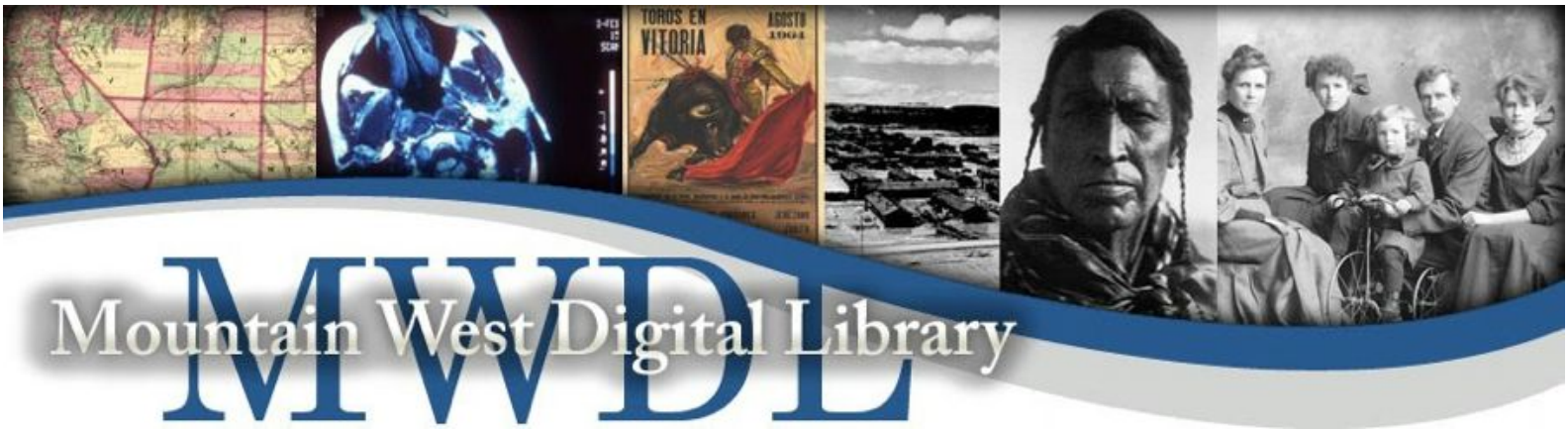

Best Practices

for

Geospatial Metadata Creation

for

MWDL Cultural Heritage Digital Repositories

Final Report of the
Mountain West Digital Library Geospatial Discovery Task Force
July 2015

MWDL Geospatial Discovery Task Force members

Co-Chairs

Kristen Jensen, Utah Department of Heritage and Arts
Liz Woolcott, Utah State University

Phase One Contributors

Anna Neatrour, Mountain West Digital Library *(Team Lead)*
Liz Woolcott, Utah State University Merrill-Cazier Library *(Team Lead)*
Michelle Olsen, Snow College, Richfield Campus Library *(Team Lead)*
Bert Granberg, Utah Automated Geographic Reference Center
Christian Sarason, OCLC
Devin Becker, University of Idaho Library
Greta Bahnemann, Minnesota Digital Library
Jeremy Myntti, University of Utah, Marriott Library
Ken Rockwell, University of Utah Marriott Library
Kristen Jensen, Utah Department of Heritage and Arts
Paula Mitchell, Southern Utah University
Ray Matthews, Utah State Library
Robyn Keeling, Utah Geological Survey Library
Sandra McIntyre, Mountain West Digital Library
Silvia Southwick, University of Nevada Las Vegas
Sowmya Athithan, Utah Valley University

Phase Two Contributors

Anna Neatrour, Mountain West Digital Library *(Team Lead)*
Dustin Olson, Utah State University Merrill-Cazier Library (Team Lead)
Kristen Jensen, Utah Department of Heritage and Arts *(Team Lead)*
Rachel Wittmann, Clemson University Library *(Team Lead)*
Bert Granberg, Utah Automated Geographic Reference Center
Christian Sarason, OCLC
Devin Becker, University of Idaho Library
Greta Bahnemann, Minnesota Digital Library
Heidi Orchard, Utah Department of Heritage and Arts
Jeremy Myntti, University of Utah Marriott Library
Ken Rockwell, University of Utah Marriott Library
Liz Woolcott, Utah State University Merrill-Cazier Library
Paula Mitchell, Southern Utah University Sherratt Library
Sandra McIntyre, Mountain West Digital Library
Silvia Southwick, University of Nevada Las Vegas Library

Phase Three Contributors

Kristen Jensen, Utah Department of Heritage and Arts *(Team Lead)*
Liz Woolcott, Utah State University Merrill-Cazier Library *(Team Lead)*
Anna Neatrour, Mountain West Digital Library
Catherine McIntyre, Utah Valley University Library
Greta Bahnemann, Minnesota Digital Library
Jeremy Myntti, University of Utah Marriott Library
Ken Rockwell, University of Utah Marriott Library
Rachel Wittmann, Clemson University Library
Ray Matthews, Utah State Library
Sandra McIntyre, Mountain West Digital Library
Scott Eldredge, Brigham Young University, Harold B. Lee Library
Silvia Southwick, University of Nevada Las Vegas Libraries

Additional Members

Amy Rudersdorf, Digital Public Library of America
Anne Morrow, University of Utah Marriott Library
Andy Wesolek, Clemson University Libraries
Becky McKown, Brigham Young University Harold B. Lee Library
Cheryl Walters, Utah State University Merrill-Cazier Library
Dave Morrison, University of Utah Marriott Library
Devin Becker, University of Idaho Library
Dorotea Szkolar, former geospatial metadata intern, MWDL
Elizabeth Perkes, Utah State Archives
Gina Strack, Utah State Archives
Geri Ingram, OCLC
Jason Roy, University of Minnesota Libraries
Kinza Masood, University of Utah Marriott Library
Lorelei Rutledge, University of Utah Marriott Library
Matt Pierce, University of Utah Quinney Law Library
David Monk, LDS Church History Library
Nick Hayen, Mountain West Digital Library
Rebekah Cummings, Mountain West Digital Library
Robyn Krohn, Utah State Library
Christopher Vinson, Clemson University Libraries
Glee Willis, University of Nevada Reno Libraries

Table of Contents

Best Practices for Geospatial Metadata Creation for MWDL Cultural Heritage Digital Repositories

Final Report of the
Mountain West Digital Library Geospatial Discovery Task Force

Submitted: July 31, 2015
Approved:

Executive Summary

In July of 2013, the Mountain West Digital Library Geospatial Discovery Task Force convened with the purpose of developing a standard format for recording geospatial metadata and developed the following Task Force charge:

1. **Identify existing** geospatial metadata practices used by Mountain West Digital Library (MWDL) partners and other digital libraries, as well as other disciplines and organizations.
    a. Identify areas for potential development and improvement (e.g., key issues and problems)
    b. Identify easily achievable practices that MWDL partners can implement in the interim before a final recommendation is made.
    c. Identify examples of map-based search Interfaces, successful and unsuccessful
2. **Develop guidelines** for MWDL partners to use in **standardizing metadata practices** to optimize geographically-based discovery of digital resources.
    a. Suggest steps for addressing key issues and problems
    b. Recommend revisions to the MWDL Dublin Core Application Profile to accommodate geospatial metadata
3. Develop **recommendations and instructions for creating map-based search interfaces** or augmenting existing interfaces with map-based search functionality.
4. **Identify and share tools** related to geospatial metadata and the creation of map-based search interfaces.
5. Create the following **deliverables**.
    a. An online bibliography of resources
    b. Guidelines for MWDL geospatial metadata practices
    c. Recommendations and instructions for creating map-based search interfaces or search functionality
    d. Final report on the task force's activities, findings, and recommendations

During the course of this process, the task force identified two limitations on the initial charge:

1. Technical requirements and recommendations for map-based interfaces were beyond the scope of the task force, since they depended too heavily on the type of digital repository used by an institution and not solely on the format of geospatial metadata recorded.
2. Geospatial metadata recommendations would need to focus primarily on “place name” metadata, since this was the most applicable type of metadata for cultural heritage objects. GIS and geospatial metadata represented as data sets were outside the scope of the majority of content provided by MWDL repositories.

The final recommendation of the task force was two-fold: 1) that MWDL partners use the GeoNames database as the primary recommended controlled vocabulary to create a geospatial metadata entry and 2) at minimum, geospatial metadata would present in a hierarchy from smallest entity to largest, with the hierarchy split by commas. For example, Boise, Idaho would be represented as:

ex.: ***Boise, Ada County, Idaho, United States***

Additionally, MWDL partners are highly encouraged to provide the URI following the name hierarchy in order to begin laying the groundwork for future involvement in linked data initiatives. All place name metadata referring to the same place should be separated by commas, only. Semi-colons should only be used to indicate separate and distinct places within the same geospatial field. An example entry for Boise, Idaho would look like this:

ex.: ***Boise, Ada County, Idaho, United States, http://sws.geonames.org/5586437/***

Additionally, if any institution would like to include latitude and longitude, these coordinates can be included after the place name hierarchy and URI. An example would include:

ex.*:* ***Boise, Ada County, Idaho, United States, http://sws.geonames.org/5586437/, 43.6135, -116.20345***

To aid in the creation of geospatial metadata, instructions were developed for using the GeoNames database. These included examples of the construction of geospatial metadata for a variety of place types, and included instructions on how to include multiple geospatial metadata elements, such as the URI and latitude and longitude. Please see Appendix III for a link to the GeoNames Instructions). The full list of recommendations is highlighted in Section 1 of this document.

Methodology

These recommendations were developed over the course of more than two years from the efforts of the Mountain West Digital Library (MWDL) Geospatial Discovery Task Force. The task force was constituted following the research work of MWDL graduate school intern Dorotea Szkolar in the summer of 2012. Dorotea documented the widely varying practices of the MWDL collections partners in managing geospatial metadata, and described emerging international standards. The task force was formed in July 2013 and was composed of 39 members from MWDL partner repositories and non-MWDL repositories from around the country.

The task force operated in three phases. During the first two phases, the task force split into three subgroups that analyzes specific research topics. The first phase was conducted from July through December 2013 and included subgroups that 1) examined Dorotea Szkolar's report on geospatial metadata, 2) developed suggestions for "low-hanging fruit" or recommendations that could be implemented easily before the final conclusions of the task force, and 3) looked for examples of successful map-based interfaces for digital repositories. The findings were reported to the whole task force and archived on the Mountain West Digital Library Geospatial Discovery Task Force website. (See Appendix I).

The second phase was conducted from January through July 2014 and also split into three subgroups that took the information from the first phase subgroups and researched recommendations further. Those subgroups looked at 1) controlled vocabulary options, 2) formats for recording coordinate information, and 3) the technical specifications for the previously identified successful map-based interfaces. The reports and meeting minutes of these subgroups were recounted to the whole task force and archived at the MWDL Geospatial Discovery Task Force website (see Appendix I). The preliminary findings of the first two phases were reported at several conferences, including the Digital Library Federation in October 2014.

The third and final phase was conducted from January 2015 through July 2015. During this final phase, the task force acted on the recommendations of the previous phase and looked into the analysis of controlled vocabularies. Due to the straight forward nature of the work in the third and final phase, the task force did not split into subgroups. The work conducted during the third phase included selecting the GeoNames database as the recommended controlled vocabulary for MWDL partners and developing instructions on how to format the metadata entry and use the GeoNames database.

During the course of this process, several recommendations were made for future task force groups to look at. Section 2 - Future Task Force Activities and Section 3 - Recommendations for UALC Digitization Committee Actions in the recommendation list below outlines possible future tasks to be done.

Final Recommendations

1. Recommendations for Standardizing Geospatial Metadata

1.1. GeoNames is the preferred geographic database from which to derive controlled vocabulary terms, uniform resource identifiers (URI)s, and latitude and longitude information for metadata creation.

1.1.1. The preferred format for geospatial metadata will follow this pattern: place name hierarchy, URI, Latitude, Longitude, with the place name hierarchy being the minimal accepted format.

Examples:

Minimal accepted format:
Aurora (historical), Mineral County, Nevada, United States

Highly recommended format:
Aurora (historical), Mineral County, Nevada, United States, http://sws.geonames.org/5499519/

Additional accepted format:
Aurora (historical), Mineral County, Nevada, United States, http://sws.geonames.org/5499519/, 38.28714, -118.9007

1.1.2. Place name hierarchy should be presented from smallest to largest, with places spelled out, separated by commas.

1.1.3. Latitude-longitude coordinates should be expressed as decimal degrees, without directional letters (“N”, “W”, etc.), and separated by commas.

1.1.4. When referring to more than one place, place semi-colons only between each unique place while retaining the commas between the elements that describe each place.

Example:
Aurora (historical), Mineral County, Nevada, United States, http://sws.geonames.org/5499519/**;** Phoenix, Maricopa County, Arizona, United States, http://sws.geonames.org/5308655/

1.1.5. Additional geospatial information (such as street addresses) or metadata that needs to be entered in a different order than what is prescribed in 1.1.1 should be placed in unmapped metadata fields.

1.2. A metadata value in a field mapped to the spatial coverage refinement (dcterms:spatial) is recommended for all records harvested by MWDL. The mapping to dcterms:spatial can be done at the collection level. The OAI provider

for the repository hosting the collection should support provision of qualified Dublin Core.

1.3. A metadata value in a field mapped to the spatial coverage refinement (dcterms:spatial) is highly recommended for all records in new collections harvested by MWDL.

1.4. Where converting legacy data may be too difficult, partners can add an additional separate field mapped to the Dublin Core term spatial (dcterms:spatial) with basic, minimal geospatial metadata **at least at county or county equivalent** (e.g., parish, borough, shire) level, in accordance with GeoNames.

1.5. Since MWDL contributors may need to use varied controlled vocabularies, it is recommended that conformance to these standards be highly recommended but not enforced. Failure to adhere to these standards will not lead to collections being excluded from harvest. However, MWDL and DPLA may not be able to interpret spatial information that is expressed outside the parameters of these standards.

1.6. All standards and practices adopted by the metadata review board should be compliant with applicable ISO standards for geographic metadata.

2. Future Task Force Activities

2.1. Based off the recommendations for a preferred controlled vocabulary, create actionable plans for Collections Partners and Member Repositories to deal with legacy geospatial data including:

2.1.1. Creating a list of common find and replace scenarios that may be useful for all who contribute their collections to MWDL.

2.1.2. Performing a global search and replace of simple geospatial metadata with more in-depth formatted spatial metadata when possible.

2.1.3. Having a subgroup look at the top five strategies currently in use by MWDL partners to assign geospatial metadata and estimate what would need to be done to convert legacy data.

2.2. Have a subgroup look at adopting and/or developing a gazetteer of regional place names that are missing from international controlled vocabularies.

2.3. Review the formatting and syntax of Points and Boxes, particularly in regards to the DCMI Box/Point Encoding Schemes.

2.4. Look further into GeoJSON vs. KML for presenting spatial data on map interfaces.

3. Recommendations for UALC Digitization Committee actions

3.1. Formally revise the MWDL Dublin Core Application Profile element “spatial” to reflect the above recommendations.

3.2. As part of MWDL’s future work with updating the MWDL Dublin Core Application Profile and General Guidelines for Digital Collections Metadata, MWDL should foster a larger discussion on the topics of subject coverage vs. geospatial coverage, and how the two areas may relate.

Appendices

Appendix I - Geospatial Website
https://sites.google.com/site/mwdlgeospatial/home

Appendix II - Reports and Meeting Minutes of the Entire Task Force
https://sites.google.com/site/mwdlgeospatial/home/meeting-minutes

Appendix III - GeoNames Instruction
https://docs.google.com/document/d/1vmEy1vPcaKfhY0FAiKUgnVAvn5S6iDxBErf1X7ikuoA/edit#heading=h.ju4w3x7re9z1