages evaluation of internal data sources already described with standardized metadata. For example, the University Libraries on the Twin Cities campus will be moving away from dSpace/Dublin Core for AgEcon and the University Digital Conservancy to Fedora/MODS over the next year. In this case, it has already been decided that shifting to Fedora/MODS would meet users' and Libraries needs more effectively.

Current Practices

The group did not attempt to evaluate every local system, but identified and collected data about each system described, including the extent to which it contained unique data and whether or not it was already being shared or exposed. Local sources were inventoried on a spreadsheet. However, the spreadsheet obscures the relationship of exclusivity to extent of exposure for local sources. Nor does it provide it provide room for discussion of local sources. Thus we developed this grid of exclusivity to extent of exposure to help clarify where certain sources fall. The sources noted in this diagram are intended as examples, and are not comprehensive. The idea of “sharing” as defined in this diagram, describes the discovery dimension only. The group identified other related dimensions (not shown) that could also be mapped, including the Delivery dimension and Manipulation dimension (the extent to which objects can be manipulated if it accompanies the metadata).

Dimensions of Exclusive and Shared Local Data Sources for Discovery



The Holmes database is exclusive to the University Libraries, but it’s not readily exposed to external aggregators while MNCAT content is often not at all exclusive to the University Libraries, but is probably our single most exposed local data source. This should come as no surprise - catalog content is created in accordance with internationally followed standards. Additionally, resource sharing tools like OCLC’s WorldCat only work if participating libraries expose their own work and vice versa.

At the same time, we do have good examples of exclusive local data sources which are also highly exposed. Notably, these data sources, such as AgEcon, the University Digital Conservancy and the University of Minnesota - Crookston Yearbooks, reside on the dSpace platform which comes optimized for Dublin Core.

In addition to sharing data with external systems, there is a recognition that data from local systems need to be available for re-use by other local systems. There is some evidence of overlap and sharing of data between local systems already, noting that UMedia, in particular, is a local aggregator of valuable Libraries content such as photographs from the Libraries' many archives as well as digitized objects from the collections such as historical maps from the James Ford Bell Library. In these cases, users can discover the resources via the collection's website or from the UMedia repository. However, the group is not aware of any standard guideline followed by University Libraries staff for setting up a local data store that includes a step for considering how to share the data.

The group recommends that the Libraries consider a full analysis of each data store for its potential to contribute exclusive data, offset by cost and other factors. This work should include a review of each source in conjunction with the steps identified in the Best Practices section below. In many cases, the value of exposing the data does not rise to the level of effort that would be needed to meet many of the recommendations in Best Practices. So the Libraries should be selective and look to identify high yield pay-offs, rather than engage in a comprehensive effort.

External Aggregators

***Objective:** Investigate and make recommendations for how the Libraries can best utilize large data aggregations in offering discovery services to our users. Identify external systems to be targeted in our efforts to expose our metadata and holdings and discuss how to accomplish these objectives:*

- Consider the potential impact of availability through a given resource as well as the effort required on our part

- Review the process of fulfillment for all of the key external search interfaces.

Make recommendations for optimizing fulfillment, if needed.

A growing number of external data aggregations offer institutions and individuals the opportunity to share their data resources with the larger community. The University of Minnesota participates in some of these efforts, and will likely participate in more of them in the future. Some of these participatory relationships are known to and mediated by the University Libraries. Others may be driven more by departmental, disciplinary, or individual preferences, with varying degrees of prominence within its community of users. In addition to participating via contributions, resources discoverable in selected external data aggregations should be available for scoping as laid out in the Vision above.

External data aggregators fit in to several categories:

- General metadata aggregations (examples: WorldCat; Google Scholar; OAIster; Primo Central; Google)
- General data object aggregations (examples: HathiTrust; Wikipedia; Internet Archive; Flickr Commons)
- Disciplinary aggregations (examples: AgEcon Search; ArXiv (for physics); Earth Prints)
- Form aggregations (examples: Digital Dissertations; MERLOT (learning objects); ArtSTOR; ArchiveGrid)

- Topical aggregations (examples: Minnesota Reflections; EthicShare)

When deciding which aggregations the libraries should contribute to and which should be featured for our own users, several factors need to be considered:

- **Ease of access.** Can the aggregator's contents be freely accessed, or is some form of authentication required?
- **Performance.** Can the aggregator's contents be searched successfully in a variety of modes (by words in full text; by author/title/subject/etc. keywords; for a summary view or a more granular view of contents; for known items)? Can results be ordered in appropriate ways (relevance rank, date, etc.)? Can useful views and versions of metadata (e.g., facets and term browse lists) and data objects (e.g., thumbnails and full size views) be consistently retrieved?
- **Coverage.** Does the aggregation include most of the content items that a user would expect, or at least a sufficient and representative number, or are there gaps and exclusions which might disappoint users?
- **Ease of contribution.** Can metadata and/or resources be contributed automatically via a protocol like OAI-PMH, or is more focused effort required?
- **Standardization.** Does the aggregator accept resources and data files in a variety of formats and standards, or must the contributing institution adhere to a particular set of standards?
- **Adoption.** Has the aggregator been adopted as a primary aggregation within one or more of our user communities?
- **Cost.** Is the service free, or are there costs to contribute metadata or resources, or to search and retrieve them?
- **Rights.** Does the aggregator manage rights responsibly? Are there significant limitations on the use that may be made of the aggregated resources?
- **Preservation.** Does the aggregator offer a reliable commitment to preserve contributed objects?
- **Maintenance.** Does the aggregator engage in processes to ensure currency and accuracy of the metadata and data objects in its care? Can contributing institutions make corrections to their contributions?
- **Secondary distribution.** Are metadata and objects contributed to an aggregator harvested by other aggregators for wider exposure and use?
- **Impact on local services.** Will exposure in an external aggregator promote increased use of local services? Can local units absorb the expected increase in demand?
- **Local interest.** Will the aggregator serve the needs of a particular local interest group? Will it provide information about the local setting or institutional collaborations?

Data Aggregation Categories

This factor analysis of the components by which different aggregation services should be judged can be correlated with the aggregation categories identified earlier to highlight some of the differences among these aggregation types. As discussed below, different factors may weigh more heavily in the case of different types of aggregators. As a basis for this analysis, the study group for external aggregators assembled a list of high profile and representative services which aggregate metadata and data objects (see Appendix 6). A separate spreadsheet (Appendix 7) indicates the categories represented by each of the profiled aggregation services.

General metadata aggregations

Two trends in this type of aggregation deserve note. One trend is the emergence of vended web-scale discovery systems which aggregate metadata from many sources into a single indexing and searching environment (e.g., Primo Central and Summon). The other trend is the integration of smaller aggregations into larger, established general aggregations (e.g., the merging of OAIster and Networked Digital Library of Dissertations and Theses into WorldCat). Both trends demonstrate the strong appeal of very broad and inclusive discovery environments. It is also interesting to note that both OAIster and NDLTD are form-based aggregations, which may be proving less independently sustainable than disciplinary or topical aggregations.

General data object aggregations

General data object aggregations are distinguished from metadata aggregations by the fact that they contain digital versions of the objects that their metadata records describe. This can enable them to offer full text indexing and other search options beyond those possible for metadata aggregations. It may also place a greater responsibility for good preservation and rights management. Ease of access to the digital objects also becomes crucial. (Examples: AgEcon Search; Digital Dissertations; Wikipedia)

Disciplinary aggregations

The extent to which a disciplinary aggregator is adopted by scholars in its target field should clearly be a crucial factor. The more scholars in a particular discipline come to rely on an aggregator for scholarly communications, the more of a hub the aggregator will become. In many cases this adoption will indicate successful or at least satisfactory performance on several of the other measures—ease of access, performance, and ease of contribution, and coverage. Other factors, such as costs, rights management, and preservation, may figure less prominently in scholarly communities' adoption behavior. When evaluating a disciplinary aggregator, libraries should weigh the importance of these other factors to the libraries' mission, and seek ways to ensure that they are also taken into account. (Examples: AgEcon Search; ArXiv (for physics); Earth Prints)

Form aggregations

Coverage should be a crucial factor for form aggregations, and a problematic one. The synergies among scholars that drive disciplinary aggregators are lacking when it comes to aggregations based on form, and success may depend more on the decisions and actions of libraries, archives, and cultural heritage institutions. Coverage is a likely indicator of ease of contribution, but may not indicate success on the other factors. Performance and preservation should also be carefully evaluated, and good secondary distribution may have a higher value than would be needed from a disciplinary aggregator. (Examples: Digital Dissertations; MERLOT (learning objects); ArtSTOR; NDLTD)

Topical aggregations

“Topical” is defined here as a focus on a narrower range of specified interest than would be the case with disciplinary aggregations. Thus, while a disciplinary aggregation might take philosophy as its focus, a more narrowly focused aggregation like EthicShare would be categorized as topical. Aggregations based on assembling information about a particular place or person or corporate body would also be categorized as topical. Because of their

narrow focus, local interest should be an important factor in weighing participation in and presentation of topical aggregations. (Examples: Minnesota Reflections; EthicShare)

Also noted: Entity/vocabulary registries

Entity/vocabulary data services aggregate the contributions of participating institutions and individuals to provide greater standardization and web-accessible identification of entities and concepts. These include both traditional name and subject authority files (LCNAF/LCSAF), emergent identity registries (ORCID, ISNI registry), and any entity-defining source with structured data (e.g., Wikipedia). By leveraging the data in these services with linked data models for declaring semantically defined relationships between resources on the one hand and entities and concepts on the other, greater synergy can be achieved and discoverability generally can be expanded. Contributing to linked data stores and the semantic web could enable wider access to local resources.

DBpedia can serve as a case study for the possibilities inherent in open registries. DBpedia is a harvested store of linked data statements based on typed or recognizable relationships present in Wikipedia. For example: Wikipedia includes an article on the University of Minnesota. It also includes an article on Ken Mauer in which Mauer's "Education" is expressed by a link to the University of Minnesota article. From this, DBpedia generates an RDF triple on Mauer's DBpedia page (http://dbpedia.org/page/Ken_Mauer) linking him to the University, and on the University's DBpedia page (http://dbpedia.org/page/University_of_Minnesota) adding Mauer's name to a list of names identified as having received their education at the University.

DBpedia's store of RDF triples is one of the most heavily used sets of linked data in the linked data environment. As such, the URIs derived from Wikipedia/DBpedia and used in other sources to represent the University of Minnesota in online descriptions of other entities expand the use of the Wikipedia data relationships well beyond their use in Wikipedia. As part of managing discovery, the University of Minnesota might want to take a more proactive approach to managing the record of its relationships being derived from Wikipedia. A similar case could be made for other emerging registries in the linked data environment, e.g., the emerging International Standard Name Identifier registry; but DBpedia is the most prominent source of this type of data at this point.

Summary

Just as with any other category of resources provided by the library, a variety of external aggregators is necessary to meet a diverse set of user needs. Selection of particular systems or datasets for the Libraries should be based on a cross-analysis of the type of aggregation and criteria for participation. It remains to be seen which aggregations will prove to be independently successful, and which will lapse or be integrated into larger discovery environments; but it is likely that both general and specialized aggregations will continue to be useful to the Libraries' user communities.

It should be noted that the team did not complete one of the objectives in the charge, to "review the process of fulfillment for all of the key external search interfaces [and] make recommendations for optimizing fulfillment, if needed." Lacking the capacity for in-depth examination of specific external systems and aggregators, the group did not feel able to make

such recommendations. We suggest that this objective be carried forward and addressed whenever the Libraries choose to participate in specific aggregators.

Recommendations and strategies for exposing our local data to these external systems follow in the Case Studies and Best Practices sections.

Case Studies

It's one thing to recommend that local data sources follow best practices and be exposed to other systems for use. It's another to think about what that would really look like. During the course of its work, the group realized that there is not necessarily a hard and fast line between internal data sources and external aggregators, nor are even the most exposed data sources exposed completely. Below, the group presents several case studies to illustrate opportunities for sharing data with external systems, re-using internal data locally and pulling external data sources into local systems.

To help the deliberation process, the group used the table below to organize the relationships between data sources, discovery environments and users. The table shows how a given mechanism operates on a source/environment pair and describes the intersection of users and content. For example, MNCAT as a local data source (the catalog) represents the intersection of local content and local users. Simultaneously, MNCAT also has an external discovery dimension in that MNCAT data is made available for use by external systems like WorldCat, placing it at the intersection of local content and global users. Case studies illustrating different aspects of the table follow below.

	Discovery Environment	
	Local	External
Internal Data Sources <ul style="list-style-type: none"> · Catalog · Finding Aids · Digitized resources · Locally-created databases 	<u>Mechanism:</u> Provide central portal(s) that eliminate silos <u>Intersection:</u> Local Content, Local User	<u>Mechanism:</u> Push out/open up <u>Intersection:</u> Local Content, Global User
External Data Sources <ul style="list-style-type: none"> · Externally-managed metadata, e.g., journal articles · HathiTrust API · WorldCat link 	<u>Mechanism:</u> Pull/Harvest, Load, Integrated Search, Separate Search. Could be one-time or ongoing, ingested records, or data accessed via an API, or provided as a separate search interface. <u>Intersection:</u> Remote Content, Local User	<u>Mechanism:</u> University Libraries works with external entities like HathiTrust to develop standards and provide governance. <u>Intersection:</u> Community Content, Community Users

The first case study shows how we have yet to fully capitalize on our own internal tools by showing the points of tension with digitized projects from the East Asian Libraries which are stored in a College of Liberal Arts database and inconsistently represented in MNCAT.

The second case study describes an instance in which the same local data source - archival finding aid documentation - is exposed, but in a different way to to different tools, some of which are local, like MNCAT, and some of which are not, like OCLC's ArchiveGrid.

The third case study discusses metadata that exists in the University Libraries MNCAT catalog, that has been marked for local use only (i.e., should not be exported to other databases). This example shows an ongoing need to regularly review choices around exporting, as restrictions change over time and as new contracts are negotiated. Records not exported from MNCAT are stored locally, but they aren't always locally created (for example, many records are purchased from vendors) and that difference plays a role in why they haven't been exported.

The fourth case study describes a technical opportunity to disclose catalog holdings to search engines, such as Google, via the Primo system. More needs to be learned about how the implications of such a choice, but the example illustrates another step toward enabling nontraditional access points to Libraries holdings, which has been a key message from the local Discoverability work. However, some of the same issues relative to non-exported MNCAT records are at play here, as well as an open question about whether this would be an effective tool for our users. Finally, our decision on an issue like this would most likely have to be made without knowledge of how Google might change its search algorithms.

The fifth case study addresses a particularly new issue: as the University Libraries participates more in multi-institutional projects which include digitized objects to which the University Libraries may not have local access, the University Libraries have the option to begin providing our users with records of content the University Libraries don't own, provided those records can be tied to delivery options.

1. Capitalize on Local Tools

The East Asian Library sponsored a project to digitize and add English subtitles to a collection of 13 classic Chinese and Japanese films owned by the University Libraries. The films are cataloged in MNCAT, but there is no link from the catalog record to the digitized version, which is currently available only via the College of Liberal Arts Digital Content Library (a local data source at the Twin Cities campus, but not one managed by the University Libraries). There isn't even a note that a digitized version is available.

Presumably the digitized versions of the films will eventually also be available via the Libraries UMedia Repository, but right now the only way to find these resources is to search the Digital Content Library or to follow a link from the East Asian Library web page.

Even if these films were available in the UMedia Repository, the current Libraries web site implies to the user that it is possible to search for all video materials from the "video and media" tab. However, at this time the search for videos on the Libraries website "video and media" tab only searches MNCAT, not UMedia. Many video resources owned by the Libraries and now being digitized have never been cataloged in MNCAT and are only available via UMedia.

The problem could be addressed in a variety of ways - from adding links to existing MNCAT records to moving the digitized films to the UMedia Repository to revising the search function on the Libraries website. This is an example of a case where silos based on format need to be addressed, and where the mechanism that provides delivery of the resource (i.e. the digitized version) is disconnected from the discovery mechanism.

2. Consistent Exposure of Valuable Content

Archival finding aids at the University of Minnesota are encoded in XML using EAD, then made available for searching via DLXS. Unfortunately, DLXS does not effectively expose finding aid data to Google. Archives units frequently resort to mounting redundant HTML versions of finding aids on their individual unit web sites. While this is an effective strategy for ensuring discovery via external search engines such as Google, it has undesirable side effects. For example, it results in potentially confusing inconsistent displays and branding of finding aids that vary from unit to unit. It also limits the ability to link or connect users to archival collections in other units, since typically such a link takes the user to an environment that only has finding aids from one unit.

Archival collections that have been cataloged in MNCAT are exposed to external search portals via WorldCat, but the majority of archival collections have not been cataloged. (Note: A project completed in 2009 batch-created about 1200 records from EAD to MARCXML to MARC, but approximately 3000 EAD finding aids have no corresponding MARC record.) The University of Minnesota also contributes its EAD finding aids to OCLC's ArchiveGrid. However, ArchiveGrid harvests the XML files and displays them using the stylesheets supplied by the individual archival units, rather than the stylesheet developed for DLXS, so display is inconsistent across the various archival collections that are part of the University Libraries.

Again, there are numerous options for resolving these issues such as fully cataloging finding aids, modifying or moving away from DLXS or revising finding aid stylesheets harvested by ArchiveGrid.

3. Records in the general catalog that are not shared

Some bibliographic records in Aleph are marked with a code that identifies them as needing to be excluded from exporting to other systems - the 965\$aNOEXP flag. This designation is followed by a \$b code that further defines the collection. Over half a million records are currently flagged for non-distribution, primarily due to contractual restrictions. However, the list below shows some opportunities presented by changes in the data itself, in the status with the vendor, or in the use of the external system. The University Libraries export records representing new holdings regularly to OCLC and metadata to Google and HathiTrust on a project basis. Given the shift from viewing OCLC's WorldCat primarily as a resource-sharing environment to a discovery environment for local users (featured on the Libraries home page), the logic behind not contributing records for electronic resources may not apply any more.

Code (\$b)	# of records in MNCAT	Description	Contract Information
CIS	53,828	Congressional Information Sources (Law Library microform collection)	According to Law Library, there are restrictions in the license.
CJK	1415	Records with Chinese, Japanese, and Korean vernacular characters. The master record used to exist in RLIN.	NA. These records should be cleaned up and records available for export.
EVANS	37,374		Restricted by NewsBank license.
HEIN	9		These records are probably errors that need to be cleaned up
HEINLC	204	Hein Legal Classics (online)	According to Law Library, this contract prohibits the uploading of records
HEINWT	1452	Hein World Trials (online)	According to Law Library, this contract prohibits the uploading of records
HIS	3320	ACLS History ebooks	There is nothing in this contract that restricts exporting these records
LICREC	15,570	"License Restricted" for Early English Books Online ebooks	Redistribution is restricted by contract.
MM	77,036	"Major Microform"	Records were part of WorldCat Collection Sets, which prohibits distribution to third parties. UMN holdings are not uniformly set in WorldCat.
NE	1657	Records from the ARTFL Project; converted from NOTIS	No contract was found that would prohibit sharing these records.
NET	13,173	NetLibrary records	It is possible that the original contract with NetLibrary discusses what the Libraries can do with records. However, the current license with OCLC and Baker & Taylor does not mention records. The records provided by YBP are not set to "No Export."
SFX	85,986	e-journal records exported from SFX and enriched from MARCIt	There is no legal restriction on exporting these records, although the dynamic nature of this collection presents issues. They were initially excluded from the OCLC export because licensing didn't permit ILL for most resources. There may also be significant issues with accurately matching MarcIt records to OCLC records as a result of the diverse sources and levels of detail in MarcIt records.

UMDAILRC	108	Duluth records from the American Indian Learning Center	Setting these records as unavailable for export is a local decision. The Duluth Library did not want these resources to be available in OCLC because they are not owned by the library.
USDOC	276,288	Government Document records from MARCIVE	Library is prohibited from transferring records to other entities by contract. However, MARCIVE is now setting new holdings in OCLC for the Libraries.
	<i>ca 550,000 records</i>		

4. Holdings exposed via Primo Sitemap

Primo can generate sitemap files that expose local holdings to public search engines such as Google, Yahoo, and Bing. Holdings records would also be available to the university website search engine. Once a search engine discovers and indexes the sitemap file, the Primo holdings records have the potential to show up in user's search results.

For example, Katholieke Universiteit Leuven (K.U.Leuven) created a sitemap file for their Primo installation in June, 2010. The sitemap file has been indexed by Google, so their holdings can now be discovered in a standard web search. Figure 1 depicts Google search results that link the user to full record display page in the K.U.Leuven Primo site, which has been branded as "Limo".

Although sitemap files have the potential to make holdings discoverable by more users, there are a few noteworthy caveats. First, the question of legality arises, as with the preceding example of records not exported from MNCAT. Even though the original MARC metadata has been transmuted into the Primo PNX format, many of the original access points would still exist and be exposed. It is unclear if contracts allow for the automated sharing of data that has been transformed or enriched as part of a local process. Second, creating a sitemap file does not necessarily guarantee that a search engine will add the holdings to its index, nor can it affect the results ranking within the search engine. Furthermore, once a sitemap file is created, Primo has no control over how quickly the holdings are indexed (or re-indexed) by the various search engines. However, some search engines, such as Google, provide tools for submitting new sitemaps and monitoring the indexing progress. Also, UMN relies on a Google Search Appliance for domain-wide web searching, and unless we made specific arrangements, the appliance would also pick up the sitemap, potentially injecting a quantity of records into the site search that could render it useless because library results would dominate all University of Minnesota results.

The group recommends further evaluation of this opportunity.

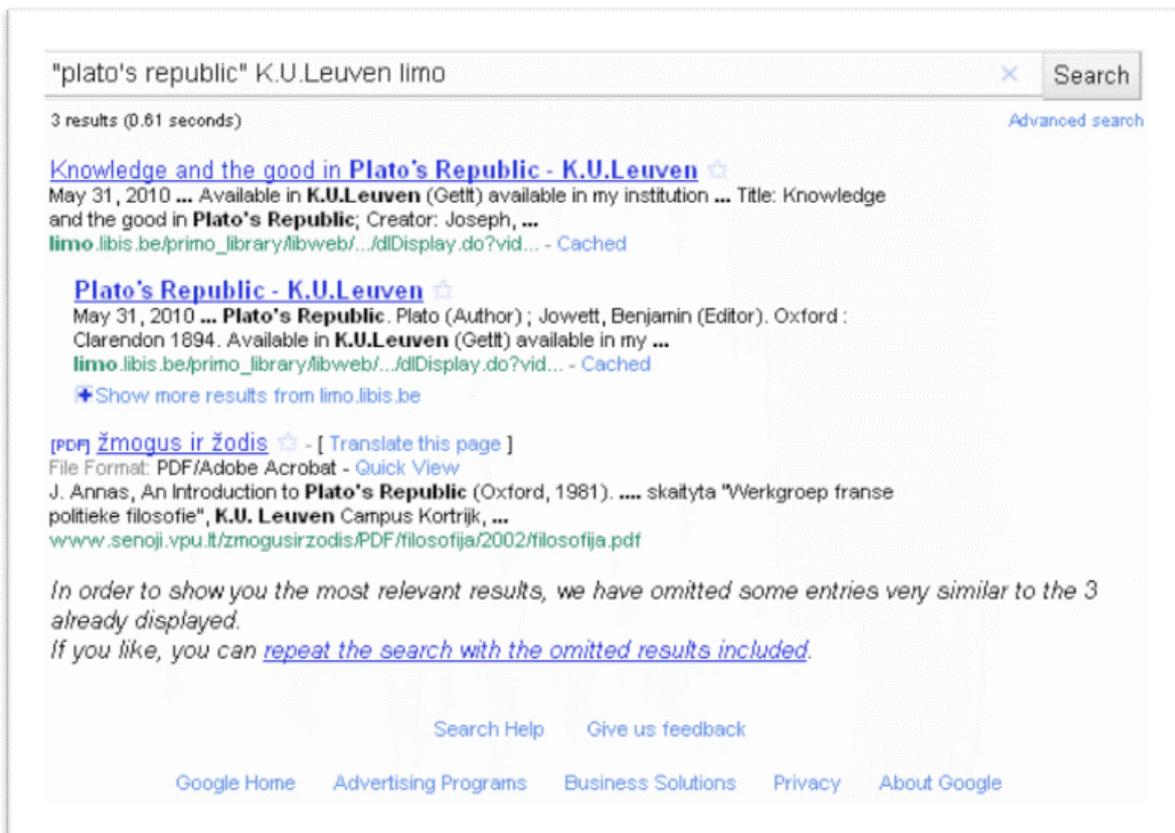


Figure 1: Google search results listing holdings at K.U.Lueven

5. Enhance discovery of relevant resources not locally owned or managed

Several recent University of Minnesota Libraries staff groups have explored the issue of selecting, describing, and providing access to quality materials that are available online or via interlibrary loan and which may meet a user's information need, but which are not owned by the university in physical format.

A recent UMN report on "Linking from MNCAT to Digital Surrogates in HathiTrust" discusses this problem in the context of HathiTrust materials, and analyzes the feasibility of different options to give users access to content available in HathiTrust. The current configuration enables users to discover materials owned by the University of Minnesota Libraries and link to the digitized version in HathiTrust (full-text or snippet) via the HathiTrust Bib API. However, this configuration requires that a bibliographic record describing the material already exists in MNCAT, limiting discovery to only those materials that were contributed to the project by the university.

The materials in HathiTrust represent quality collection development choices by the university's peer institutions, suggesting that discovering these resources would also be beneficial to UMN searchers, assuming that the delivery requirements for a particular user's information need can be adequately met.

Taken together, these case studies illustrate the relationships between

- local tools such as MNCAT and the UMedia Repository,
- local data such as MNCAT records or finding aids,
- local discovery interfaces such as the Libraries websites,
- external data such as commercially produced records imported into MNCAT, and
- external tools such as OCLC's ArchiveGrid and the UMN-TC College of Liberal Arts Digital Content Library.

Best Practices

Objective: Codify best practices for exposing our data.

The first phase of the Discoverability initiative demonstrated that use of library resources does not always originate with library managed services. Libraries need to leverage this awareness by identifying ways to enlarge and enhance the libraries' presence in external discovery environments. The Libraries should devise innovative ways to expose and optimize the discoverability of library resources in these systems and increase the visibility of library contributions to these systems' users.

The best practices toward exposing the Libraries' metadata that are proposed below are organized under three headings: metadata, systems, and community. Distinct kinds of policy and process decisions are needed in each of these areas. Because discoverability requires the expenditure of resources, there is a need to make trade-offs and to prioritize. Decisions about discoverability services are practical only to the extent that library resources can sustain these efforts.

Metadata

Many data formats, content standards, software platforms are currently in use by the university's wide and diverse user population, and the expectation is that these factors will change over time. Defining best practice cannot consist of selecting a single optimal model - it must address the need for flexibility in data presentation, transformation and communication as a primary criterion.

Description. Descriptive metadata is intended to make the object recognizable to users and to provide a basis for assessing its potential value. To ensure this goal is achieved for a large user population, descriptive metadata should be delivered with all necessary contextual information. This can mean ensuring that collection level information appears at the item level; the spelling out of abbreviations and avoidance of jargon; providing information about item extent and format; and tracing the object's relation to other information resources. Some elements of descriptive metadata (e.g., title, form, date) are commonly selected for brief representations of an object. For best practices, these critical descriptive elements should always convey useful information, either transcribed from the object or supplied by the metadata creator. Currently, this is not always the case.

Access. Some metadata elements are intended to be access points, i.e., to assist users in locating relevant objects through searching. Access points function better when they include a uniform data string for retrieval. This can be either a resolvable URI or a text string of some

kind, or both. Uniformity of access data, often ensured through reference to a controlled vocabulary, enables users and systems to retrieve records more precisely. Controlled vocabularies can also facilitate discovery via synonyms as well as preferred terms, and guide summary-level navigation of a collection, surfacing relationships which the metadata of individual objects will not reveal. Facilitating the searching and navigation of controlled vocabularies themselves may also be advisable in some situations to take full advantage of their benefit.

Relevance ranking. Relevance is generally expressed by the ranked order of objects in a result set. The numeric rank value is determined algorithmically based on weighted factors, often derived from data in and associated with the metadata object. Relevance can be assessed based on a combination of many different factors, e.g., the presence of the user's search term in an object's metadata or full text; the presence of the search term in a controlled vocabulary cluster; transaction-based evidence of relevance to a search term as judged by other web objects' links and other users choices; frequency of use; and so on. The process of making local metadata available to external systems should include a consideration of the elements that would facilitate optimal relevancy ranking by those systems.

Rights. Users need to be informed of any restrictions on the uses they can make of the objects they discover. Metadata can provide guidance about the rights status of objects and any steps needed to pursue particular types of use. In some systems, rights status is a criterion for filtering search results. Information about rights associated with an object should be included in any exported data. Optimally, this data would allow external systems to derive rights status in recognition that status is a dynamic attribute.

Metadata maintenance. As standards and controlled terminologies evolve, changes need to be reflected in metadata records that make use of them. Consistency with standards enhances discoverability by aligning metadata with external patterns of use and simplifying transformations of metadata by applications using other standards and vocabularies. Metadata should be kept up to date both in the home repository and made available or pushed to external sites.

Metadata exchange. When exporting or importing metadata, reformatting and other kinds of data translation may be required to make the data interoperable with other systems. Adherence to shared standards will facilitate and simplify data transformations. In some cases, offering multiple versions of metadata records based on standard transformations will be the most effective way of managing exchanges of discovery information.

Systems

Platform management. Exposing metadata requires that local systems interoperate with other systems. Local customization that hinders the ability to facilitate exchange of data should be limited.

Service continuity. As libraries move into creating and publishing unique digital objects for the scholarly community, they raise expectations and undertake an obligation to maintain discovery and access for these objects. The process of exposing data to external systems

should include a consideration of service continuity where discovery is coupled with expectations for delivery.

User authentication. While libraries prefer to provide open and unfettered access to resources, this is not always possible. Some kinds of access will be limited by licensing agreements or other considerations and will require users to authenticate their rights to the use the resource. Satisfaction with discovery and delivery services may thus depend on quick and efficient models for user authentication. Exposed metadata should include any relevant information that would assist a searcher with during the process of delivery, and should enable external systems to offer choices based on availability.

Community

Discovery community participation and development. The digital discovery environment offers diverse opportunities for participation and contribution by the Libraries. Collaborating with other service providers to maximize the exposure of local resources and local access to partners' resources will benefit both the campus and the larger community. Contributing innovative models for data, software, and processes to the larger community promises similar benefits. Community participation and development should effect to distribute local data where useful and to simplify the work of integrating community-level changes into the local discovery environment.

Open documentation. In order to make use of exposed metadata, other systems often need access to documentation about the local collection--metadata application profiles, controlled term mappings, and format options. Some of this information may also be of use to end users. Accurate and accessible documentation of library policies, practices, and data and service models will assist both end users and collaboration and development partners in making good use of the libraries' resources.

References:

Foulonneau, Murial & Riley, Jenn (2008). Metadata for Digital Resources.
Bruce, Thomas R. & Hillmann, Diane I. (2004). The continuum of metadata quality: defining, expressing, exploiting. *Metadata in Practice*, 238-256.
RLG (2005). Descriptive Metadata Guidelines for RLG Cultural Materials, http://www.oclc.org/research/activities/past/rlg/culturalmaterials/RLG_desc_metadata.pdf, viewed Sept. 21, 2010.

Conclusion

The vision of a new discovery environment, that surfaced from the work of the phase 2 Discoverability group, suggests that a synthesis of tools and services need to be coordinated in such a way to enable users to discover, access, and interact with relevant data from internal, external, owned, licensed, and freely-available data sources. This vision evolved as data was gathered and evaluated on the Libraries' user communities, the amalgam of systems and interfaces currently at play, and the options available for integrating and exposing data.

The interviews, conversations, and feedback from the peer institutions indicate the progress made by other institutions in the direction of integrated discovery environments. The Libraries' approach to developing its new discovery environment favors the needs and wants of the user communities and should be recognized as unique in comparison to the approaches of its peers. Next steps will, therefore, need to be inclusive of the broad user community: selectors/liasons, for example, may be consulted about external systems that might be included in scopes, while researchers may be able to advise on key datasets.

To invigorate the vision for a new discovery environment at the University of Minnesota, the Libraries will need to consider options and strategies for designing and coordinating the discovery environment that address the key concepts that emerged from the group's research and analysis: scopes; integrated metadata layer; exposing internal data to external systems; external data aggregations; object manipulation and personal curation; data-driven decisions; ongoing ability to review and develop the discovery environment.

The Vision of Discoverability Phase 2 and the recommendations from each of the sub-groups are intended to guide the next steps for the Libraries. Many of the recommendations suggest the need for further investigation in order to determine compatibility, value and impact. And most of the recommendations will require supplemental inquiry that will want to be reviewed in relation to results of information gathered on other recommendations. Some of these complexities and cross-analysis are directly referred to in the recommending statements. In addition to the Vision detailed above, the final recommendations from the Phase 2 Discoverability group that are interlaced throughout the report are noted, here, in abbreviated form:

- 1) Standardize data from the internal data stores to facilitate better sharing of the content to determine if the data is still viable and valuable.
- 2) Consider a full analysis of each data store for its potential to contribute exclusive data to external aggregations, offset by cost and other factors.
- 3) Generate well-described, standardized metadata for local collections and data sources that can be exported, harvested, or made available for crawling by external systems.
- 4) Select systems or datasets based on a cross-analysis of the type of aggregation and criteria for participation.
- 5) Carry forward with the objective to examine specific external systems and aggregators.
- 6) Remain responsive to user needs and make data-driven decisions by favoring systems that provide consistent data sufficient for measurement and analysis of user behavior and interactions throughout the discovery environment.
- 7) Develop a dynamic process for conducting ongoing reviews of the environment as a means to accommodate technological changes and the evolution of the discovery and delivery expectations. These reviews must encompass user needs, data mapping, relevance rankings, and source quality.
- 8) Evaluate the opportunity to use Primo for exposing local holdings to public search engines such as Google, Yahoo, and Bing; and that will also expose holdings records to the university website search engine.
- 9) Balance the potential gains from adoption of a discoverability service against the practical requirements of implementing and sustaining the service.

The list of concluding recommendations is clearly much more than a quick to-do list. There are many variables that need to be considered and variables that will need to be considered in correlation to determine potential and effectiveness. Initially, the Libraries may want to measure the fitness of particular systems and configurations against the vision, and revisit NOEXP records for possible output to external aggregators such as WorldCat. Establishing procedures for creating scopes for specific user communities and tasks would further enlighten other decision-making that will ultimately add value to the faceted future of the University Libraries' discovery experience.

APPENDIX 2: USER COMMUNITY PERSONAS

Affiliate Users

Undergraduates

Now that I'm a junior I work differently than I used to. Before I started focusing on my major I had a lot required classes that weren't very interesting, so I did what I had to do, but no more. Now though I really enjoy my classes and put more effort into my assignments. Some things haven't changed though. I'm used to using Google to look up stuff and it gives me good results, so I still use it most of the time. I might use library stuff more, but I always get confused on the library site and never get the results I expect. I have a lot things I have to do, so I don't spend a lot of time looking around for stuff. There's just so much stuff out there! I've found it's good to listen to my teachers about which resources to use because they point us to stuff that they say are what they want us to use. I forget how I found it, but Google Books that's pretty cool when it has the book I want in it.

From the literature

- More interested in broad subject-based access than other groups (Desktracker stats)
- Aware of qualitative difference between published research and general Internet sites. (Hampton-Reeves et al., 2009)
- Very reliant on library catalogs, databases, and staff advice (Hampton-Reeves et al., 2009).
- Perception of research led by assignments (Hampton-Reeves et al., 2009).
- Start research with course readings, followed by Google, scholarly research databases, and the catalog (Head & Eisenberg, 2009).
- Value efficiency and utility of research strategy over depth and completeness. “Conceptualize the information-seeking part of research as a practice learned by rote” (Head & Eisenberg, 2009).
- Increasingly using limited preview in Google Books rather than going to the library (Hampton-Reeves et al., 2009).
- Students who use discipline-specific databases have a better experience accessing research, but are then dependent on database holdings (Hampton-Reeves et al., 2009).
- When using a search engine, many students click on the first search results. In one study “[o]ver a quarter of respondents mentioned that they chose a Web site because the search engine had returned that site as the first result suggesting considerable trust in these services” (Hargittai, Fullerton, Menchen-Trevino, & Thomas, 2010)

From our colleagues

Kate Peterson, Information Literacy librarian

- I'm very skeptical of the statement that students are “very reliant on library catalogs, databases, and staff advice,” unless they are referring to information provided by the instructor in which case I do think students are reliant on what their instructors tell them or give them.
- I don't think the Google Books are yet showing up in the top few results in Google so students still aren't finding this, but I believe this trend is probably.

References

Hampton-Reeves, S., Mashiter, C., Westaway, J., Lumsden, P., Day, H., & Hewertson, H. (2009). Students' use of research content in teaching and learning. *Report for the Joint Information Systems Council*.

Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication*, 4. Retrieved from <http://ijoc.org/ojs/index.php/ijoc/article/view/636>

Head, A. J., & Eisenberg, M. B. (2009). How college students seek information in the digital age. *Project Information Literacy Progress Report*. Washington DC, Information School, University of Washington.

Graduate Students

Now that I'm a graduate student I spend a lot of time doing the same things that faculty do - practice, I suppose. Certainly my research is much more in depth than it used to be or that my students do in my classes. I'm so busy that in I have to fit my research into any gaps I can find - it's hard to sit down and really get organized. I've heard other people mention databases that they use from the library, but I don't know how they ever found them. There are so many choices from the library and, like I said, I'm working around lots of other responsibilities so I end up getting frustrated. It would be great if I could search lots of databases at once. That way I could really make the most of the citations I'm working from. It would also help me when my research spreads into other disciplinary areas or focuses on data I'm not familiar with. What I really need to do get my research published and published in the right place and I'd love it if there were some easy tool from the library to help me with that.

From the literature

- More interested in numeric data than other groups (Desktracker stats)
- Often work from/need help with citations (Desktracker stats)
- Identify themselves more closely with academic researchers than undergraduates (Hampton-Reeves et al., 2009).
- Often lack basic research skills and work by trial and error. Largely unaware of many library services offered (University of Minnesota Libraries, 2006).

From our colleagues

Kate Peterson, Information Literacy librarian

- Graduate students are often overwhelmed by the availability of information/discovery and don't know when to stop (or how to be systematic) and thus want tools that can, for example, search multiple databases together or the 'one library search' idea, at least as a way to improve efficiency. This is mostly a social sciences and interdisciplinary challenge.
- This point from the Life Sciences profile is also relevant to many other graduate students: "Work is increasingly collaborative. Obstacles include different vocabularies, not knowing core journals, and not knowing where to publish (Marcus, Ball, Delserone, Hribar, & Loftus, 2007)

References

Hampton-Reeves, S., Mashiter, C., Westaway, J., Lumsden, P., Day, H., & Hewertson, H. (2009). Students' use of research content in teaching and learning. *Report for the Joint Information Systems Council*.

University of Minnesota Libraries. (2006). A multi-dimensional framework for academic support. Minneapolis: University of Minnesota. Retrieved from http://www.lib.umn.edu/about/mellon/UMN_Multi-dimensional_Framework_Final_Report.pdf

Faculty

I really appreciate the effort that the library puts into buying resources for the university. While I do still spend some time browsing library shelves, I couldn't live without electronic journals these days. As a result I practically never go to the library anymore. In so many ways electronic formats make my life easier; especially when it comes to working with colleagues at other institutions. I wish that there were more tools specifically created to help support collaboration across institutions especially when collaborators are at places where the teaching/research focus differs. One thing that isn't easier with electronic materials is dealing with copyright. Wow. The rules really vary from what you can use to teach with to how you handle other people's work in writing to how you handle your own work. This is an area where I definitely could use some guidance!

From the literature

- More IP concerns (publishing, use of materials in class) than other groups (Desktracker stats)
- Across disciplines, increasingly comfortable with electronic content over print (Schonfeld & Housewright, 2010)
- Library's role as a gateway becoming less important to them (Schonfeld & Housewright, 2010).
- Value the library's role as a buyer most highly among Gateway, Archive, Buyer, Teaching Support, and Research Support (Schonfeld & Housewright, 2010).
- Faculty at research institutions have very different information seeking practices than faculty whose primary responsibility is teaching (E-journals: Their use, value and impact, 2009; Niu et al., 2010).
- Need tools that allow them to collaborate with colleagues at other institutions (University of Minnesota Libraries, 2006).
- Still consider shelf browsing "extremely" or "somewhat" important (University of Minnesota Libraries, 2006).

References

Schonfeld, R. C., & Housewright, R. (2010). Faculty survey 2009: Key strategic insights for libraries, publishers, and societies

E-journals: Their use, value and impact (2009). Research Information Network.

University of Minnesota Libraries. (2006). *A multi-dimensional framework for academic support*. Minneapolis: University of Minnesota. Retrieved from http://www.lib.umn.edu/about/mellon/UMN_Multi-dimensional_Framework_Final_Report.pdf

Non-affiliate Users

High School Students

I've only been on campus a couple of times, when I was looking for stuff for my History Day projects. I'd found a lot of stuff about my topic on the Web, but my teacher said I needed to have "primary source" stuff and maybe look at the books in the list at the end of the Wikipedia article I used. I wouldn't have known where to go or what to do if my teacher hadn't been there with people who work at the library. My friend who takes PSEO classes says he uses the library web site once in a while, when his professor tells the class what to look for.

From the literature

Personal interview with Amy West:

- History Day participants are likely the largest single group of high school users. Libraries staff members do some advance work to prepare. Preparing targeted online content and promoting it would be useful.
- History Day participants do very mediated searching with their instructors and library staff. Typically, students are looking for “traditional” library materials, archival collections, and non-archival collections such as government publications.

Other sources:

- Tweens “...strategies involved multiple sources, including interpersonal, media, and Internet search tools. Tweens thus employed a kind of information bricolage, gathering and assembling ready-at-hand information from varied persons and media in the course of a single problem, provided they were motivated to do so.” (Meyers, Fisher, & Marcoux, 2009)

References

Meyers, E. M., Fisher, K. E., & Marcoux, E. (2009). Making sense of an information world: The everyday-life information behavior of preteens. *Library Quarterly*, 79(3), 301-341.

Engineering Professionals

When I was a student at the U, we didn’t use the library much. We did use books with tables of properties of materials and characteristics of airfoils, though, and now those books or on the shelves at the library. Every once in while I need to look up something like that, and it’s easier to drive over to campus and pull the book. Once, a librarian at the desk showed me how to use some online source to look up the properties I need. I’d have to be on campus to use it, though, so I’d rather just look it up in the book I know.

I’ve also called the library to find out whether they have a certain standard before asking my company to purchase it for me. Those things are expensive—if it’s already here, there’s no point in us buying it again.

From the literature

No specific literature found. Points are culled from an interview with Tony Ihrig, Physical Science & Engineering Library reference coordinator.

- Most often looking for standards and tables of information.
- When on site, usually know exactly what they’re looking for and just need to verify that we have it and find it on the shelves.
- Occasionally seek help surrounding an invention: sources for manufacture of specific elements, standards, and patent searching.

Commercial Users

I know what I want and I often need it fast -- when I know the specific item I am looking for I just want to find and retrieve that piece. It is frustrating when I need to click on multiple links to narrow my search. . Sometimes I like to check the library’s catalog to see if they have the specific resource I need, but there are other situations where it is easier for me to send a topic and have InfoNOW do the initial research, present the results, and deliver the content (articles, chapters, books, etc.). My informational needs vary according to the breadth and depth of research that I am conducting, this means that I may need information from the library as well as archival collections. I don’t understand why I can find items in the library catalog and it takes a couple of days to receive the information.

From the literature

No specific literature found. Points below are culled from an interview with Todd Fenton of InfoNOW on June 18, 2010:

- Clients usually know exactly what they want (have citations from other sources, have already done database searches, etc...) and want to confirm that it is in our holdings. Other times they have a topic and expect us to do the searching for them, then request specifics from that request.
- May search Worldcat/OCLC or catalog to see if the UL has the item the information is in, but other times just send requests regardless if we have the piece.
- 1-2 requests a month for archival materials from Andersen; about 30 per month from MLAC – demonstrates need for materials not available in open stacks.
- Muddling of library collections/catalogs confuses users.
 - Do not understand that Crookston and Duluth are different (management and location) from the Twin Cities Libraries – see University of Minnesota and don't see the difference in the campuses.
- Googled discovery environment not what users want – need to be able to pinpoint information.

Archives

I've used archival material quite a bit over the years. I participated in History Day in middle school and then starting as an undergraduate I got interested in the history of science in general and now I'm a professor specializing in the history of computer science. Even though my needs and interests have changed over the years, I still find I need help from archivists all the time. Even when I get the system down at one archive, I find that this knowledge doesn't transfer to other archives. I love that archivists are beginning to make more information about what they've got available online, but I find I still need their help to understand even that information, let alone to get hold of the materials. I really wish I could use something like Facebook or Twitter to get help from archivists as the need arises rather than having to make appointments and have formal meetings.

From the literature

- Various user groups and skill levels (e.g. archivists, researchers, genealogists, etc.)
- Different approaches in understanding the structure of EAD finding aids -- need to provide a context for users who want to search for a specific item vs. those who want to browse (Nimer & Daines, 2008).
 - Context needed for navigation of finding aid and also for understanding entries and archival terminology (Chapman, 2010).
- Social navigation as a possible tool to inform users of recent acquisitions or discoveries (Nimer & Daines, 2008).
 - Social navigation also provides a way for archivists to propagate information
- Dynamic web environment to assist users in navigating finding aids (Chapman, 2010).
- Location of finding aid materials – need to clearly communicate to users how they can access an item(s) (Chapman, 2010).

From our colleagues

Ryan Bean, YMCA Archives

- The literature is spot on about the various user groups with diverse skill levels.
- I would also add diverse needs, (ie. Someone looking for a photo of a parent vs. a scholar researching their next book).
- The comment about Facebook/Twitter is appropriate for users with one-off questions, (ie. when did..., do you have...). However, since our materials don't leave the building and

scholarly researchers do need to visit if they are planning any sort of in depth research, I find the many technology points curious. I have never heard a researcher make those comments. What I have heard is that if they know we use those tools, they think it is cool. I would also add that in general, archive researchers know what they are looking for/hope to find.

References

- Nimer, C., & Daines, J. G. (2008). What do you mean it doesn't make sense? redesigning finding aids from the user's perspective. *Journal of Archival Organization*, 6(4), 216-226.
- Chapman, J. C. (2010). Observing users: An empirical analysis of user interaction with online finding aids. *Journal of Archival Organization*, 8(1), 4-30.

Health Sciences Disciplines

Clinicians

I need evidence-based, vetted, authoritative information while I am working with patients or elsewhere in the clinic setting. I rely on UpToDate and other point of need resources to quickly search for and find information. When I can't find something I find myself asking my colleagues to see if they have any suggestions, but this can be tough because of our very busy schedules. I rely on remote access to library resources since it is difficult for me to visit the library. I prefer to use resources that I can easily access on my smartphone or mobile device.

From the literature

- Access to resources and information at the point of need (bedside, clinic, hospital) for questions that arise from patient/client interactions— need based learning (Ebell & Shaughnessy, 2003; Slawson & Shaughnessy, 2005).
 - Asynchronous learning is very appropriate (McCann, Schneiderman, & Hinton, 2010)
- Filtered/distilled information/ vetted (e.g. evidenced-based information) (McCord et al., 2007; Slawson & Shaughnessy, 2005)
 - Need vetted/reviewed information
- Convenience, habit, and speed of use needed for successful information seeking (Dawes & Sampson, 2003)
- Information seeking skills based on information needs – foraging/”keeping up,” hunting/”just in time, and combining information evidence with patient-centered care (Slawson & Shaughnessy, 2005)
- Timeliness – need information in a timely manner (Green & Ruff, 2005)
- Consult with colleagues (Dawes & Sampson, 2003)
- On the spot search training/resources (Van Duppen et al., 2007)

References

- Dawes, M., & Sampson, U. (2003). Knowledge management in clinical practice: A systematic review of information seeking behavior in physicians. *International Journal of Medical Informatics*, 71(1), 9-15. doi:DOI: 10.1016/S1386-5056(03)00023-6
- Ebell, M. H., & Shaughnessy, A. (2003). Information mastery: Integrating continuing medical education with the information needs of clinicians. *Journal of Continuing Education in the Health Professions*, 23(Suppl 1), S53-62.
- Green, M. L., & Ruff, T. R. (2005). Why do residents fail to answer their clinical questions? A qualitative study of barriers to practicing evidence-based medicine. *Academic Medicine*, 80(2), 176-182.

- McCann, A. L., Schneiderman, E. D., & Hinton, R. J. (2010). E-teaching and learning preferences of dental and dental hygiene students. *Journal of Dental Education*, 74(1), 65-78.
- McCord, G., Smucker, W. D., Selius, B. A., Hannan, S., Davidson, E., Schrop, S. L., Rao, V., & Albrecht, P. (2007). Answering questions at the point of care: Do residents practice EBM or manage information sources?. *Academic Medicine*, 82(3), 298-303.
- Slawson, D. C., & Shaughnessy, A. F. (2005). Teaching evidence-based medicine: Should we be teaching information management instead?. *Academic Medicine*, 80(7), 685-689.
- Van Duppen, D., Aertgeerts, B., Hannes, K., Neirinckx, J., Seuntjens, L., Goossens, F., & Van Linden, A. (2007). Online on-the-spot searching increases use of evidence during consultations in family practice. *Patient Education & Counseling*, 68(1), 61-65.

Professional Students - Health Sciences

I need access to information at the point-of-need and from various settings (in the lab, out in the field, or in a clinic setting) in a timely manner. I often need information that I can apply to the situation at hand; for example I may be working on a case and need to verify something to support a decision or a conclusion. It is easier to remember something when I am able to reference resources on the spot, if I try to search for something later it is difficult for me to remember to details of the circumstance. I prefer to begin my searching in a centralized location; a place to find out where to find out is useful.

From the literature

- Learning information seeking skills for practice-based application (Slawson & Shaughnessy, 2005)
- Integration of resources to support and inform decisions (Nayda & Rankin, 2008)
- Remote access, nature of work/research means that the students typically are not in the library
- Access to information in clinic setting (or point of need) ideal for retaining and using of information (Green & Ruff, 2005)
- Point of need for information varies: formal clinic setting, but also more “field type” work.” Public Health is an area where this is a primary issue because of availability of online resources, deverticalization, inadequate resources available, etc. (Rever et al., 2007)
- Access to information at point-of-care/bedside (Green & Ruff, 2005)
- “... useful to have one key clinical question and to answer it right on the spot because you need to act fairly soon and you have like three other patients to see and you may not get back to that clinical question in a timely enough manner if you don’t have the resources right there to be able to pursue that question. “
- Centralized sources of information (Green & Ruff, 2005)
 - Need for centralized information service or “place to find out where to find out” (Rever et al., 2007)
- Need information in a timely manner (Green & Ruff, 2005)
- Timely and evidence-based information at point of need/bedside (McCord et al., 2007)
 - Integration of evidence-based resources into clinical setting (McCann et al., 2010; Yew & Reid, 2008)
 - Timely and vetted/accurate information (Nayda & Rankin, 2008)

References

- Green, M. L., & Ruff, T. R. (2005). Why do residents fail to answer their clinical questions? A qualitative study of barriers to practicing evidence-based medicine. *Academic Medicine*, 80(2), 176-182.
- McCann, A. L., Schneiderman, E. D., & Hinton, R. J. (2010). E-teaching and learning preferences of dental and dental hygiene students. *Journal of Dental Education*, 74(1), 65-78.
- McCord, G., Smucker, W. D., Selius, B. A., Hannan, S., Davidson, E., Schrop, S. L., Rao, V., & Albrecht, P. (2007). Answering questions at the point of care: Do residents practice EBM or manage information sources?. *Academic Medicine*, 82(3), 298-303.
- Nayda, R., & Rankin, E. (2008). Information literacy skill development and life long learning: Exploring nursing students' and academics' understandings. *Australian Journal of Advanced Nursing*, 26(2), 27-33.
- Rever, D., Tuner, A. M., Madhavan, A., Rambo, N., Bugni, P. F., Kimball, A., & Fuller, S. S. (2007). Understanding the information needs of public health practitioners: A literature review to inform design of an interactive digital knowledge management system. *Journal of Biomedical Informatics*, 40, 410-421.
- Slawson, D. C., & Shaughnessy, A. F. (2005). Teaching evidence-based medicine: Should we be teaching information management instead?. *Academic Medicine*, 80(7), 685-689.
- Yew, K. S., & Reid, A. (2008). Teaching evidence-based medicine skills: An exploratory study of residency graduates' practice habits. *Family Medicine*, 40(1), 24-31.

Researchers/Fellows - Health Sciences

I need timely information delivered to my setting fast. My preferred resources for information are: journal articles, web pages, databases, and personal communication. Sometimes I will use conference proceedings, but not as often as the other resources. Because of the research I am conducting I need cross disciplinary information that is available in a variety of resources and formats.

From the literature

- Learning information seeking skills for practice-based application (Slawson & Shaughnessy, 2005)
- Timeliness – need information in a timely manner (Green & Ruff, 2005)
- Most used resources (in order): journals, Web pages, databases, and personal communication. Conferences and proceedings used less often (Hemminger, Lu, Vaughan, & Adams, 2007).
- Crossdisciplinary information and/or varied resources and format (Rever et al., 2007)

References

- Green, M. L., & Ruff, T. R. (2005). Why do residents fail to answer their clinical questions? A qualitative study of barriers to practicing evidence-based medicine. *Academic Medicine*, 80(2), 176-182.
- Hemminger, B. M., Lu, D., Vaughan, K. T. L., & Adams, S. J. (2007). Information seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*, 58(14), 2205-25.
- Rever, D., Tuner, A. M., Madhavan, A., Rambo, N., Bugni, P. F., Kimball, A., & Fuller, S. S. (2007). Understanding the information needs of public health practitioners: A literature review to

inform design of an interactive digital knowledge management system. *Journal of Biomedical Informatics*, 40, 410-421.

Slawson, D. C., & Shaughnessy, A. F. (2005). Teaching evidence-based medicine: Should we be teaching information management instead?. *Academic Medicine*, 80(7), 685-689.

Other Disciplines

Arts and Humanities

In my area, a digital tool has to really be worth using to justify the time to learn about it. The tools I value help me with databases and chasing down citations in works I'm already using. Much of my work involves working with journals and monographs in a really deep way. Catalog records and the tools that help with them only go so far - some of my best research finds come from serendipitous discoveries on library shelves. I also use primary sources lots. However, here, I work with archivists quite a bit because you can't browse archives like you can libraries.

From the literature

- Rarely consult with librarians, but often with archivists – presumably due to importance of archival materials in research (Barrett, 2005).
- Value serendipity in information seeking; find catalog records inadequate thus like to browse shelves (Barrett, 2005).
- Use databases at the beginning of the research process; citation chasing from middle to end (Barrett, 2005).
- Archival research is increasingly important, and “challenging due to the idiosyncratic organization of archives and the range and variety of materials housed within them” (University of Minnesota Libraries, 2006).
- Slower than others to adopt digital tools like citation management systems, slow to seek assistance (University of Minnesota Libraries, 2006).

References

Barrett, A. (2005). The information-seeking habits of graduate student researchers in the humanities. *Journal of Academic Librarianship*, 31(4), 324-331.

University of Minnesota Libraries. (2006). *A multi-dimensional framework for academic support*. Minneapolis: University of Minnesota. Retrieved from http://www.lib.umn.edu/about/mellon/UMN_Multi-dimensional_Framework_Final_Report.pdf

Visual Arts

I use the library for a couple of different things: Sometimes when I'm doing work for a class, I need to do research on a particular artist or the art of a certain period. Then I'll search the catalog for books or ArtSTOR images or videos for that person or period or whatever. But when I'm creating my own art, I use the library in a totally different way. I'll go to the library and just start walking down the aisles looking for something that catches my interest. Or if I think I might want to capture something about, for example, the sounds of city life, I might look for reports on noise in an urban area that have figures that might inspire me.

From the literature

- “Compulsive browsers” (Pacey, as cited in (Hemmig, 2008))

- Besides art, interested in resources related to literature, history, geography, biology, and children's literature. (Hemmig, 2008)
- "Serendipity is considered 'not only inevitable but desirable'" (Budd, as cited in (Hemmig, 2008))
- "Students use the library to learn about their discipline, faculty use it for inspiration and ideas" (Hemmig, 2008)
- "Other areas of interest" list is missing a few things: Art students and faculty are also interested in film, the environment—ecology, architecture, biology—genetics, history, religion—spirituality—philosophy.

From our colleagues

Deborah Ultan Boudewyns, Librarian for art and art history

- "Other areas of interest" list is missing a few things: Art students and faculty are also interested in film, the environment—ecology, architecture, biology—genetics, history, religion—spirituality—philosophy.

References

Hemmig, W. S. (2008). The information-seeking behavior of visual artists: A literature review. *Journal of Documentation*, 64(3), 343-62.

Engineering

Engineering Grad Student: I didn't use the library when I was an undergrad, but now my advisor is telling me I need to keep up on anything new that touches on our research area. Not just for me, either: I'm supposed to screen anything that might be related and pass on the really relevant stuff to him. I guess most of this comes from the library somehow, but usually I just go to the websites for the professional organizations and journals where my research group publishes and look for anything new that looks relevant. I've used some books, too, but more and more of those are online. I like that because I don't usually read every page, just the parts that are really important for my work. I'm starting to look for stuff to help with my dissertation, too. I spend a lot of time looking for the things cited in a lot of what I read.

Engineering Faculty Member: I don't much need the library, except for the books they buy and the journals they subscribe to. I check the tables of contents for the main journals in my field online, and keep up on what people are talking about at conferences. When I need to know about something new to me, I'll usually start with a web search. I can find out a lot by using Google, and Google Scholar helps me find academic work in that field. Then I can connect to the full text from the links there. I have colleagues who swear by the article databases the library pays for for keeping up on what's new. But I don't have time to do a lot of searching.

From the literature

- Faculty are extremely unlikely to start research with the library catalog. Much more likely than other groups to start a general purpose web search (Schonfeld & Housewright, 2010).
- Faculty still use TOC and journal browsing for current awareness (Bennett & Buhler, 2010).
- Assistant professors do substantial database searching; Full professors rely more on interpersonal communication through conferences and connections outside the university (Bennett & Buhler, 2010).
- Patterns of use (browsing habits, methods of keeping up, attention to journal name, use of library databases) differ by discipline, even within engineering fields (Bennett & Buhler, 2010).

From our colleagues

Meghan Lafferty, librarian for Chemical Engineering and Material Science

- Chemists also use TOC and journal browsing for current awareness
- In addition to using interpersonal communication and connects to keep current, full professors have graduate students to do the legwork of keeping current for them.
- Chemical Engineering and Materials Science (CEMS) is a good example of how patterns of use can vary by discipline. The field a researcher comes from is definitely a factor; the department includes chemists, electrical engineers, chemical engineers, physicists, and mathematicians.

References

Bennett, D. B., & Buhler, A. G. (2010). Browsing of E-journals by engineering faculty. *Issues in Science and Technology Librarianship*, 61.

Schonfeld, R. C., & Housewright, R. (2010). *Faculty survey 2009: Key strategic insights for libraries, publishers, and societies*.

Law

Law Student: I thought I knew how to search for information, but searching for legal information is different. Finding the legal information I need is difficult since there is a lot of information out there, sometimes I think I spend more time going through my search results than I need to. My instructors tell me that I need to search the print digest in addition to online resources, but I much rather rely on Lexis Nexus, Westlaw, Heinonline, or even Google. It's easiest for me to search online using keywords and descriptions of the specific cases that I am interested in.

Faculty: I rely on others to assist me with my research needs. When I know that I am looking for I will contact a librarian and request research assistance, and sometimes I used student research assistants. In addition to Lexis Nexus and Westlaw for information I use Blawgs and go directly to the publisher to find the most recent information. My colleagues also offer suggestions of articles since we have similar research interests. I will search the library catalog, but this is when I know what I am looking for.

From the literature

Students:

- When first arrive at law school student typically do not have the strong information seeking skills required for the complexities and attention to detail required for legal researching (Jones, 2009)
- Problem is efficient allocation of attention to the right information (especially since law student work in an environment where information is abundant) (Jones, 2009)
 - Value for use of secondary sources (e.g. Westlaw and/or LexisNexis) for cost- and time-savings/efficiencies (Jones, 2009)
- Paradigm shift of reliance on print to online materials (Lihosit, 2009)
 - Potential impact of Computer-Assisted Legal Research (CALR) – searching based on computer code rather than legal principles and policies. (Lihosit, 2009)
- Attorneys develop knowledge based on distributed social networks (e.g. research tools used by attorneys is dependent on what other attorneys in their network are using) rather than individual and controlled textual sources (Lihosit, 2009)
- Natural language searching and searching based on facts of a case instead of legal concepts (Peoples, 2005)
 - Preference for searching for everything using a computer, instead of using the print digest for rules and the computer for facts – users are unaware of some of the shortcomings of electronic research

- Resources used: LexisNexus, Westlaw, Heinonline, Google, research guides, faculty, libraries and the library catalog (primarily for known item searching)

Faculty:

- Reliance on individualized services for research -- The first thing many of our faculty members do when they are seeking information is to contact a reference librarian and request research assistance (Lewis, 2002)
 - Libraries have used the three-tier process described in Lewis' article as a way to support faculty research
- Many faculty members rely on law student research assistants
- Information from colleagues and social networks (Lihosit, 2009)
- Resources used: Blawgs, current awareness sources offered by publishers & scholarship repositories, Google, J-STOR, LexisNexis/Westlaw (and other major or specialty databases depending on their area of interest), library catalog (known item searching), and conferences and workshops also a source of discovery

References

- Jones, J. M. (2009). Not just key numbers and keywords anymore: How user interface design affects legal research. *Law Library Journal*, 101(1), 7-30.
- Lewis, S. H. (2002). A three-tiered approach to faculty services librarianship in the law school environment. *Law Library Journal*, 94(1), 89-100.

Physical/Life Sciences

The library? Only if I have no other choice. Most of the reading I do is from recent journals, and those I can get online. I go to publishers I already know, or sometimes just do a search in Google Scholar. If the article I want is in, for example, an Elsevier journal, I can link right to ScienceDirect. If I'm on campus, I can get to the full text in a couple of clicks. It's annoying, though, that I don't get full text if I'm off campus.

Another annoyance is trying to share articles with my colleagues at other institutions. I do a lot of work with other people around the world, and if we're gathering literature on a topic or starting to write an article, we don't have a good way of sharing articles other than emailing PDFs.

From the literature

- Faculty are extremely unlikely to start research with the library catalog. Much more likely than other groups to start a general purpose web search (Schonfeld & Housewright, 2010).
- Use e-journals very heavily (E-journals: Their use, value and impact, 2009)
- "Just four months after ScienceDirect content in physics was opened up to Google, more than a third of all traffic arrived via this route." (E-journals: Their use, value and impact, 2009)
- The more research-intensive the institution (for a given field), the shorter the sessions on the publishers' platforms. Likely related to a preference for using gateways (defined by Elsevier in this study as any time the user goes directly to an article (PubMed, etc.)). (E-journals: Their use, value and impact, 2009)
- Prefer newer articles (E-journals: Their use, value and impact, 2009).
- Regard physical library buildings as last resort (Marcus et al., 2007).
- Work is increasingly collaborative. Obstacles include different vocabularies, not knowing core journals, and not knowing where to publish (Marcus et al., 2007).

From our colleagues

Meghan Lafferty, Librarian for Chemistry

- I definitely agree about the catalog, but I'm not sure that that's very new in fields very reliant on journals.
- Chemists are still pretty heavy users of databases like SciFinder Scholar, Beilstein, and Gmelin. Structure searching is not an insignificant factor with that.
- Although new literature is very important, older literature is still useful and used in chemistry.
- Another obstacle with increasingly collaborative work is being able to share resources across institutional boundaries due to licensing issues.

References

E-journals: Their use, value and impact (2009). Research Information Network.

Marcus, C., Ball, S., Delserone, L., Hribar, A., & Loftus, W. (2007). *Understanding research behaviors, information resources, and service needs of scientists and graduate students: A study by the University of Minnesota Libraries*. Minneapolis: University of Minnesota Libraries. Retrieved from <http://www.lib.umn.edu/about/scieval/Sci%20Report%20Final.pdf>

Schonfeld, R. C., & Housewright, R. (2010). *Faculty survey 2009: Key strategic insights for libraries, publishers, and societies*.

Social Sciences

I'm primarily concerned with finding journal articles. Ever since I was introduced to JSTOR, I've used it almost exclusively. Lately though, I've found it less helpful because there don't seem to be materials in there that are recent enough. I also use a fair amount of data although the kind of help I usually need - how to work with statistical software or determining the right method - isn't something I go to the library to get. I'm using primary sources more, but it's really time-consuming because it seems like each archive is organized differently. Even though I visit the library myself for my own research, I do feel it's important to send my students to the library.

From the literature

- Ph.D students focus more on journal literature (Fleming-May & Yuro, 2009).
- JSTOR has been incredibly successful at branding itself as the go-to source (Fleming-May & Yuro, 2009).
- Have somewhat contradictory attitudes regarding physical library – feel undergrads they teach should go there, but rarely do so themselves (Fleming-May & Yuro, 2009).
- Archival research is increasingly important, and “challenging due to the idiosyncratic organization of archives and the range and variety of materials housed within them” (University of Minnesota Libraries, 2006).
- Many rely on data sets and/or statistics (University of Minnesota Libraries, 2006).

References

Fleming-May, R., & Yuro, L. (2009). From student to scholar: The academic library and social sciences PhD students' transformation. *Portal-Libraries and the Academy*, 9(2), 199-221.

University of Minnesota Libraries. (2006). *A multi-dimensional framework for academic support*. Minneapolis: University of Minnesota. Retrieved from http://www.lib.umn.edu/about/mellon/UMN_Multi-dimensional_Framework_Final_Report.pdf

APPENDIX 3: PEER INSTITUTION MATERIALS

Email introduction:

The University of Minnesota Libraries is planning for our second phase of enhanced discoverability for our users. One goal is to gather information in three key areas:

Current vision for a local discovery environment

What local resources can or should be exposed to external discovery systems

What external repositories might we integrate into our local discovery environment

As a part of this preparation we want to survey what we believe to be institutions at the front edge of the curve of adoption and innovation. We want to know more about your experiences and activities to help us in our planning – and potential others as well. We intend to make all of our planning documents for Phase 2 available openly as we did for Phase 1 (<https://wiki.lib.umn.edu/WebServices/Discoverability>).

We have developed a set of 7 survey questions that we would ask you to answer. The survey is at this URL:

Hopefully it won't take more than a few minutes to answer these questions.

We want to thank you, in advance, for your time and assistance. If you have any questions, please feel free to contact: _____

Questions:

1. Do you have any reports or research on your institution's work in the area of discovery and delivery to which you could point us?
2. Has your institution developed guidelines or principles for your discovery layer which you could share with us?
3. What tools, systems (and strategies) have you implemented to promote discovery across collections (to move away from "silos")? If you are using specific tools or systems (e.g. Summon, Primo, Blacklight, etc.)? Can you briefly describe the aspects of your collection, environment, and/or user needs that led you to choose this approach?
4. Are there external data sources that you feel are important to integrate into your local discovery environment (e.g. HathiTrust records)? If so, what are those sources and how are you planning to integrate them?
5. Have you identified any user groups that you feel require a specialized discovery environment? Do you segment your discovery environment for any specific user groups?
6. Where (i.e. to what services) is your library exporting data or making it available to be harvested? E.g. WorldCat, REPEC, Google Scholar, Digital Dissertations, OAI sources
7. Are there specific data aggregators to which you feel are critical to send your data, i.e. aggregators that you prioritize when you share your data (e.g. OCLC WorldCat, Archive Grid, EBSCO)?
8. If we have further questions, can we contact you or can you direct us to someone else to talk to?

If you have recommendations about other institutions that are doing innovative work in this area or which you believe it may be helpful for us to contact, please let us know.

As we have done for our previous work on this project, we intend to make all of our planning documents for this phase available openly (<https://wiki.lib.umn.edu/WebServices/Discoverability>).

Bibliography of Recent Reports and Planning Documents from Some Peer Institutions

Indiana University

Digital Project Planning Services homepage, with many links

<http://www.dlib.indiana.edu/services/digitalProjectPlanning/index.shtml>

University of Rochester Libraries

Metadata to Support Next-Generation library resource Discovery: Lessons from the eXtensible Catalog, Phase 1

<https://urresearch.rochester.edu/institutionalPublicationPublicView.action?institutionalItemId=532>
1

Smithsonian Institution

FY 2010 to FY 2015 Smithsonian Institution Information Technology Plan

<http://www.si.edu/ocio/PDFs/SITP.pdf>

University of Michigan

“University of Michigan chooses the Summon Service...”

<http://www.lib.umich.edu/media/news/university-michigan-chooses-summon-service-improve-discovery-library-collections>

Article Discovery Working Group – Final Report (January 2010)

<http://www.lib.umich.edu/files/adwg/final-report.pdf>

Article Discovery Working Group Supplemental report (April 2010)

<http://www.lib.umich.edu/files/adwg/supplemental-report.pdf>

Dartmouth

An Evaluation of Serials Solutions Summon As a Discovery Service for the Dartmouth College Library (November 2009)

http://www.dartmouth.edu/~library/admin/docs/Summon_Report.pdf

Next Generation Library systems at Dartmouth College (September 2008)

<http://www.dartmouth.edu/~library/admin/docs/NextGenLibraryReport.pdf>

OhioLINK

Looking For a Comprehensive Discovery Layer (2008)

<http://dltj.org/article/discovery-layer-itn/>

More About OhioLINK’s Discovery Layer Desires (blog entry, November 24, 2008)

<http://dltj.org/article/discovery-layer-itn-2/>

North Carolina State University

Virtual Shelf Index (Project home page with many links)
<http://www.lib.ncsu.edu/dli/projects/virtualshelfindex/>

Digital Library Initiatives Access & Discovery Projects homepage (with many links)
<http://www.lib.ncsu.edu/dli/projects/accessanddiscovery.html>

California Digital Library

Rethinking How We Provide Bibliographic Services for the University of California
<http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf>

Publications Page - Links to other, more recent reports and information:
<http://www.cdlib.org/news/publications/>

MN-OBE

Social Networking: MN-Obe Reference & Instruction Presentation (2007)
<http://mnobepresentations.pbworks.com/>

MnObe Thinks about the Future of the Catalog (good set of links)
<http://mnobefuturecat.pbworks.com/FrontPage>

University of Nevada – Las Vegas

There were a series of articles on their approach and efforts in *Information Technology and Libraries* December 2009

University of Wisconsin

Resource Discovery (homepage)
<http://cuwlwiki.wetpaint.com/page/Resource+Discovery>

CUWL Resource Discovery Exploratory Task Force
http://uwlib.uwsa.edu/committees/userservices/documents/resourcediscovery_report2009uscc.htm

Blog with links on Forward - their UW System resource discovery experiment
<http://forward.library.wisconsin.edu/moving-forward/>

Resource Discovery at UW Libraries (June 2008)
<http://uwlibdiscovery.blogspot.com/>

University of Chicago

Faceted Browsing Project homepage – Includes a report on the selection of Aquabrowser and other links
<http://www.lib.uchicago.edu/staffweb/depts/ils/projects/faceted-browsing/>

Faceted Browse Study (October 2008)

<http://www.lib.uchicago.edu/staffweb/depts/ils/projects/faceted-browsing/faceted-browse-usability-rept.pdf>

Also, see the "Virtual Wayfinding Study" from 9/2009 appended to their email interview response

Report on Personas, based on work from Johns Hopkins University

2010 survey results <http://www.lib.uchicago.edu/e/surveys/2010/index.html>, particularly comments related to e-resources

<http://www.lib.uchicago.edu/e/surveys/2010/comments/2010eresources.html>

Proposed Data Source	What is this resource?	Scope/Description (does this duplicate other data? Enhance that data? Is this new data?)	Currently being updated/maintained?	Metadata standard/format	Who creates/maintains the metadata?	Digitized object avail? Formats?	Records/data already avail to external systems? (Either targeted "pushing" or available for harvesting "pulling")	Comments	Local contact	What rights information does source provide?	What preservation/retention policy does the source have?
UMD Digital Commons http://d-commons.d.umn.edu:8180/jsui/	A digital repository of University of Minnesota Duluth materials. Initial materials being added to the repository include the American Indian Learning Resources Center Tribal Newspaper Collection and papers and photographs from Glensheen and the Congdon family.	New data	Yes	DSpace/DublinCore	UMD Technical Services staff	Yes, TIFF and PDFs	No, however Dspace supports OAI-PMH v. 2.0 as a data provider	Currently nearing the end of a 1 year pilot project so there are a limited number of records in the system at this time.	Shixing Wen	?	?
UMD and TC campus: Digital Content Library (DCL) http://dcl.umn.edu/	http://dcl.umn.edu/ A combined resource of the Twin Cities College of Liberal Arts (CLA) and the College of Design (CDes). There are almost 200,000 learning objects from different disciplines in image, video, and audio formats. Access is available to all U faculty, students, and staff. The collections are password protected because of copyright considerations. All materials must be used for educational purposes only	New data from a variety of sources purchased, licensed, donations, and copystand photography	Yes	VRA Core 3.0 and Cataloging Cultural Objects (CCO) as a data content standard	Authorized staff in Duluth/TC	Yes - jpeg, TIFF, Quicktime, Flash 8 codec, MP3	No	Go to http://dcl.umn.edu/static_content_items/standards to see complete list of file formats and sizes. Uses VRA	Go to http://dcl.umn.edu/static_content_items/about for complete contact information. Shixing Wen for UMD contact	Must be used for educational purposes only	?
Minnesota Reflections: Minnesota Digital Library	Collection of more than 45,000 images and documents depicting the history of Minnesota, contributed by 100+ historical societies, public libraries, special archives, universities and colleges, including the U of M	Some (most?) U of M contributions may also be in Umedia and possibly other repositories	Yes	Dublin Core	Contributors	JP2, HTML	Discoverable through Google. OAI harvesting available. Records have been pushed to WorldCat.		Jason Roy	?	?
UMD: Song Index http://www.d.umn.edu/lib/ref/music/songdb/index.php	This database provides access to more than 15,000 songs contained in 331 anthologies held by UMD library. Users may search by first line of song, first line of chorus, and composer. Search results list the citation and call number of the anthologies containing the song.	Provides enhanced searching capabilities for titles found in Aleph. http://www.d.umn.edu/lib/ref/music/songdb/index.php	Yes	Metadata entered into a MySQL database. Does not follow any national standard or format	Pam Enrici	Citations only	No	Data stored in a MySQL database	Darlene Morris, Pam Enrici (data input), Dan Filipiak (data export)	?	?
UMD: LibGuides content http://libguides.d.umn.edu/umguides	Approximately 100 web pages containing subject guides and library information for faculty and students.	New data	Yes	XML	Library subject specialists	html, jpeg, pdf	No	U of Iowa has written code to pipe the data to Primo. Code available via EL Commons developers corner	Darlene Morris		
UMD: Finding Aids http://libarchive.d.umn.edu	Finding aids for materials in the UMD Northeast Minnesota Historical Center (NEMHC) and in the future for materials in the UMD Archives	New data	Yes	EAD and MARC	NEMHC staff	Yes, TIFF and jpeg	Will import MARC records to Aleph and Primo. 1800 photographs and map resources are also in Minnesota Reflections	Uses Archon open source software	Darlene Morris/Pat Maus/ Mags David		
UMC: Yearbooks http://yearbooks.umcrookston.edu/	From 1910 to 1999 the school produced yearbooks, which chronicled the year's activities, students and staff. The yearbooks not only offer a history of the faculty, staff, and students who taught, worked, and studied at the institution but they also offer an interesting history of fashion, architecture, sports, arts, and curriculum of the different eras.	digitized version of print	no - publication ceased	DSpace/DublinCore	library staff	yes - pdf	Uses Google Search Appliance to make website data searchable		Owen Williams	?	?

Proposed Data Source	What is this resource?	Scope/Description (does this duplicate other data? Enhance that data? Is this new data?)	Currently being updated/maintained?	Metadata standard/format	Who creates/maintains the metadata?	Digitized object avail? Formats?	Records/data already avail to external systems? (Either targeted "pushing" or available for harvesting "pulling")	Comments	Local contact	What rights information does source provide?	What preservation/retention policy does the source have?
UMC: Northwest Monthly http://nwmonthly.uncrookston.edu	The Northwest Monthly was a publication of the University of Minnesota's Northwest School of Agriculture, located at Crookston Minnesota. It began publication in December 1916. In 1943 it changed its name to Northwest School News and its publication to nine times a year. The last issue was published in 1968	digitized version of print	no - publication ceased	DSpace/DublinCore	library staff	yes - pdf	Uses Google Search Appliance to make website data searchable		Owen Williams	?	?
SFX	SFX is an OpenURL resolver and knowledgebase of electronic resources	The knowledgebase contains our e-journal holdings, some e-books and free e-resources	Yes	No standard; MySQL database	Local Technical Services departments	Facilitates access to digital objects	Yes, available to Google Scholar through a regular export; Exported for enrichment by the MARCIt service and import into MNCAT	Important collection of our holdings; currently synced to Verde e-resources system and will be the base for Verde data in the future		?	?
UDC	University Institutional Repository	this is a unique database	yes	DC	distributed	yes	yes	Google searches objects and records	Beth Kaplan & Philip Herold	?	?
AgEcon	Article database of applied economics with strong international emphasis	this is a unique database	yes	multiple / homegrown	Julie Kelly & Louise Letnes	yes	yes	Google searches records	Julie Kelly	?	?
MNCAT (including records from external vendors where contractual issues may prevent exposure)	Primary catalog of holdings	Includes bibliographic data from vendors and other external entities (e.g., gov docs, OCLC), original records	Yes	MARC	Local Technical Services departments; Enterprise Technology	No	Exported to OCLC WorldCat on a regular basis; Exported to Google and HathiTrust	The only descriptive metadata we have for a large portion of our holdings	?	?	?
MNCAT electronic journal and selected ebook records; free resources	A subset of MNCAT: Records exported from SFX and enriched via the MARCIt service	A subset of MNCAT: Records exported from SFX and enriched via the MARCIt service	Yes	MARC	The process is owned by Enterprise Technology	No; contains links out	Exported to OCLC WorldCat	A cataloger should comment on the metadata quality	?	?	?
MNCAT Duluth holdings that can be shared externally	A subset of MNCAT: Records exported from SFX and enriched via the MARCIt service	Holding records designated within a Duluth sub library	Yes	MARC	Duluth Technical Services	No	Exported to OCLC WorldCat monthly		?	?	?
MNCAT Morris holdings that can be shared externally	A subset of MNCAT: Records exported from SFX and enriched via the MARCIt service	Holding records designated within a Morris sub library	Yes	MARC	Morris Technical Services	No	Exported to OCLC WorldCat monthly		?	?	?
MNCAT Crookston holdings that can be shared externally	A subset of MNCAT: Records exported from SFX and enriched via the MARCIt service	Holding records designated within a Crookston sub library	Yes	MARC	Crookston Technical Services	No	Exported to OCLC WorldCat monthly		?	?	?
MNCAT HSL holdings that can be shared externally	A subset of MNCAT: Records exported from SFX and enriched via the MARCIt service	Holding records designated within an HSL sub library	Yes	MARC	HSL Technical Services	No	Exported to OCLC WorldCat weekly	weekly or monthly?	?	?	?
MNCAT Law Library holdings that can be shared externally	A subset of MNCAT: Records exported from SFX and enriched via the MARCIt service	Bibliographic records designated within a Law Library sub library	Yes	MARC	Law Library Technical Services	No	Exported to OCLC WorldCat weekly	weekly or monthly?	?	?	?
MNCAT Twin Cities holdings that can be shared externally	A subset of MNCAT: Records exported from SFX and enriched via the MARCIt service	Holding records designated within a TC sub library	Yes	MARC	Twin Cities Technical Services	No	Exported to OCLC WorldCat weekly		?	?	?
DLXS	TC archival finding aids	The most comprehensive source of info on U of M archival records holdings, describing 4000-5000 collections; Some collections also described in less detail in MNCAT	Yes	EAD	ASC units	No	Discovery through Google is unreliable. Therefore, many ASC units maintain HTML versions linked to their web sites to enhance discoverability. Data is also harvested into Archives Grid	Comprehensive descriptions	Jason Roy	?	?

Proposed Data Source	What is this resource?	Scope/Description (does this duplicate other data? Enhance that data? Is this new data?)	Currently being updated/maintained?	Metadata standard/format	Who creates/maintains the metadata?	Digitized object avail? Formats?	Records/data already avail to external systems? (Either targeted "pushing" or available for harvesting "pulling")	Comments	Local contact	What rights information does source provide?	What preservation/retention policy does the source have?
UMedia/Images	Repository of curated and user-contributed digitized object and rich media	This is both a replacement and an extension of the previous IMAGES system. Umedia will provide campus support for the discovery of still images, audio and video.	Yes	MODS	Contributors	JP2, Flash for audio and video	Discoverable through Google	Varies. Data for curated material is more complete than user-contributed is likely to be. In Beta	Jason Roy	?	?
Uthink Blogs	Movable Type MU installation containing thousands of current and abandoned blogs created by members of the UMN community. Includes many class blogs.	This data is not collected or indexed elsewhere (except by web search engines, UMN search appliance). Contains blog post data and associated metadata (timestamps, tags, author info).	Yes.	DC?, RSS	Users	Yes - HTML, RSS	Blogs available for crawling by web search engines, UMN search appliance		Shane Nackerud		
PRIMO as its own entity if we indeed decide to ingest content without going through ALEPH first	Piped records from Aleph MNCAT	[checking on exact definition; should be all current holdings for Twin Cities and Duluth]	Yes	PNX (Primo Normalized XML)	Normalization rules are set within Primo	No	No; contains links out to some resources; we could change the robots.txt file to allow access to a structured xml file by crawlers		?	?	?
MLAC - Other libraries' items in storage	?	Holdings for items in closed stacks	Yes	MARC	Twin Cities Technical Services	No	Not sure if these are included in the export to OCLC		?	?	?
Libdata	Subject and research guide creation and management tool. Contains thousands of records for items in our collections and on the open web as well as hundreds of different groupings of those items. Used to manage mappings of subjects to UMN affinity strings.	There is a large overlap with catalog data.	Yes, somewhat. Resources are not regularly updated or weeded.	Custom, RSS	Library staff	Yes? - RSS	Subject guides available for crawling by web search engines, UMN search appliance.		Shane Nackerud		
MHAPO	http://map.lib.umn.edu/mhapo/index.html	Provides access to over 50,000 digital copies of Minnesota air photos found in our collection and available from Landview.	Yes	Custom XML	Map Library	Yes, jpeg	Contains MN county and year; could add uniform description to all records that might help relevant discovery if exposed to external systems	Check on connection with MDL	Ryan Mattke; Kristi Jensen	?	?
Digitized plat maps and atlases	http://map.lib.umn.edu/platmaps	?	?	?	Map Library	Yes, jpeg		Will be going to MDL using their template	?	?	?
Historic Map and Atlas Holdings	http://map.lib.umn.edu/selectedhistoric	Holdings for selected historical items	Yes		Map Library				?	?	?
[Note: this may not be enough data to warrant keeping on the list]											
Selected Minnesota Maps	http://umedia.lib.umn.edu/taxonomy/terms	Available in U Media Beta	?	?	?	?			?	?	?
Charles Babbage Institute Oral Histories	transcripts of interviews with computing pioneers - http://www.cbi.umn.edu/oh/index.phtml	MySQL3 database of oral histories with links to transcripts	Yes, records added when new oral histories created	LCSH for subject headings	CBI	.pdf transcripts (Some audio recordings also available, but these are in Umedia and not linked to this database)	Discoverable in Google. All oral histories are also cataloged in MNCAT (through a separate, non-automated process).	Plans have been discussed to transfer it either to UMedia or Digital Conservancy. Currently the MNCAT records include info on the associated audio recordings but no links to the PDF transcripts.	Stephanie Crowe	?	?
EFW quickref	custom shadow catalog - http://www.lib.umn.edu/apps/qref/search.phtml	provides interactive searching of printed sources within this library's reference collection and of relevant subject related web sites	Yes and No	n/a	EFW staff	no	no, not from this interface	This and the other EFW/Forestry databases are probably going to be discontinued. This one has been migrated to del.icio.us.	Linda Eells	?	?
Map Happy	Google Maps interface for map catalog	custom shadow catalog	yes; irregular	marc & polar > decimal conversion; calc centerpoint	Lisa	sometimes	if you can read the SQL, but not an explicit intent; google is pulling		Lisa Johnston	?	?

Proposed Data Source	What is this resource?	Scope/Description (does this duplicate other data? Enhance that data? Is this new data?)	Currently being updated/maintained?	Metadata standard/format	Who creates/maintains the metadata?	Digitized object avail? Formats?	Records/data already avail to external systems? (Either targeted "pushing" or available for harvesting "pulling")	Comments	Local contact	What rights information does source provide?	What preservation/retention policy does the source have?
Universal Sherlock Holmes	special.lib.umn.edu/rare/ush/ush.html	Web version of reference text on Sherlock Holmes	Yes and no (see comments)	?	?	HTML	Discoverable in Google. Also cataloged (and linked to) in MNCAT.	Appears that updates/supplements are being done as separate PDFs and listed at: http://special.lib.umn.edu/rare/holmes.phtml#Research%20Tools%20and%20Useful%20Links	Tim Johnson	?	?
Urban forestry database	http://forestry.lib.umn.edu/bib/urban.html	Indexes publications relating to the history of urban forestry; urban forest legislation; the benefits of urban forests; selection and planting of trees; maintenance of the urban forest; planning and management; and urban forestry programs.	Covers journal articles, books, conference proceedings, government documents.	DC	Database produced by the U of Minn Forestry Library.	no	no	All records in the Forestry Databases describe items that are owned by the University of Minnesota Libraries.	Linda Eells	?	?
Social sciences in forestry database	http://forestry.lib.umn.edu/bib/SSiF.html	This database indexes publications in 41 subject areas relating the social sciences to forestry. Among them are: legislation, policy and planning, management, economic development, investment and finance.	Covers journal articles, books, conference proceedings, government documents.	DC	Database produced by the U of Minn Forestry Library.	no	no	All records in the Forestry Databases describe items that are owned by the University of Minnesota Libraries.	Linda Eells	?	?
Trail planning, construction, and maintenance database	http://forestry.lib.umn.edu/bib/trls.html	Indexes publications relating to the planning, design, construction, and maintenance of all types of trails. Includes trail use studies.	Covers journal articles, books, conference proceedings, government documents.	DC	Database produced by the U of Minn Forestry Library.	no	no	All records in the Forestry Databases describe items that are owned by the University of Minnesota Libraries.	Linda Eells	?	?
Tropical forestry conservation & development database	http://forestry.lib.umn.edu/bib/trps.html	Subjects indexed include: Tropical forest resources; deforestation; conservation; indigenous peoples; management; policy; trade and industrial development; nontimber forest products; research, education, and training; and history.	Covers journal articles, books, conference proceedings, government documents.	DC	Database produced by the U of Minn Forestry Library.	no	no	All records in the Forestry Databases describe items that are owned by the University of Minnesota Libraries.	Linda Eells	?	?
ETRC (Electronic Text Resource Center)	http://etrc.lib.umn.edu/	Electronic text project developed by the Libraries a number of years ago.	No	TEI	No one	encoded texts	?	This is an old system that has been frozen and is no longer actively supported. Long term availability of this application is undetermined at this time.	Jason Roy	?	?
Pachyderm projects	online exhibits and learning objects	Provides users with the ability to create more contextualized learning objects-incorporating digital media in order to tell a story.	Yes. Support for the Pachyderm system is through central OIT.	Dublin Core	Contributors	image, audio, video	Probably not since the learning object is Flash-based.	Many digital images embedded, but no way to access them individually.	Jason Roy	?	?
Omeka projects	online exhibits and learning objects, http://gallery.lib.umn.edu; http://brickhouse.lib.umn.edu	Online exhibit building tool.	Yes	Dublin Core	Contributors	Image, audio, video, PDF	Discoverable through Google	Many digital images embedded, but no way to access them individually. Metadata varies by project	Jason Roy	?	?

Proposed Data Source	What is this resource?	Scope/Description (does this duplicate other data? Enhance that data? Is this new data?)	Currently being updated/maintained?	Metadata standard/format	Who creates/maintains the metadata?	Digitized object avail? Formats?	Records/data already avail to external systems? (Either targeted "pushing" or available for harvesting "pulling")	Comments	Local contact	What rights information does source provide?	What preservation/retention policy does the source have?
Catalog of Extant Ming Gazetteers	http://eastasian.lib.umn.edu/mingindex_phtml	The catalog covers over a thousand of extant Ming gazetteers compiled and published between 1368 and 1644 found in selected reference works. All entries are in both Chinese characters and pinyin with searchable fields of title, author, publication date, keyword, reign period, Ming and modern administrative divisions. At this point, however, the search can only be performed by using pinyin not Chinese characters.	?	?	?	?	?	Includes title in Chinese characters and pinyin, description, creator, subjects/keywords, publication info, period, series, physical location, source book and set titles in Chinese characters and pinyin, Genre, Language, and holding info	?	?	?
Classic Chinese Films Translation and Subtitling Project	http://eastasian.lib.umn.edu/ccf.phtml	The Project translated and subtitled seven Chinese classical films for the support of teaching and research.	?	VRA	?	Digitized film/video	?	Part of CLA Digital Content Library. Uses VRA	Su Chen	?	?
Chinese Ming History Seen Through Illustrated Maps	http://eastasian.lib.umn.edu/mingmap.phtml	digitized version of print	?	?	?	Maps	?	Includes title in Chinese characters and pinyin, compiler, in Chinese characters publication date, period of reign in Chinese characters, Ming and modern administrative divisions in Chinese characters and pinyin, brief introduction, original copy holders, and reproduction	Su Chen	?	?
Discovering Chinese Culture Through Arts	http://eastasian.lib.umn.edu/chiculture_arts.phtml	The project digitizes and catalogs 1,260 Chinese art objects selected by the renowned ceramic scholars Harry Garner and Margaret Medley from 75 museums around the world found in Chinese Art in Three-Dimensional Colour(donated by Mr. Jonathan R. Gross to the East Asian Library) published by the Asia Society for the Gruber Foundation in 1969.	?	?	?		?		Su Chen	?	?
Japanese Americans in Manzanar 1942-1945	Newspaper http://eastasian.lib.umn.edu/mz.phtml	Manzanar Free Press is a newspaper published by Americans of Japanese descent interned in the Manzanar Relocation Center of Interned Pacific Coast Residents of Japanese Descent between 1942 and 1945.	?		?	Digital images	?	Appears to be a single issue of a newspaper	Su Chen	?	?
Japanese Classic Films Translation and Subtitling Project	http://eastasian.lib.umn.edu/jccf.phtml	The Project translated and subtitled six Japanese classical films for the support of teaching and research.	?	VRA	?	Digitized film/video	?	Part of CLA Digital Content Library. Uses VRA	Su Chen	?	?
Learning Japanese Online	http://eastasian.lib.umn.edu/learning_j_p.phtml	Learning Japanese Online (LJO) is designed to provide self-studying tools for Japanese language learners.	?		?	digitized magazine text	?	Learning object incorporating a digitized magazine	Su Chen	?	?

Proposed Data Source	What is this resource?	Scope/Description (does this duplicate other data? Enhance that data? Is this new data?)	Currently being updated/maintained?	Metadata standard/format	Who creates/maintains the metadata?	Digitized object avail? Formats?	Records/data already avail to external systems? (Either targeted "pushing" or available for harvesting "pulling")	Comments	Local contact	What rights information does source provide?	What preservation/retention policy does the source have?
Photographs of Pre-1949 China in the YMCA Archives	http://eastasian.lib.umn.edu/pp.phtml	717 black and white photos taken by YMCA employees in China between 1896 and 1949 are digitized and cataloged in this project. These photographs, gathered by ordinary young American college graduates, provide rare views of 17 Chinese cities, documenting activities such as sports and physical education, industrial training, public health campaigns, science lectures, and protest movements. Presented as a group, the photographs comprise an extremely valuable resource for scholars and students who are interested in history in general and Chinese history in particular.	?	VRA	?	digitized images	?	Part of CLA Digital Content Library: Already part of Umedia. Uses VRA	Su Chen	?	?
Propaganda through Popular Media: Film, Photograph, Poster, and Slide from China, Japan and Korea	http://eastasian.lib.umn.edu/propaganda.phtml	The project aggregates a group of visual materials including film, photograph, poster, and slides, centered on the theme of propaganda and its promotion through popular media from China, Japan and Korea. These unique visual resources can be accessed online for authorized users.	?	VRA	?	digitized images	?	Part of CLA Digital Content Library. Uses VRA	Su Chen	?	?
Ames Library Exhibits	http://ames.lib.umn.edu/exhibits/ames-ex-0.phtml	These were physical display case exhibits that were also posted concurrently online	No; set up by George Swan and Don Johnson	No	N/A	Yes, jpeg	Ames staff don't really point people here	Available to web search engines	?	?	?
Travelers' Accounts	http://ames.lib.umn.edu/ciguide.phtml	Database of records pointing to specific entries in Ames collection items around this topic	No; set up by George Swan and Don Johnson	Custom set-up	N/A	No	These resources are available in MNCAT, except where Don added unique descriptions in the notes fields	Pulled from MNCAT; not exported	?	?	?
Women in Pre-Independent India	http://ames.lib.umn.edu/wiguide.phtml	Database of records pointing to specific entries in Ames collection items around this topic	No; set up by George Swan and Don Johnson	Custom set-up	N/A	There are some digitized objects	These resources are available in MNCAT, except where Don added unique descriptions in the notes fields	Pulled from MNCAT; not exported	?	?	?
CBI oral history database	transcripts of interviews with computing pioneers - http://www.cbi.umn.edu/oh/index.phtml	?	?	LCSH	?	.pdf transcripts	MNCAT		CBI	?	?

APPENDIX 5: INTERNAL DATA SOURCE PROFILES

Many of these resources are currently available only through unit web pages. With the recent Libraries website redesign, which de-emphasizes those unit-level portals in favor of a more unified approach, these resources are even more obscured. At this stage, the Libraries are attempting to present a single portal for discovery, at least on the surface. However, the infrastructure to support it a single portal isn't really yet in place. Presumably this is only an awkward "growing stage" as the Libraries continue to work towards eliminating silos, but in the short term, more problems may have been created in the process of solving others.

University Digital Conservancy

URL: <http://conservancy.umn.edu>

Description: The UDC is a venue for faculty to deposit copies of their works for long-term preservation and open access and centralized, searchable access to institutional digital resources that would have traditionally gone to the University Archives.

Resources Included: Faculty work, institutional archives, theses and dissertations

Metadata & Access: UDC uses Dublin Core and Google searches full text and metadata of records.

Sent to External Sources: No

Local Contacts: Beth Kaplan & Philip Herold

AgEcon

URL: <http://ageconsearch.umn.edu/>

Description: AgEcon Search is a free, open access repository of full-text scholarly literature in agricultural and applied economics, including:

Resources Included: Working papers, Conference papers, Journal articles

Metadata & Access: Mostly homegrown metadata; Google crawls records but not full-text.

Sent to External Sources: No

Local Contacts: Julie Kelly, Louise Letnes

EFW quickref

URL: <http://www.lib.umn.edu/apps/qref/search.phtml>

Description: Provides interactive searching of printed sources within this library's reference collection and of relevant subject related web sites; acts as a mini shadow catalog.

Resources Included: MNCAT record data

Metadata & Access: No standard metadata format. Access is via website only.

Sent to External Sources: No

Local Contacts: Linda Eells

Map Happy

URL: <http://www.lib.umn.edu/apps/maphappy/>

Description: Google Maps interface to Borchert Map Library collection. Acts as shadow catalog.

Resources Included: MNCAT records plus some additional data

Metadata & Access: MNCAT records plus some additional data

Sent to External Sources: Not sent to Google, but viewed via Google Maps

Local Contacts: Lisa Johnston

Forestry Databases

Urban forestry database

Social sciences in forestry database

Trail planning, construction, and maintenance database

Tropical forestry conservation & development database

URL: <http://forestry.lib.umn.edu/bib/index.html>

Description: All records in the Forestry Databases describe items that are owned by the University of Minnesota Libraries. Goal is to bring to light subject-based journal literature in local collection.

Resources Included: MNCAT records for journals, books, conference proceedings, government documents

Metadata & Access: No standard metadata; access via website only.

Sent to External Sources: No

Local Contacts: Linda Eells

East Asian Film Digitization and Subtitling Projects

Chinese Classic Films Translation and Subtitling Project, Japanese Classic Films Translation and Subtitling Project

URL: <http://eastasian.lib.umn.edu/ccf.phtml>, <http://eastasian.lib.umn.edu/jccf.phtml>

Description: Projects to digitize, translate and subtitle classic Chinese and Japanese films owned by the U of M Libraries, to support research and teaching on campus. Digitized films with subtitles are hosted (and searchable) through the College of Liberal Arts Digital Content Library

(<https://dcl.umn.edu/>) and also accessible through a portal on the East Asian Library web site.

Resources Included: Digitized motion pictures

Metadata & Access: Films in CLA's DCL are cataloged using VRA Core. Films are Source material is cataloged in MNCAT but with no link to the digitized version in the DCL. Access requires U of M authentication.

Sent to External Sources: No

Local Contacts: Su Chen

Learning Japanese Online

URL: http://eastasian.lib.umn.edu/learning_jp.phtml

Description: Online learning object incorporating 5 years of Nihongo Journal, a monthly magazine for Japanese learners. The project incorporates PDF files and audio files. The site indexes the the magazine material according to pedagogical purpose and subject matter, as well as allowing users to browse the magazine by individual issue.

Resources Included: Digitized journal, including both print and audio material

Metadata & Access: No standard metadata; access via website only. Access requires U of M authentication.

Sent to External Sources:No

Local Contacts: Su Chen

Catalog of Extant Ming Gazetteers

URL: <http://eastasian.lib.umn.edu/mingindex.phtml>

Description:Catalog of over 1000 Ming dynasty gazetteers compiled and published between 1368 and 1644 in 10 different reference works.

Resources Included: Bibliographic data in English and Chinese, including data on Ming and modern administrative divisions for each gazetteer

Metadata & Access: No standard metadata; access via website only

Sent to External Sources: No

Local Contacts: Su Chen

Chinese Ming History Seen Through Illustrated Maps

URL: <http://eastasian.lib.umn.edu/mingmap.phtml>

Description: Index of 500+ digitized maps of Ming dynasty-era China

Resources Included: Digitized maps, corresponding descriptive data, including bibliographic data for the source of each map

Metadata & Access: No standard metadata; access via Chinese language website only.

Sent to External Sources: No

Local Contacts: Su Chen

East Asian Media Digitization Projects

Discovering Chinese Culture Through Arts

Photographs of Pre-1949 China in the YMCA Archives

Propaganda Through Popular Media: Film, Photographs, Posters, and Slides from China, Japan, and Korea

URL:

<http://eastasian.lib.umn.edu/chiculturearts.phtml>

<http://eastasian.lib.umn.edu/pp.phtml>

<http://eastasian.lib.umn.edu/propaganda.phtml>

Description: Collections of East Asian-related photographs, lantern slides, posters, and films digitized and made available through the College of Liberal Arts Digital Content Library. While the data is part of the larger DCL, the materials are brought together as a collection through the web page portals on the East Asian Library web site.

Resources Included: Digitized photographs (including photographs of 3-dimensional art objects), lantern slides, posters, and films.

Metadata & Access: VRA, access is via the DCL with some of the content requiring U of M authentication

Sent to External Sources: No

Local Contacts: Su Chen

DLXS (Archival Finding Aids)

URL: <http://discover.lib.umn.edu/findaid/>

Description: Database of 8000+ EAD finding aids describing archival collections. Some are also represented by MNCAT catalog records (though in less detail), but many are only described in DLXS. Projects are currently underway to link from some finding aids to digitized versions of the material they describe.

Resources Included: Archival finding aids

Metadata & Access: EAD

Sent to External Sources: EAD files are harvested by ArchiveGrid; Discovery through Google is possible, but unreliable.

Local Contacts: Kris Kiesling, Jason Roy

UMedia/Images

URL: <http://umedia.lib.umn.edu/>

Description: System for the management and delivery of digital objects and rich media, including both curated and user-contributed material.

Resources Included: Digitized maps, illustrations, photographs, art, posters, videos, and audio files. User-contributed material is not necessarily owned by the University of Minnesota.

Metadata & Access: MODS
Sent to External Sources: Discoverable through Google
Local Contacts: Jason Roy

Minnesota Reflections

URL: <http://www.mndigital.org/reflections/>
Description: Collection of 45,000+ images and documents depicting the history of Minnesota, including about 225 from various University of Minnesota Libraries collections and the University of Minnesota Duluth. Most (all?) should also be in UMedia.
Resources Included: Digitized photographs and architectural drawings
Metadata & Access: Dublin Core
Sent to External Sources: Discoverable through Google, OAI harvesting available. Records have been pushed to WorldCat
Local Contacts: Jason Roy

Electronic Text Resource Center (ETRC)

URL: <http://etrc.lib.umn.edu/>
Description: Online versions of various books and writings (in some cases excerpts) in the humanities, marked up in XML as part of a project that is no longer active.
Resources Included: XML text
Metadata & Access: TEI
Sent to External Sources: Discoverable by Google; Not linked to MNCAT records
Local Contacts: Jason Roy?

Charles Babbage Institute Oral Histories

URL: <http://www.cbi.umn.edu/oh/index.phtml>
Description: Transcripts of oral history interviews with computing pioneers
Resources Included: PDF transcripts
Metadata & Access: LCSH for subject headings; Access is via a MySQL database linked from the CBI web page; Oral histories are cataloged in MNCAT but not linked to the online transcripts
Sent to External Sources: Discoverable via Google
Local Contacts: Stephanie Crowe

UThink Blogs

URL: <http://blog.lib.umn.edu/>
Description: Movable Type MU installation containing thousands of current and abandoned blogs created by members of the UMN community. Includes many class blogs.
Resources Included: Blog posts, images, podcast multimedia files
Metadata & Access: XML (RSS), free-tagging
Sent to External Sources: Web search engines, blog directories
Local Contacts: Shane Nackerud

Libdata

URL: https://www.lib.umn.edu/libdata_admin/login.phtml (internal)
Description: Subject and research guide creation and management tool. Contains thousands of records for items in our collections and on the open web as well as hundreds of different groupings of those items. Used to manage mappings of subjects to UMN affinity strings.

Resources Included: Brief records for electronic and print materials in ULibs collections and on the open web, collections of those records in subject, course guides.

Metadata & Access: Custom scheme, XML (RSS), REST API

Sent to External Sources: MyU Portal

Local Contacts: Shane Nackerud

Charles Babbage Institute Oral History Database

URL: <http://www.cbi.umn.edu/oh/index.phtml>

Description: Transcripts of interviews with pioneers in computing.

Resources Included: PDF documents

Metadata & Access: LCSH

Sent to External Sources: MNCAT

Local Contacts: Stephanie Crowe

Universal Sherlock Holmes

URL: <http://special.lib.umn.edu/rare/ush/ush.html>

Description: Online version of reference text on Sherlock Holmes, updated with supplements

Resources Included: HTML and PDF text

Metadata & Access: Main text is cataloged in MNCAT. Supplemental PDFs are not linked to main text or cataloged in MNCAT. Can only be found via Special Collections web page.

Sent to External Sources: Discoverable via Google

Local Contacts: Tim Johnson

Digital Commons (UM Duluth)

URL: <http://d-commons.d.umn.edu:8180/jspui/>

Description: A digital repository of University of Minnesota Duluth materials.

Resources Included: Initial content will include tribal newspapers from the American Indian Learning Resources Center and papers and photographs from Glensheen and the Congdon family.

Metadata & Access: Dublin Core and DSpace

Sent to External Sources: No

Local Contacts: Shixing Wen; Darlene Morris

Digital Content Library

URL: <http://dcl.umn.edu/>

Description: A combined resource of the Twin Cities College of Liberal Arts (CLA) and the College of Design (CDes). Access is available to all U faculty, students, and staff. The collections are password protected due to copyright considerations. All materials must be used for educational purposes only.

Resources Included: There are almost 200,000 learning objects from different disciplines in image, video, and audio formats. U of M Duluth library contributes images on an on-going basis.

Metadata & Access: VRA Core 3.0 and Cataloging Cultural Objects (CCO) Searchable via the website

Sent to External Sources: No

Local Contacts: Go to http://dcl.umn.edu/static_content_items/about for complete contact information. Shixing Wen is Duluth campus contact

Song Index (UM Duluth)

URL: <http://www.d.umn.edu/lib/ref/music/songdb/index.php>

Description: A database that allows users to search for songs by first line of song, first line of chorus, and composer.

Resources Included: 15,000 songs contained in 331 anthologies held by UMD library

Metadata & Access: MySQL database. Searchable via webpage

Sent to External Sources: No

Local Contacts: Pam Enrici, Dan Filipiak

LibGuides (UM Duluth)

URL: <http://libguides.d.umn.edu/umguides>

Description: Web pages created by library staff

Resources Included: Subject guides and library instruction.

Metadata & Access: Hosted database outputs data in XML format

Sent to External Sources: No, in the future records could be imported into Primo via pipes using code created by University of Iowa library staff

Local Contacts: Darlene Morris

Finding Aids (UM Duluth)

URL: <http://libarchive.d.umn.edu>

Description: Finding aids for materials from the University of Minnesota Duluth

Resources Included: The collections of the Northeast Minnesota Historical Center (NEMHC) are currently being added. The collections of the UMD Archives will be added in the future.

Metadata & Access: EAD and MARC records are being created. Access is via Archon open source software

Sent to External Sources: No, in the future MARC records will be imported into Aleph/Primo

Local Contacts: Darlene Morris, Pat Maus, Mags David

Yearbooks (UM Crookston)

URL: <http://yearbooks.umcrookston.edu/>

Description: Digitized versions of the Crookston campus yearbooks

Resources Included: Yearbooks dating from 1910 to 2000

Metadata & Access: Dublin Core – Dspace access.

Sent to External Sources: Google search appliance is used to make data searchable

Local Contacts: Owen Williams

Northwest Monthly (UM Crookston)

URL: <http://nwmonthly.umcrookston.edu/>

Description: Digitized version of the Northwest Monthly, a publication of the U of MN Northwest School of Agriculture, located at Crookston.

Resources Included: Publications covering the period 1916 to 1968

Metadata & Access: Dublin Core – Dspace access.

Sent to External Sources: Google search appliance is used to make data searchable

Local Contacts: Owen Williams

SFX Electronic Journals

URL: http://tc.liblink.umn.edu/sfx_local/a-z/default

Description: The SFX system, branded as “Find It” at UMN, contains electronic journals holdings for all campuses. Some electronic books are also activated in SFX. The system enables discovery of materials via the E-Journals list and delivery of materials via OpenURL link resolving.

Resources Included: Primarily electronic journals; some e-books

Metadata & Access: The SFX KnowledgeBase and administrative data is stored in a MySQL database based on a custom structure developed by the Ex Libris Group.

Sent to External Sources: Yes, an SFX job makes data available to be harvested by external sources, e.g., Google Scholar.

Local Contacts: Janet Arth and Cecilia Genereux

Primo Discovery

URL:

http://prime2.oit.umn.edu/primo_library/libweb/action/search.do?dscnt=0&vid=TWINCITIES&dstmp=1282780441630&fromLogin=true

http://prime2.oit.umn.edu/primo_library/libweb/action/search.do?dscnt=0&fromLogin=true&vid=DULUTH&dstmp=1282780569844&fromLogin=true

Description: Primo is a system that provides a discovery layer for one or more sources of data. The University of Minnesota - Twin Cities and Duluth campuses offer Primo as the default interface to the library catalog. These search interfaces are branded as MNCAT and QuickSearch, respectively. Primo is capable of ingesting data from multiple sources and metadata structures. At this time, the University of Minnesota is only piping Aleph catalog material into Primo.

Resources Included: All publically-accessible data included in the Aleph catalog and owned/made available by the Twin Cities or Duluth campuses.

Metadata & Access: Primo will ingest various metadata formats, including MARC, Dublin Core, EAD, etc. All metadata is converted into the PNX format (Primo Normalized Record).

Sent to External Sources: Primo data is not currently exposed to any external sources. However, we could modify our version 3 configuration to expose our holdings to crawlers.

Local Contacts: Naun Chew, Jeff Peterson

Specialized Resources from the Borchert Map Library

URL:

<http://map.lib.umn.edu/mhapo/index.html>

<http://map.lib.umn.edu/platmaps>

<http://map.lib.umn.edu/selectedhistoricmn>

<http://umedia.lib.umn.edu/taxonomy/term/713>

Description: Staff in the Borchert Map Library have actively worked to make unique resources in those collections available to users in digitized formats. Areas of emphasis have included: air photos, plat maps and atlases, and other selected Minnesota maps, including historic resources.

Resources Included: The MHAPO project provides access to over 50,000 digital copies of Minnesota air photos found in the local collection and available from Landview

(<http://www.dnr.state.mn.us/maps/landview.html> -- confirm link with Kristi/Ryan). Digitized plat maps and atlases are available through the Map Library website and will be made available soon in the Minnesota Digital Library. Selected digitized map holdings are available through the U Media Archive.

Metadata & Access: The MHAPO data is in a custom XML format. Selected plat maps and atlases will be described using the MDL template. Resources already available in the U Media Archive are [need to get this information from Stephen or Jason]

Sent to External Sources: Selected digitized objects are available in the Minnesota Digital Library and through U Media Archive. The MHAPO project may represent an opportunity to expand access.

Local Contacts: Kristi Jensen, Ryan Mattke

Specialized Resources from the Ames Library

URL:

<http://ames.lib.umn.edu/exhibits/amesex-0.phtml>

<http://ames.lib.umn.edu/ciguide.phtml>

<http://ames.lib.umn.edu/wiguide.phtml>

Description: By varying degrees, several historic projects have collected metadata, digitized objects, and enhanced descriptions for specialized resources in the Ames Library. These projects included: Ames Library Exhibits, Travelers' Accounts, and Women in Pre-Independent India.

Resources Included: The Ames Library Exhibit collection presents objects in .jpeg format that were part of physical display case exhibits that were also posted concurrently online. This data is discoverable by web search engines as a part of the Libraries' website. The Travelers' Accounts and Women in Pre-Independent India databases each point to specific physical resources in the Ames collection around that topic. There are some digital objects included in the Women in Pre-Independent India system.

Metadata & Access: The Library Exhibits data is only available in HTML ([verify]) and does not use a formal metadata structure. The other two resources exist in databases and also use a common custom structure. None of these online collections are being maintained.

Sent to External Sources: Surrogate records for the Travelers' Accounts and Women in Pre-Independent India collections are already available in the Aleph catalog. However, none of the added descriptions and notes for these objects are available outside these custom databases.

Local Contacts: David Faust

APPENDIX 6: EXTERNAL DATA AGGREGATOR PROFILES

AgEcon Search

<http://ageconsearch.umn.edu/>

Description: AgEcon Search is a free, open access repository of full-text scholarly literature in agricultural and applied economics.

Resources included: Over 40,000 items in nine languages, primarily working papers, conference papers, and journal articles, all in PDF format.

Metadata and access: Contributors provide Dublin Core metadata records. Find and browse searching is done against this metadata. There is no indexing on the site of full-text content.

University of Minnesota participation: The University hosts the AgEcon Search database (currently in Dbase, eventually in Fedora) and actively solicits contributions to it from researchers, conferences, and scholarly organizations.

Local contacts: Julie Kelly, Louise Letnes

Archive Grid

<http://www.archivegrid.org>

Description: ArchiveGrid is an important destination for searching for historical documents, personal papers and family histories held in archives throughout the world. ArchiveGrid is an online service that provides access to detailed archival collection descriptions. It includes nearly a million descriptions of archival collections held by thousands of libraries, museums, historical societies and archives worldwide.

Resources included: Finding aid information derived from MARC records in WorldCat (and formerly from RLIN) and harvested from contributing institutions which request crawling of online finding aids.

Metadata and access: ArchiveGrid parses data from the source metadata to compose an ArchiveGrid metadata record, including links to the holding institution. Where the source indicates access restrictions, these appear also in ArchiveGrid. Contribution is free, but access to the full set of ArchiveGrid metadata is through paid subscription. Many ArchiveGrid records are also included in WorldCat.org and can be searched without charge.

University of Minnesota participation: We both contribute to and subscribe to ArchiveGrid.

Local contacts: Charles Spetland for subscription; Kris Kiesling for contributed data

ArXiv.org

<http://arxiv.org>

Description: ArXiv.org is a repository for open-access hosting of pre-print, e-print, and research papers in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics.

Resources included: Documents and related files in the following formats: (La)TeX, AMS(La)TeX, PDFLaTeX, DOCX (Word 2007), PDF, PostScript, HTML with JPEG/PNG/GIF images.

Metadata and access: Custom metadata, API and OAI-PMH access

University of Minnesota participation: Andrew Odlyzko, School of Mathematics, is listed on the Advisory Committee for computer science in ArXiv; documents by University of Minnesota scholars appear in the database.

Local contacts: Unknown

Digital Dissertations

<http://proquest.umi.com/pgdweb>

Description: Digital Dissertations is a repository of analog and digitized versions of dissertations with their corresponding metadata. Originally owned by University Microfilms, Inc (UMI), it is currently run by ProQuest. In addition to dissertations, a few authors opt to add their master's theses to Digital Dissertations.

Resources included: The repository includes dissertations from 700 universities world-wide dating back to 1938. Most of the older dissertations are in microfilm format. Increasingly universities are contributing in digital formats. When requests for older dissertations are processed, ProQuest generates a digitized version and adds it to the repository.

Metadata and access: Dissertations are searchable by various data elements, including author, title, subject, and abstract. As a participant in the CIC consortial agreement with ProQuest, the University of Minnesota has free access to dissertations produced at CIC institutions.

University of Minnesota participation: The University contributes digital and hard copy dissertations to Digital Dissertations along with author-supplied metadata. University Archives considers the Digital Dissertation copy of a university thesis to be archival. Many of the university's digital dissertations and theses are also available from the University Digital Conservancy.

Local contacts: Charles Spetland, Jim Stemper

Google Books

<http://books.google.com/>

Description: Google Books makes searchable the full text of books through agreements with publishers (who may submit PDFs or print copies for scanning) and scanning projects with libraries. Depending on copyright status and permissions that have been granted, users are able to read very brief excerpts, slightly longer excerpts, or entire books. Books in the public domain may be downloaded as PDFs; some titles are also available in EPUB format.

Resources included: More than 10 million books and magazines in many languages and from all time periods.

Metadata and access: All books scanned through the Library Project are submitted to Google with MARC metadata supplied by the owning libraries. Primary access is through <http://books.google.com>, where individual users can search, browse, track and organize books of interest, read and write reviews, purchase, or find a book in a library. APIs are available at <http://code.google.com/apis/books/>

University of Minnesota participation: As a member of the CIC, the University of Minnesota is participating with scanning currently underway. About 1 million U of MN-owned volumes are expected to be scanned. In order to participate, the University Libraries made its full MARC catalog available to Google, which then provided selective lists of titles desired for the scanning project. Titles scanned by Google will also be available through HathiTrust.

Local contacts: Jerrie Bayer (logistics); Sue Hallgren (general questions); Charles Spetland (collection related questions).

Google Documents (part of Google Books)

<http://books.google.com/>

Description: The Google Government Documents project is a CIC initiative that is part of the larger Google Books project. There is no separate interface or access for government documents. Detailed information on the CIC project is available at

<http://www.cic.net/home/projects/Library/BookSearch/GovDocs.aspx>.

Resources included: Over 100,000 federal documents are currently included. Minnesota's contribution is complete; other CIC libraries are still in process.

Metadata and access: Like other resources included in Google Books, documents were submitted to Google with MARC metadata supplied by the owning libraries. Primary access is through <http://books.google.com>, where individual users can search, browse, track and organize books of interest, read and write reviews, purchase or find a book in a library. APIs are available at <http://code.google.com/apis/books/>

University of Minnesota participation: The University of Minnesota submitted 85,000 federal documents that were duplicates in our collection. The project is complete. The documents are also included in HathiTrust, where they are in the public domain category with full access available.

Local contacts: Kirsten Clark

ICPSR (Inter-university Consortium for Political and Social Research)

<http://www.icpsr.umich.edu/icpsrweb/ICPSR/>

Description: ICPSR is an international consortium of about 700 academic institutions and research organizations. It provides leadership and training in data access, curation, and methods of analysis for the social science research community.

Resources included: ICPSR maintains a data archive of more than 500,000 files of research in the social sciences. It hosts 16 specialized collections of data in education, aging, criminal justice, substance abuse, terrorism, and other fields. Resources include both documentation on research studies (reports, code books) and datasets from those studies.

Metadata and access: ICPSR metadata can be searched for known items and browsed by topic, geography, investigator, and series. Access to documentation is freely available. Access to datasets is generally restricted to ICPSR member institutions.

University of Minnesota participation: The University is a contributing member of ICPSR with access to ICPSR datasets.

Local contacts: Amy West

OAIster

<http://oaister.worldcat.org>

Description: "OAIster is a union catalog of millions of records representing open archive resources that was built by harvesting from open archive collections worldwide using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Today, OAIster boasts more than 23 million records representing digital resources from more than 1,100 contributors." –Description from OCLC, <http://www.oclc.org/oaister> .

Resources included: OAIster is an aggregation of harvestable metadata, not of resources. The resources represented are in a wide variety of publication and media types.

Metadata and access: The OAIster metadata records are harvested as Dublin Core metadata, the mandatory base-level format for OAI-PMH. Some complex objects are represented at the object level, others at the page level.

University of Minnesota participation: When the university exposes Dublin Core metadata records for its digital collections via OAI-PMH, the metadata is harvested by OAIster and incorporated into the OAIster search database. For example, objects in the Images/UMedia database can be found via OAIster.

Local contacts: John Butler

Research Papers in Economics (RePEc)

<http://repec.org>

Description: An open, volunteer, decentralized collection of scholarship in the field of economics. Institutions can establish local RePEc repositories of metadata and digital objects searchable through the RePEc search database, Econpapers, <http://econpapers.repec.org> .

Resources included: 910,000 items including working papers, journal articles (often metadata only, access via a local resolution service), software components, and author and institution contact listings. AgEcon Search contributes its metadata to the RePEc database.

Metadata and access: Econpapers searches all the metadata for scholarly objects. The format of the records appears to be local. Authors may register their identities with the RePEc author registration service and declare which articles should be linked to their registered identities.

University of Minnesota participation: AgEcon Search contributes its metadata to the RePEc database.

Local contacts: Julie Kelly; Louise Letnes

Social Explorer

<http://www.socialexplorer.com>

Description: Social Explorer is a web-based aggregation of U.S. demographic data from 1790 to present. The primary product is a web application that creates demographic maps and reports. Users can also download data to use on their own. A free scaled-down version is available (access to Census 2000 data/maps only), as well as a premium subscription edition.

Resources included: Includes historical census data for the United States from 1790 to the present at census tract, county, state, and national levels. Based on this data, Social Explorer makes available interactive thematic maps of historical census data and customized reports for any selected area. Data is continuously added.

Metadata and access: Access through Social Explorer website. Reports can be downloaded as Excel spreadsheets or in CSV format. Maps generated by Social Explorer can be saved in a variety of image formats or exported as PowerPoint slides.

University of Minnesota participation: The National Historical Geographic Information System (NHGIS) is a project of the Minnesota Population Center at the University of Minnesota. The NHGIS project made historical GIS possible by digitizing maps of United States back to the first census in 1790. Social Explorer worked closely with the NHGIS for many years to digitize some of the historical data for various cities around the country.

Local contacts: Wendy Thomas

Wikimedia Commons

<http://commons.wikimedia.org>

Description: Wikimedia Commons is a repository of media files available for use without license or attribution. Consisting primarily of images, the Commons also includes audio and video files. Images used in Wikipedia articles are hosted in the Commons.

Resources included: Images, audio, video

Metadata and access: Custom metadata, open wiki-style access

University of Minnesota participation: Unknown

Local contacts: Unknown

Wikipedia

<http://en.wikipedia.org>

Description: Wikipedia is a multilingual, web-based, free-content encyclopedia project based on an openly-editable model. Wikipedia's articles provide links to guide the user to related pages with additional information. Wikipedia is written collaboratively by largely anonymous Internet volunteers who write without pay. Anyone with Internet access can write and make changes to Wikipedia articles (except in certain cases where editing is restricted to prevent disruption and/or vandalism). Users can contribute anonymously, under a pseudonym, or with their real identity, if they choose. Since its creation in 2001, Wikipedia has grown rapidly into one of the largest reference websites, attracting nearly 78 million visitors monthly as of January 2010.

Resources included: More than 16,000,000 articles in more than 270 languages. More than 3.4 million articles in English.

Metadata and access: Wikipedia uses a very simple markup based on UseModWiki. Wikipedia's full text is searchable. Wikipedia also has browsable indices, including an alphabetical index and a category index (based on category tags for each article). Broad categories can also be accessed via high-level LC and Dewey classification lists.

University of Minnesota participation: It appears that there has been some ad-hoc linking by ASC staff. Some peer institutions have had success generating traffic to their sites, in particular archives and special collections, by adding links to relevant Wikipedia articles.

Local contacts: Unknown

WorldCat

<http://www.worldcat.org>

Description: WorldCat is the world's largest network of library content and services. WorldCat serves as a "master" catalog of library materials. WorldCat.org is a portal to the WorldCat catalog and a supporting program of data syndication.

Resources included: MARC records for anything cataloged by a participant, with an emphasis on traditional library materials in all formats, and including both physical formats and electronic resources. Includes 193 million bibliographic records for items in formats including books, videos, serial publications, articles, recorded books and music, electronic books, sheet music, genealogical references, cultural artifacts, digital objects, and Web sites. Includes items in more than 470 languages and multiple scripts.

Metadata and access: Records are MARC format. Contributions to WorldCat are governed by OCLC standards and guidelines for both content and coding. Access available through Worldcat.org as well as "basic" and "search" APIs and other web services. See

<http://www.oclc.org/services/web/default.htm> and http://worldcat.org/devnet/wiki/Main_Page. -

University of Minnesota participation: The University of Minnesota is an OCLC member and submits its bibliographic holdings to WorldCat via batch on a regular basis.

Local contacts: Christine DeZelar-Tiedman, Sue Koelmel (billing)

External Aggregations	Do we contribute metadata? By what method?	Do we contribute digital objects? By what method?	Can we update metadata or replace digital objects?	Who creates the metadata and/or digital objects?	What metadata content standards are used?	What metadata formats or schemas are used?	What file formats are used for digital objects?	How do we search and retrieve data from the aggregator?	What preservation/retention purpose does the aggregator serve?	What rights information does the aggregator provide?	Local contact or person with expertise?
HathiTrust	Yes.	Yes. Google scans are sent to HathiTrust	Yes	UL, other libraries contributing records, Google Scan project	AACR2, LC SHM, others	MARC, DC	PDF	Web interface. MNCAT API to check for copy in Hathi Trust	May be part of last copy in CIC decisions	Maintain copyright information	John Butler (admin); Chew Chiat Naun (data)
Google Scholar	No	No	NA	Publishers and indexes				Web interface	None	Rights mediated by link resolver (SFX)	
Google Books	Yes?	Yes. Scheduled shipments for scanning	??	UL			PDF	Web interface. Links in MNCAT.	??	Rights determined by Google algorithm	
WorldCat	Yes. Scheduled batchloads of records	No	Yes	UL	AACR2, LC SHM, others	MARC, DC	NA	Web interface. [OCLC Connexion client, Z39.50 for staff]	Global master bibliographic database	OCLC	Chew Chiat Naun
Internet Archive	??	??	Yes	Contributor	??	??	Various text, image, and data formats			Rights specified by contributor	
Minnesota Digital Library	Yes	Yes	Yes	Minn cultural heritage institutions	Local	DC, MODS forthcoming	PDF, ??	Web interface	Web interface. Included in Video and Media databases		
Digital Dissertations	Yes. Metadata files sent to ProQuest	Yes. Files sent to ProQuest	No	UM Grad School prior to reorganization	Digital Dissertations topical headings; spelling out of symbols in titles	Dublin Core, MARC	PDF and others	Web interface. Included in our list of databases	Regarded as archival for U of Minn dissertations	ProQuest embargoes content it considers to be not cleared for copyright	Charles Spetland; Jim Stemper

External Aggregations	Do we contribute metadata? By what method?	Do we contribute digital objects? By what method?	Can we update metadata or replace digital objects?	Who creates the metadata and/or digital objects?	What metadata content standards are used?	What metadata formats or schemas are used?	What file formats are used for digital objects?	How do we search and retrieve data from the aggregator?	What preservation/retention purpose does the aggregator serve?	What rights information does the aggregator provide?	Local contact or person with expertise?
ArtSTOR	Yes	Yes. WW1 and WW2 posters and postcards.	??	Digital Collections Unit				Web interface. Included in our list of databases			
CAMIO (Catalog of Art Museum Images Online)	No	No (but Walker and MIA do)	??	Contributing institution	??	DC?	image files	Web interface. Included in our list of databases		Rights cleared by CAMIO	No current license
ArchiveGrid	Yes. OAI-PMH with notice to ArchiveGrid	No	Yes?	Contributing institution	DACS, AACR2	EAD, HTML, MARC, text	Generally points only to finding aid	Web interface. Included in our list of databases	None	Rights info contained in finding aid or MARC record	Kris Kiesling; Charles Spetland
Wikipedia	Links added by some ASC staff	No articles known create by staff	Yes	anyone	Wikipedia standard	HTML	NA	Web interface; Google	None	None	NA
Social Explorer (Mapping census data)											
AgEcon Search	Yes. Locally hosted. Authors contribute metadata	Yes. Locally hosted. Authors contribute objects	Yes	Authors, AgEcon staff	Local	DC	PDF and others?	Web interface			Julia Kelly
RePEc (Economics)	Yes. AgEcon metadata harvested via OAI-PMH	No	??	AgEcon Search primarily authors	Local	DC	NA				
ArXive (Physics)	No	??	Yes				.pdf				
MERLOT (Multimedia Educational Resource for Learning and Online Teaching)	Yes?	Yes. Individual membership s	Yes	Contributors, including U of M faculty	??	??	??	MERLOT search interface. Not included in our list of databases	??	Rights specified.	

External Aggregations	Do we contribute metadata? By what method?	Do we contribute digital objects? By what method?	Can we update metadata or replace digital objects?	Who creates the metadata and/or digital objects?	What metadata content standards are used?	What metadata formats or schemas are used?	What file formats are used for digital objects?	How do we search and retrieve data from the aggregator?	What preservation/retention purpose does the aggregator serve?	What rights information does the aggregator provide?	Local contact or person with expertise?
Networked Digital Library of Theses and Dissertations (NDLDT)	No. NDLDT uses OAI-PMH	No. NDLDT collects only metadata	??	Authors, library staff	Local and varied	DC	PDF	Web interface. Included in our list of databases	None	Specified by contributing institution or author	
OAlster	Yes. OAI-PMH	No	Yes	Archives and Special Collections; Technical Services; Digital Collections	Local and varied	DC	Various image and text formats	WorldCat search; OAlster search	Includes preservation metadata	Rights information is provided by institutions holding the objects	John Butler
Flickr Commons	No	No	Yes	Content providers, users	EXIF		JPEG, TIFF, PNG				
Wikimedia Commons	No	No	Yes	anyone			JPEG, TIFF, PNG, SVG, MIDI, OGG				
ICPSR (Soc Sci datasets) [Other dataset repositories]	Yes?	Yes? Sometimes required by terms of grant funded research?	??	Contributor	Datasets are accompanied by explanatory code books.	??	Various	ICPSR web interface; locally held codebooks are in MNCAT	Preservation site for datasets; policy posted on website	Access to non-governmental research is limited to ICPSR members	Amy West
[Other disciplinary repositories, cf. http://oad.simmons.edu/oadwiki/Disciplinary_repositories]											