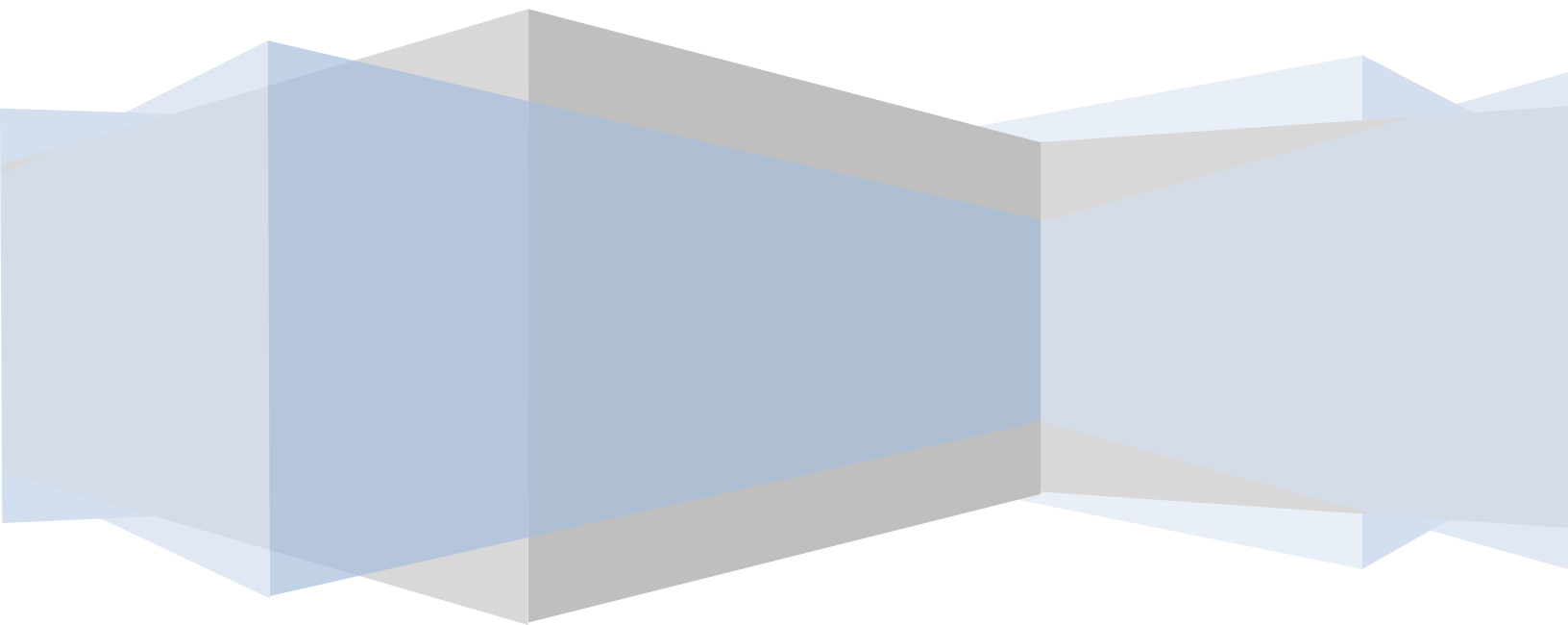


Minnesota Historical Society

# Search Engine Replacement RFI

April 13, 2009

Responses Due May 15, 2009



## 1 INTRODUCTION

The intent of the Search Engine Replacement Request For Information (RFI) is to solicit industry input to validate our roadmap and to gather information, including cost estimates, that can help us focus on a subset of possible solutions for content ingestion and search, as a precursor to our Request for Proposal (RFP) process.

## 2 BACKGROUND

### 2.1 THE MINNESOTA HISTORICAL SOCIETY

The Minnesota Historical Society (MHS) ([www.mnhs.org](http://www.mnhs.org)) is an independent, nonprofit corporation created by the Legislative Assembly of the Territory of Minnesota in 1849. The mission of the Society is to foster among people an awareness of Minnesota History so that they may draw strength and perspective from the past and find purpose for the future. To achieve this objective, the Society provides opportunities for all people of all ages to learn about the history of Minnesota, collects and cares for materials that document human life in Minnesota, makes them known and accessible to people in Minnesota and beyond, and encourages and executes research in Minnesota history.

MHS operates a non-circulation library and approximately 26 historic sites with trails, historic and modern buildings, and exhibits across the state. The Society employs approximately 275 full-time employees and 400 part-time employees, with over 1,838 volunteers contributing over 44,100 hours of labor annually.

Although organized as a non-profit, the MHS is responsible for certain state functions, including the State Historic Preservation Office and State Archives, and receives approximately 55 percent of its annual revenue of approximately \$45 million from state appropriations. The remaining 45 percent comes from private donations, federal grants, and earned revenue.

The MHS is governed by its officers and an Executive Council elected by the membership. The Executive Council appoints the Director, who has the responsibility of directing the Society in accordance with its policies.

In fiscal year 2008 which ended on June 30, 2008, the MHS's web site received more than 8.5 million visits.

### 2.2 THE GREAT RIVERS NETWORK

With the support of a grant from the Bush Foundation in September 2006, the MHS proposes to become the digital content and service center for the cultural heritage of the Upper Midwest. With an expanded technological capacity, the MHS will:

- Build a regional network to be known as the Great Rivers Network (GRN) ([www.greatriversnetwork.org](http://www.greatriversnetwork.org)) that will integrate access to the collections of diverse institutions and facilitate resource discovery;
- Enhance its capabilities through more robust, standards based, integrated systems that utilize best of breed solutions;
- Increase the quality and quantity of available digital content, in partnership with a variety of institutions;
- Sustain the infrastructure through the use of solutions based on open standards; and
- Ensure the necessary financial base by creating a demand for valued products that will result in ongoing funding and by offering particular goods and services that will provide direct financial underwriting.

---

## 2.2.1 GRN PARTNERS

The GRN partners vary in size and capacity. There are State Archives, county historical societies, and academic institutions. The county organizations are smaller, with limited technology budgets ranging from \$50 to \$13,000, with the average being \$2,040. The other partners are more apt to have grant funded projects. A profile of each partner is available in Appendix A.

---

## 2.2.2 AUDIENCES

The GRN provides the following audiences with access to the partner organizations' collections:

- For K-12 educators, it provides primary sources, research materials, and instructional resources
- For students, it provides primary sources and research materials
- For family historians, it provides primary sources and research materials for a specific person
- For the public, it provides personal enrichment and general information about events, hours, and places
- For commercial users, it provides a tool for locating material for use in publications and other products
- For internal staff, it provides an administrative tool for the cataloging and management of information and digital resources

## 2.2.3 TYPES OF CONTENT

The content repositories searched by GRN are in disparate locations, in heterogeneous format and storage system, and large in scale. In general the content can be broken into 3 types:

1. **Nominal Data** – which are about a person or event in a person’s life. Typically, they are searched using names and dates and structured by name, date, geo-location, and event type fields. Examples include indexes of birth, death, and marriage, as well as census and cemetery records. Some collections have digital representations of the actual record.
2. **Standards-Based Data** – are organized in, or can be exported to, standard data formats such as OAI, EAD, EAC, MARC, XML schemas, and Dublin Core. PastPerfect (<http://www.museumsoftware.com/>) collections are included in this category since the program does provide some standard fields for each type of data it manages. PastPerfect is a software application used by many smaller organizations to manage their museum and archival collections.
3. **Other Content** – includes everything else such as web pages, pdf documents, photo collections, map collections, audio, video, documents, event calendars, etc. Some of this information may actually be stored in structured databases or spreadsheets, but it is not in a standard-based format.

The table in Appendix A contains a list of currently identified content by partner.

## 2.3 CURRENT TECHNOLOGY ENVIRONMENT

### 2.3.1 AUTONOMY IDOL

In September 2006, MHS signed a four-year agreement to license the IDOL search engine from Autonomy. IDOL was selected because it supports indexing of a variety of digital content in different forms and in different repositories (e.g., static web pages, databases and federated searches). IDOL can perform fielded searches, customizable sorting of results, and discreet collections of information. Further, IDOL was in the upper right of Gartner’s magic quadrant. Our faith in IDOL came from the fact that it was being used by federal government agencies including the U.S. Department of Defense, British National Archives, and Library and Archives Canada.

In early 2007, MHS contracted with Perficient, Inc., to help us configure IDOL and create a Java based interface to the IDOL server. Since then, more than 5.1 million records have been indexed into IDOL from the mnhs.org web site, six MHS MS-SQL databases, and partner content from an OAI (Open Archives Initiative) harvest feed from ContentDM. A federated search, via the Z39.50 protocol, is used to search the MHS online public access catalog (OPAC). The OPAC is in MARC format in an XLibris database. It is hosted by MNPals at the Minnesota State University at Mankato.

We launched the IDOL search engine in June 2008. You can see it in action in the search box located on the header of many web pages at [www.mnhs.org](http://www.mnhs.org). We refer to this search as the HistoryFinder search. We also launched the PeopleFinder search, [www.mnhs.org/peoplefinder](http://www.mnhs.org/peoplefinder). This search is focused on finding people in several vital statistic databases as well as our photo and art database.

In the five months since October 2008, Google analytics reveal 221,598 pages were viewed a total of 524,438 times. There were more than 94,000 visits from 68,000 visitors, with a bounce rate of 44%. Feedback is positive and negative, with the negative focused on the difficulty of understanding how the search results are arranged and the relevancy effectiveness.

## 2.3.2 ISSUES WITH IDOL

1. **Support:** While the IDOL product is able to index multiple content sources, it has proven to be hard to support and maintain. Documentation and support have been lackluster and there appear to be few outside companies providing consulting on it.
2. **Slow Indexing:** Creating and maintaining the indexes take a long time. While the bulk of our data only needs to be re-indexed infrequently, it takes up to 2 weeks to index just the death records index. Further, there is no mechanism to easily apply changes to index entries; if the content on a particular record is modified, via change or delete operation, customized programs must be written to update the corresponding IDOL index entry.
3. **Cost:** The ongoing costs for using IDOL may prove to be unsustainable under our current funding model.
4. **Autonomy's Market Position:** With the dominance of Google, and Microsoft's acquisition of Fast Search and their attempt to acquire Yahoo!, the search engine marketplace is becoming concentrated in a few providers. Autonomy may lose market position and the ability to attract technologists to continue to improve and support its products.

## 3 INFORMATION BEING REQUESTED

This is identical to first paragraph and not needed

### 3.1 DETAIL

The following sets of requirements are meant to provide more detail about our current environment and goals for our future solution. Please address them in your response.

#### 3.1.1 CONTENT INGESTION

This is a list of characteristics of content that is ingested into the GRN.

1. from a variety of sources
  - a. data from partner sites not under MHS control
  - b. data from databases
  - c. federated systems, APIs, or web services
2. defined standards, including
  - a. Open Archive Initiative (OAI)
  - b. Encoded Archival Description (EAD)
  - c. Encoded Archival Context (EAC)
  - d. MARC
  - e. Dublin Core
3. fielded information
  - a. about people "nominal indexes" (records with people's names, dates, location, event - birth, death, marriage, etc.)
  - b. bibliographic
4. data from websites
  - a. identify information on web pages, like microformats
5. data formats
  - a. spreadsheets
  - b. PDF
  - c. text documents
  - d. images
  - e. audio
  - f. video
  - g. XML
  - h. DBMS
6. support data from feeds such as OAI or RSS
7. from individuals leaving commentary about items in the system
  - a. which would be persistently linked to those items even as the system is updated and reindexed

## 3.1.2 USER SEARCHES

This is a list of characteristics of GRN search behavior.

1. supports different entry points for searching (see mnhs.org home page search box, History Finder, and People Finder)
2. using full-text searching techniques familiar from experience with Google (describe)
  - a. Phrase search ("")
  - b. Words you want to exclude (NOT)
  - c. Wild cards (\*)
  - d. Exact match (+)
  - e. Synonyms
  - f. The OR operator
  - g. The AND operator

3. using fielded techniques critical to searching nominal indexes (such as birth and death records, for example, see [people.mnhs.org/bci](http://people.mnhs.org/bci)), and library catalog (for example, see <http://mnhs.mnpals.net>)
  - a. results from fielded searches should be sortable.
4. performs federated searches to external collections
5. search within the result set
6. navigable facets for refining searches
7. results must be delivered within 2 seconds
8. save search queries

---

### 3.1.3 SEARCH RESULTS

This is a list of characteristics of search results behavior.

1. returns results contextually based on entry point
  - a. people finder searches, for example see [www.mnhs.org/peoplefinder](http://www.mnhs.org/peoplefinder)
  - b. unformatted searches, for example see the search box in the header at [www.mnhs.org](http://www.mnhs.org)
  - c. place finder searches
  - d. site searches
  - e. partner searches
2. provides heterogeneous results in one result set
3. provide direct viewing of objects in the result set
  - a. for example, images should be presented as thumbnails rather than just lists of titles
4. partners should be able to view results in a view that is "skinned" with their organizational identity

---

### 3.1.4 WEB SERVICES

This is a list of characteristics of how GRN content should be available broadly to the world wide web.

1. site maps available to crawlers
2. search API
  - a. provide "raw" results (e.g. XML, JSON)
3. easy integration into the web ecosystem
  - a. for example, queries should be expressed in the URL so they can be easily turned into deep links on other sites

---

### 3.1.5 SYSTEM TOOLS/ADMINISTRATION

This is a list of characteristics administration tasks at MHS and its partners.

1. workflow tasks
  - a. controllable indexing/harvest schedules

- b. testing changes in a test system
- c. building and modifying filters for different data types
- d. building and modifying crosswalks for fielded data
2. the system should allow partners to initiate certain changes
  - a. how would partners change the schedule
  - b. how would partners modify the skin used to present results in 3.2.3
  - c. how would partners suggest modifications in the mapping of their data
3. indexing and reindexing
  - a. workflow
  - b. speed
  - c. flexibility and time for indexing and re-indexing
    - i. all collections
    - ii. individual collections or portion of collection
    - iii. strategies for controlling the process
4. system management information
  - a. site and query statistics
  - b. harvest schedule
  - c. all collections and individual collections
    - i. granularity
5. sample of guidelines and documentation for
  - a. system
  - b. data formats
  - c. incorporation of non-standard data formats
  - d. APIs
  - e. security/user level
6. community support/user groups
7. training
8. maintenance and emergency support with references

---

## 3.1.6 SYSTEM SUSTAINABILITY

These characteristics provide the starting point of a conversation about the sustainability of your solution. Final specifications will be provided in the RFP process.

1. describe costs and maintenance
  - a. average costs in dollars of implementing and maintaining the system, including upgrades
  - b. maintenance tools available and tasks involved in running the system
  - c. major and minor releases/upgrades
    - i. frequency within last 3 years
    - ii. instructions for last two releases/upgrades with references
  - d. system enhancements



- i. components available for user modification
    - ii. components not available to user modification
    - iii. how is new functionality added into the system
    - iv. how are evolving standards and formats accommodated
2. describe support available for the system
  - a. company provided technical support
  - b. other professional support/input
  - c. user community
3. describe licensing of system and components
  - a. is source code available, for what parts, and what are the costs?
  - b. under what licenses are components of the system distributed and what are the related costs?
  - c. what restrictions would inhibit our sharing the system or components with others?

## 4 INSTRUCTIONS FOR RESPONDING TO THIS RFI

### 4.1 WHO MAY RESPOND

Responses from anyone in industry, government or academia with practical knowledge of content ingestion and search are welcome.

### 4.2 HOW TO RESPOND

One electronic copy in pdf format should be sent to [jane.wong@mnhs.org](mailto:jane.wong@mnhs.org) and [Kathryn.ludwig@mnhs.org](mailto:Kathryn.ludwig@mnhs.org). One confirming paper copy of all documents should also be sent to the postal address below.

Kathryn Ludwig  
345 Kellogg Boulevard West  
Saint Paul, Minnesota 55102

Responses to this RFI must be received at MHS no later than 2:00 PM US Central Time on May 14, 2009.

Other communication regarding this RFI should be sent to the contact listed in paragraph 4.7.

### 4.3 RFI RESPONSE CONTACT

Companies responding to this RFI shall designate a single contact within that company for receipt of all subsequent information regarding this RFI and the forthcoming series of RFPs.

## 4.4 FORMAT OF RFI RESPONSES

The following outline is offered to assist in the development of your response. You should include:

- A cover letter briefly summarizing your response, such as indicating to which areas you are responding and whether supporting documentation is included in your response.
- Your response covering any or all of the areas of information requested by this RFI.

Although the MHS will not limit the size of responses, it encourages you to consider limiting your response (not counting any supporting documentation) to approximately 25 pages. If you consider supporting documentation to be necessary, please indicate which portions of the supporting documentation are relevant to this RFI.

## 4.5 DISTRIBUTION OF RFI RESPONSES

Your information will be made available to a consultant working with the MHS for review purposes.

## 4.6 REIMBURSEMENT

The MHS will not reimburse submitters for any costs in conjunction with their responses to this RFI.

## 4.7 QUESTIONS REGARDING THIS RFI

Any technical questions regarding this RFI should be sent to:

Marj Kelly  
marj.kelly@mnhs.org  
345 Kellogg Boulevard West  
Saint Paul, Minnesota 55102

## 5 RESPONSE REVIEW PROCESS AND SCHEDULE

### 5.1 REVIEW PROCESS

This RFI is issued with the intent to survey industry to obtain information that provides guidance, which will inform preparation of the Society's RFP(s). After review of responses and presentations by invitation, the MHS will prepare for its next steps and issue one or more RFPs.

## 5.2 CLARIFICATION

To fully comprehend the information contained within a response to this RFI, the reviewing group may seek further clarification on that response. This clarification may be requested in the form of brief verbal communication by telephone; written communication; electronic communication; or a presentation.

## 5.3 RFI RESPONSE PRESENTATIONS AND DEMONSTRATIONS

RFI respondents may be invited to present their response with a technology demonstration. The purpose of this presentation would be to seek clarification of information contained within the response (as noted above); to further explore issues raised; or to further meet the goals of the RFI.

## 5.4 SCHEDULE

The schedule for responding to this RFI is as follows. Please note that early responses are encouraged.

RFI issued:	April 13, 2009
RFI responses due:	May 14, 2009
Review of RFI responses:	May 18 – June 5, 2009
Vendor Presentations	June – July 2009

## APPENDIX A PARTNER PROFILES

This appendix provides a profile of the GRN partners. This information was compiled from information provided by the partners or from information on partner web sites. Minnesota Historical Society

### Minnesota Historical Society

The MHS is the lead or development, partner. The GRN team consists of 5 MHS staff at an equivalent of 2 FTE. Two are from Enterprise Technology, one is the Director of the Library, Publications and Collections Division, and two are from the Central Collections Services Unit. At times, contracted professional services have been used to assist in IDOL configuration and programming support.

### Anoka County Historical Society, Minnesota

No partner profile provided. <http://www.ac-hs.org/>

### Blue Earth County Historical Society, Minnesota

The Blue Earth County Historical Society, <http://www.bechshistory.com/>, collects, preserves and promotes the history of Blue Earth County, Minnesota.

- Annual Budget: \$175,000
- IT Budget: 0 IT budget
- Staff: 3 full-time, 1 part-time
  - IT Staff: 0, volunteer
- Web Visits Per Year: 2007 = 15,000
- Total Requests: 2007 = 500

### Nicollet County Historical Society, Minnesota

The Nicollet County Historical Society, <http://www.nchsmn.org/>, is located in the heart of the Minnesota River Valley.. NCHS operates four sites: the Treaty Site History Center headquarters, adjacent Traverse des Sioux Historic Site, E. St. Julien Cox House, Harkin Store, and Fort Ridgely Historic Site. In 2008 the Society has served 55,000 visitors through admissions, outreach, and other services. NCHS continues to see significant growth in annual numbers served. Our archival collection is responsible for the single-greatest increase in those numbers.

- Annual Budget: \$240,000
- IT Budget: \$300
- Staff: 15 total, but only 2 full-time and 10 only part-time seasonal

- IT Staff: 0
- Web Visits Per Year: 30,000
- Total Requests: 70,000+

## Olmsted County Historical Society, Minnesota

No partner profile provided. <http://www.olmstedhistory.com/>

## State Archives Division of the State Historical Society of North Dakota

The State Archives, <http://history.nd.gov/archives/index.html>, division is responsible for the documentary collections of the State Historical Society of North Dakota, <http://history.nd.gov>. It is the official state archives, and acquires and preserves all types of research materials relating to North Dakota and the Northern Great Plains, including manuscript collections, books, periodicals, maps, newspapers, audio and video materials, and photographs.

- Annual Budget: \$55,000
- IT Budget: \$10,000 – \$10,500
- Staff: 11
- IT Staff: 0
- Website Visits: 2,076,434
- Other requests for information: 12,000 annually
- Types of content: have already provided a marriage index, will be providing more

## Digital Horizons

Digital Horizons <http://digitalhorizonsonline.org/> is an online treasure house of thousands of images, documents, video, and oral histories depicting life on the Northern Plains from the late 1800s to today. Digital Horizons was established in 2007 by a consortium including: Concordia College Archives, Moorhead, Minn.; Institute for Regional Studies, North Dakota State University, Fargo, ND.; Prairie Public, Fargo, ND.; and State Historical Society of North Dakota, Bismarck, ND. Types of content: images, videos, documents, etc. from a ContentDM based web site.

## North Dakota Genealogical Society

Types of content: have already provided a marriage index, will be providing more

## North Dakota State University

<http://library.ndsu.edu/> Types of content: have already provided a marriage index, will be providing more.

## State Historical Society South Dakota

<http://www.sdhistory.org/> in consortium with a digital library consortium consisting of universities.

## Minnesota Digital Library: Minnesota Reflections

Minnesota Reflections, [reflections.mndigital.org](http://reflections.mndigital.org), has nearly 31,000 images and documents in ContentDM, shared by more than 95 cultural heritage organizations across the state, including the GRN partners from Minnesota. This site offers a broad view of Minnesota's history for researchers, educators, students, and the public. Types of content: images, documents, videos, etc.

## APPENDIX B SUMMARY OF PARTNER CONTENT

The following table shows the types of content our partners could provide for inclusion in the GRN. The content has been put into columns based on format. Content listed in the unknown column has been listed but we do not have samples of it.

PastPerfect software – [www.museumsoftware.com](http://www.museumsoftware.com)

Partner	Nominal Data	Standards Based Data	Other Data	Unknown
Minnesota Historical Society	<ul style="list-style-type: none"> <li>-Birth Records Index 1,822,145 records MS-SQL database; records have digital image</li> <li>-Death Records Index 2,960,521 records in MS-SQL database</li> <li>-Census Records Index &gt; 5,000,000 records in MS-SQL database, records have digital image</li> <li>-Veterans Grave Index 72,105 records MS-SQL database</li> <li>-Wotr comments in MySQL</li> </ul>	<ul style="list-style-type: none"> <li>-Finding Aids (EAD) over 1,000 documents, pdf documents and some XML encoded EAD documents</li> <li>- Library Catalog (OPAC) MARC records in XLibris storage; z.39.50 connection to MNPals system</li> </ul>	<ul style="list-style-type: none"> <li>- MHS website 19,086 document/records HTML, PDF documents some data pulled from MS-SQL databases listed below</li> <li>-Events Calendar ~160 records</li> <li>-News Releases ~1,020 records</li> <li>-Duluth Lynchings 2,390 records</li> <li>-Jerome Hill Papers 2,540 records</li> <li>-MHS Online Store ~1,050 records</li> <li>-Book of Days 1,678 records</li> <li>-TimePieces 339 records</li> <li>-NRHP 1,782 records</li> <li>-Photo and Art Database 235,745 records</li> <li>MS-SQL database, many records have digital</li> </ul>	

# Search Engine Replacement RFI 2009

Partner	Nominal Data	Standards Based Data	Other Data	Unknown
			images - MN Place Names 10,093 records MS-SQL database - MN History Articles ~2,000 documents in PDF - KE Software EMu TexPress database, XML	
Anoka County Historical Society		Some collections in Past Perfect but we have not gotten samples	Photographs Club Collection 218 Rows, 7 columns Excel spreadsheet Maps 195 rows, 7 columns Excel spreadsheet	
Blue Earth County Historical Society	Will Index (1858-1899) 589 rows, 3 columns Excel spreadsheet 1947 Marriage index 535 rows, 4 columns Excel spreadsheet	Book collection Past Perfect		Sermon Index (1945) 916 rows, 3 columns Excel spreadsheet Obit index Excel spreadsheet 3-D collections Access Social Notes 1865 State Census WWII Newspaper Articles Manuscripts Yearbook Inventories



# Search Engine Replacement RFI 2009

Partner	Nominal Data	Standards Based Data	Other Data	Unknown
				Maps
Nicollet County Historical Society	<p>Fort Ridgley Cemetery Index 18 pages Word document</p> <p>Family histories 1 page Word document</p> <p>List of Persons who sought refuge at Fort Ridgley 7 pages Word document</p> <p>Index of Surnames to 1985 (newspaper announcements) 5 pages Word document</p> <p>Obits 12,046 records Excel spreadsheet</p> <p>Surname List (newspaper announcements) Excel spreadsheet</p>	<p>3-D Objects 2,300 records Past Perfect</p> <p>Photos over 10,000 records Past Perfect</p>	<p>Album databases 181 rows, 3 columns Excel spreadsheet Atlases 166 rows, 3 columns Excel spreadsheet Book inventory 1394 rows, 6 columns Excel spreadsheet Manuscript collection 674 rows, 3 columns Excel spreadsheet and Access Video lists 103 rows, 4 columns Excel spreadsheet Index to St. Peter Herald, 1985 6 pages Word document Index to St. Peter Herald's 1930 7 pages Word document Index to Old Traverse des Sioux 6 pages Word document Nicollet County Cemeteries 4 pages</p>	<p>Manuscripts</p>

# Search Engine Replacement RFI 2009

Partner	Nominal Data	Standards Based Data	Other Data	Unknown
			Word document Subject Index 5 pages Word Document	
Olmsted County Historical Society		3-D collections 9,000 records, all have images attached Past Perfect Photos 17,000 records, all have images attached Past Perfect		Newspaper index Alpha 5 Biographical index Alpha 5
Minnesota Digital Library / <i>Minnesota Reflections</i>		Photo collection ~20,000 records OAI Feed		
Digital Horizons		Photo collection ~23,000 records OAI Feed		
North Dakota State Archives	Marriage Records (Stark Co.) 12, 587 rows, 8 columns Excel spreadsheet Provided by ND			

# Search Engine Replacement RFI 2009

Partner	Nominal Data	Standards Based Data	Other Data	Unknown
	Genealogical Society Marriage Records (Cass Co.) 14,954 records Access Provided by NDSU Marriage Records 32,288 records Access Provided by ND Historical Society			