

Metadata Requirements for the Minnesota Geospatial Commons – Draft Recommendation

June 19, 2015

Authors:

Chris Cialek, MnGeo
Joella Givens, MnDOT
Mark Kotz, Metropolitan Council
Nancy Rader, MnGeo

This draft has been revised in response to comments submitted by members of the Metadata Workgroup, the Commons Project Team and the Chief Geospatial Information Officer.

Introduction

The Minnesota Geospatial Commons requires specific metadata information to function properly. To date, the Minnesota Geographic Metadata Guidelines (MGMG) have not specified mandatory elements; more specific requirements are now needed to ensure the quality and functioning of the Commons. At the direction of the Minnesota Statewide Geospatial Advisory Council, the Standards Committee Metadata Working Group has reviewed all elements in the MGMG and has developed this draft recommendation identifying elements that are **mandatory** to participate in publishing data resources to the Commons, those that are **desirable** because their inclusion improves the quality and value of the document, and those that are **optional** and thus left to the metadata author's discretion. This draft recommendation is now being forwarded for review by the Statewide Geospatial Advisory Council and the Geospatial Technical Committee.

Goal

Determine which MGMG elements are required in order for a data resource¹ to be published and accepted to the Commons, and also provide guidance on the quality and quantity of information expected in metadata to be published through the Commons.²

Governance Context

The two Commons Governance Documents referenced below (*Expectations for Commons Publishers* and *How to Become a Publisher on the Commons*) provide guidance to Commons publishers in order to assure a standard quality to the content of this statewide repository. The following guidance from those documents, in its current wording, is relevant to the goal of this effort:

1. [Expectations for Commons Publishers](#):
 - Guiding Principle: Resources published on the Commons must have sufficient documentation to allow a potential user to determine if they are fit for a particular use.

¹ This document covers **data** resources, not **application** resources (application resources currently are not required to have full metadata in the Commons, although it can be accommodated if publishers provide it). The expectation is that the application (web map, software, etc.) will provide sufficient documentation for use; in addition, the MGMG elements may not be an appropriate way to provide this documentation.

² This document is not intended to provide *complete* "best practices" guidance to writing MGMG metadata, although that is a high priority for updating MGMG training materials.

DISCUSSION DRAFT

- While a minimum level of metadata completeness is required to publish within the Geospatial Commons, complete metadata is expected.
- The Commons administrator has the authority to revoke users and publishers if they fail to comply with policies.

2. [How to Become a Publisher on the Commons](#):

How much metadata is required to publish my dataset?

While we haven't defined the minimum number of metadata fields that must be completed for a dataset to be published on the Commons, the goal is to define a level and quality of required metadata that will provide enough information for a user to evaluate a dataset while not making the documentation task too much for the data publisher. Metadata quality will be maintained by developing best practice guidance, by periodically reminding publishers to check their records and via user feedback. The Commons uses the Minnesota Geographic Metadata Guidelines.

Criteria used to decide which metadata elements are mandatory

An MGMT metadata element needs to meet at least one of the following two criteria to be considered "mandatory":

- It provides information needed for a potential user to determine fitness for their use.
- It provides information needed in order to support a Commons function (e.g., bounding box coordinates are needed to draw the extent box on the map and to support geographic searches or filtering using the map).

Categories

For the purposes of this draft recommendation, each metadata element has been placed in one of four categories in order to indicate the consequences of being excluded³:

- **Mandatory:** The data resource will not be published until this element is adequately populated ("unknown" is a valid response when the information is not available).
- **Mandatory if applicable:** The data resource will not be published until this element is adequately populated, if, and only if, the element applies to the data resource. If the element does not apply to the data resource, then it can be left blank and the data resource will be published.⁴
- **Desirable:** The data resource can be published without this element, but best practice indicates that it should be filled in so that resource users are well informed.
- **Optional:** The data resource can be published without this element; the publisher determines whether or not to fill it in.

³ Currently, several elements are automatically validated during the Commons publishing process and reports are generated with these categories:

- Error: Depending on the type of error, the data resource cannot be published either to the GDRS or the Commons until the error is fixed.
- Warning: Element is flagged and publishers are encouraged to address the issue, but the data resource can continue in the publishing process.
- Valid: No issues found during validation.

⁴ For example, the *Cell Width* and *Cell Height* elements are only mandatory for raster data formats; they do not apply to vector data formats.

DISCUSSION DRAFT

The next two sections of the document present a summary of the recommendation, followed by a more detailed description for each element of the Minnesota Geographic Metadata Guidelines. The final section touches on implementation options.

Summary recommendation

SECTION 1 - Identification	Mandatory	Mandatory if applicable	Desirable	Optional
Originator	✓			
Title	✓			
Abstract	✓			
Purpose			✓	
Time Period of Content Date		✓		
Currentness Reference		✓		
Progress	✓			
Maintenance and Update Frequency	✓			
Spatial Extent of Data			✓	
Bounding Coordinates	✓			
Place Keywords	✓			
Theme Keywords	✓			
Theme Keyword Thesaurus				✓
Access Constraints	✓			
Use Constraints	✓			
Contact person			✓	
Contact organization	✓			
Contact position			✓	
Contact address				✓
Contact city				✓
Contact state				✓
Contact postal code				✓
Contact voice phone	✓			
Contact fax				✓
Contact email	✓			
Browse graphic filename			✓	
Browse graphic description				✓
Associated datasets				✓
SECTION 2 – Data Quality				
Attribute accuracy	✓			
Logical consistency			✓	
Completeness	✓			
Horizontal positional accuracy	✓			
Vertical positional accuracy		✓		
Lineage	✓			
SECTION 4 – Spatial Reference				
Horizontal coordinate scheme	✓			

DISCUSSION DRAFT

Horizontal datum	✓			
Horizontal units	✓			
Altitude datum		✓		
Altitude units		✓		
Depth datum		✓		
Depth units		✓		
Cell width		✓		
Cell height		✓		
SECTION 5 – Entities & Attributes	Mandatory	Mandatory if applicable	Desirable	Optional
Entity and attribute overview		✓		
Entity and attribute detailed citation		✓		
SECTION 6 – Distribution				
Publisher	✓			
Publication Date				✓
Distributor contact person			✓	
Distributor contact organization	✓			
Distributor contact position			✓	
Distributor contact address				✓
Distributor contact city				✓
Distributor contact state				✓
Distributor contact postal code				✓
Distributor contact voice phone	✓			
Distributor contact fax				✓
Distributor contact email	✓			
Distributor’s dataset ID				✓
Distribution liability			✓	
Ordering instructions			✓	
Online linkage	✓			
SECTION 7 – Metadata				
Metadata date	✓			
Metadata contact person			✓	
Metadata contact organization	✓			
Metadata contact position			✓	
Metadata contact address				✓
Metadata contact city				✓
Metadata contact state				✓
Metadata contact postal code				✓
Metadata contact voice phone	✓			
Metadata contact fax				✓
Metadata contact email	✓			
Metadata standard name				✓
Metadata standard version				✓
Metadata standard online linkage				✓

DISCUSSION DRAFT

Detailed recommendation for each element

MGMG Element <i>For definitions of the elements, click on the section title link</i>	What is Required	Comments <i>Related to the recommendation for each element, not intended to provide complete best practices or examples</i>
<u>SECTION 1 - Identification</u>		
Originator	Mandatory	Users need to know the name of the organization(s) or individual(s) that developed the data set. This element is displayed on the Commons summary page for a data resource.
Title	Mandatory	Best practice is to start with the topic, followed by the geography covered, followed by a date, followed by anything else, such as originator, that's needed to distinguish this data resource from other similar ones. This element is displayed on the Commons summary page for a data resource.
Minnesota Clearinghouse ID		Never used – drop this element. Not in the Minnesota Metadata Editor (MME) .
Abstract	Mandatory	At least a short paragraph communicating the most important information about the data resource. The main point should be in the first sentence or paragraph. Users browsing the Commons will most often scan the <i>Title</i> and <i>Abstract</i> to decide whether or not to delve further. This element is displayed on the Commons summary page for a data resource.
Purpose	Desirable	For many data resources, understanding why the data was collected helps users decide whether or not it would be valuable for their purposes. This element is displayed on the Commons summary page for a data resource.
Time Period of Content Date	Mandatory if applicable	<p>Date information is mandatory; however, the information can be provided either in <i>Time Period of Content Date</i>, <i>Currentness Reference</i> or both, as appropriate. Date information is displayed on the Commons summary page for a data resource.</p> <p>This field is intended for a single date, month or year in the format YYYYMMDD or YYYYMM or YYYY. Its purpose was to enable users to search for data “older than” or “more recent than” a particular date, month or year. However, a single date, month or year does not apply for many datasets, for example, railroad map valuation footprints: https://gisdata.mn.gov/dataset/trans-rr-valuation-map-footprint where the date ranges from 1918 to 1974. Filling in “1918” would be misleading as would filling in “1974”. That information is perfectly suited, however, for the Currentness Reference field below.</p>
Currentness Reference	Mandatory if applicable	This free text field allows publishers to provide date information to the level of detail that they know it, including

DISCUSSION DRAFT

		data collected during more than one time period. If a single date in <i>Time Period of Content Date</i> covers everything, though, then there's no need to repeat it in this element.
Progress	Mandatory	Fixed domain of three choices (planned; in work; complete) The element name "Progress" and the menu choices are from the FGDC metadata standard, so we can't change those. Publishers only need to choose from a picklist, so it's an easy element to fill out. Most of the time, data won't be published until it's complete; however, that's not always the case. For example, the LiDAR hillshade layer (https://gisdata.mn.gov/dataset/elev-lidar-hillshade) was published before the state was completely covered, so this element was "in work". That alerted users that the project wasn't done, so that if their area wasn't yet covered, they should check back later when it was done. When the hillshade layer was complete, this field was updated to say "complete" so people could quickly tell that the layer was now done.
Maintenance and Update Frequency	Mandatory	Fixed domain choices and open text options cover any situation, including "irregular" and "unknown".
Spatial Extent of Data	Desirable	Some indication of geographic area should be provided to help Commons users quickly see whether or not a dataset covers their area of interest. Geographic area is most commonly found in: <i>Title</i> ; <i>Spatial Extent of Data</i> ; <i>Bounding Coordinates</i> ; and <i>Place Keywords</i> . <i>Spatial Extent of Data</i> is a useful element since it can accommodate a text description of geography that doesn't fit well into keywords.
Bounding Coordinates (4 separate fields)	Mandatory	Required for the Commons in order to support displaying a map extent and using the map to perform geographic searches. Must be provided in decimal degrees (latitude/longitude). Use http://bboxfinder.com/ for help.
Place Keywords	Mandatory	Helpful to support the filter-by-tag function of the Commons.
Theme Keywords	Mandatory	Required to support the filter-by-tag function of the Commons.
Theme Keyword Thesaurus	Optional	ISO 19115 Topic Category could be a default entry since publishing in the Commons requires that the publisher select a main ISO category (although the metadata doesn't need to contain the ISO category). We're not aware of other thesauri, but there could be some topic-specific ones.
Access Constraints	Mandatory	Users need to know whether or not there are constraints.
Use Constraints	Mandatory	Users need to know whether or not there are constraints.
Contact person	Desirable	General comment about contact information: There are three places in MGMD for contact information: 1. Person who can answer questions about the data content; 2. Person who can answer questions about how the data is distributed; 3. Person who wrote the metadata and can update it.

DISCUSSION DRAFT

		<p>Sometimes these are all the same person, but often they are different people. Each one of these has the same array of separate elements (phone, address, email...).</p> <p>Specific comment: <i>Contact person</i> is not mandatory since some organizations opt not to list a specific person. They provide a generic organization phone and email, and sometimes a contact position, e.g., “customer service representative”.</p>
Contact organization	Mandatory	
Contact position	Desirable	See comments under “Contact person” above.
Contact address	Optional	Most users contact publishers by email or phone.
Contact city	Optional	
Contact state	Optional	
Contact postal code	Optional	
Contact voice phone	Mandatory	
Contact fax	Optional	
Contact email	Mandatory	
Browse graphic filename	Desirable	
Browse graphic description	Optional	
Associated datasets	Optional	

SECTION 2 – Data Quality		
Attribute accuracy	Mandatory	Either describe any quantitative or qualitative tests done or say that attribute accuracy has not been tested. This is more valuable to a user evaluating a resource than is leaving the field blank.
Logical consistency	Desirable if applicable	This element concerns topology (either findings from tests performed to assess topology problems or topology rules included with the data); it does not apply to all data formats.
Completeness	Mandatory	Either describe aspects of the data that are not complete or say that it is complete.
Horizontal positional accuracy	Mandatory	Either describe any quantitative or qualitative tests done or say that positional accuracy has not been tested.
Vertical positional accuracy	Mandatory if applicable	Either describe any quantitative or qualitative tests done or say that positional accuracy has not been tested.
Lineage	Mandatory	List or describe information sources used and main processing steps.
Source scale denominator		A single number does not work well for most digital data. This information can be included, if applicable, in the sources described in <i>Lineage</i> . Not in MME.

SECTION 3 – Data Organization		
		We are no longer using this section. None of the elements in Section 3 are in MME.
Native Dataset Environment		
Indirect Spatial Reference		
Spatial object type		
Vendor specific object types		

DISCUSSION DRAFT

Tiling scheme		Not an FGDC element.
---------------	--	----------------------

SECTION 4 – Spatial Reference		
		<p>Implementation of spatial reference information needs to be further investigated. The crosswalk between FGDC and MIMG elements is not always smooth; MME implementation can be complex and incomplete for some coordinate systems.</p> <p>Note: publishers should also provide a projection file with their data.</p>
Horizontal coordinate scheme	Mandatory	MME issues: Several choices are missing from MME (e.g., Albers), and the HTML created by MME is missing coordinate information when “Geographic” is chosen.
Ellipsoid		Understood as part of the <i>Horizontal Coordinate Scheme</i> and <i>Horizontal Datum</i> . Dropped from MME.
Horizontal datum	Mandatory	
Horizontal units	Mandatory	
Distance resolution		Virtually no one understands or uses this element. Dropped from MME.
Altitude datum	Mandatory if applicable	
Altitude units	Mandatory if applicable	
Depth datum	Mandatory if applicable	
Depth units	Mandatory if applicable	
Cell width	Mandatory if applicable	
Cell height	Mandatory if applicable	

SECTION 5 – Entities & Attributes		
		<p>If a dataset has entities and attributes, then it is mandatory that they be listed and explained. However, not all datasets have entities and attributes, most notably imagery. Hence, the field is mandatory if applicable.</p> <p>All field names must be included and defined; it is not sufficient to just list the field names since they are often cryptic and their meaning is often unclear. Since there are several ways to provide this information (the <i>Entity and Attribute Overview</i> element, the <i>Entity and Attribute Detailed Citation</i> element or more specific field name and domain elements), no specific way is required.</p>
Entity and attribute overview	Mandatory if applicable	See comments for this section above.

DISCUSSION DRAFT

Entity and attribute detailed citation	Mandatory if applicable	See comments for this section above.
--	-------------------------	--------------------------------------

SECTION 6 – Distribution		
Publisher	Mandatory	This is an organization name since publishers to the Commons are organizations. A person's name can go in the distribution contact.
Publication Date	Optional	Optional since it is unclear what this date refers to (when the dataset was first published? When it was most recently published?), so it could be misleading. The field provides many options which seem to cover almost any scenario: "unknown", "unpublished material", "periodically revised", a date, or other free text.
Distributor contact person	Desirable	This element is not mandatory since some organizations opt not to list a specific person. They provide a generic organization phone and email, and sometimes a contact position, e.g., "customer service representative".
Distributor contact organization	Mandatory	We need a distribution contact in case there are problems distributing a particular resource (e.g., the download file is missing; an FTP site is down...)
Distributor contact position	Desirable	See comments under "Distributor contact person" above.
Distributor contact address	Optional	Most users contact publishers by email or phone.
Distributor contact city	Optional	
Distributor contact state	Optional	
Distributor contact postal code	Optional	
Distributor contact voice phone	Mandatory	
Distributor contact fax	Optional	
Distributor contact email	Mandatory	
Distributor's dataset ID	Optional	For publisher's use
Distribution liability	Desirable	It is the publisher's responsibility to determine whether they should include a distribution liability statement and how it is worded.
Transfer format name		Dropped from MME. (The Commons resource page shows available format options.)
Transfer format version #		Dropped from MME.
Transfer size		Dropped from MME.
Ordering instructions	Desirable if applicable	This element is less needed for resources distributed on the Commons.
Online linkage	Mandatory	Either link back to the Geospatial Commons or to another appropriate place. Revisit automatic wording which currently reads, <i>"I AGREE to the notice in "Distribution Liability" above. Clicking to agree will either begin the download process, link to a service, or provide more instructions. See "Ordering Instructions" above for details."</i>

SECTION 7 – Metadata		
Metadata date	Mandatory	Helps us ensure that Commons metadata is kept updated.
Metadata contact person	Desirable	This element is not mandatory since some organizations opt

DISCUSSION DRAFT

		not to list a specific person. They provide a generic organization phone and email, and sometimes a contact position, e.g., “customer service representative”.
Metadata contact organization	Mandatory	We need to know who to follow up with if metadata on the Commons needs updating.
Metadata contact position	Desirable	See comments under “Metadata contact person” above.
Metadata contact address	Optional	Most users contact publishers by email or phone.
Metadata contact city	Optional	
Metadata contact state	Optional	
Metadata contact postal code	Optional	
Metadata contact voice phone	Mandatory	
Metadata contact fax	Optional	
Metadata contact email	Mandatory	
Metadata standard name	Optional	
Metadata standard version	Optional	
Metadata standard online linkage	Optional	

Implementation

This recommendation does not specify how to enforce requirements. Not all requirements need to be enforced in the same way or with the same consequences. Some possible ways include:

- Provide updated education and training materials on metadata best practices.
- Work with new publishers to review their initial records (one-on-one assistance); do not publish these initial records until they meet requirements.
- Validate anything that can be automatically tested:
 - During metadata creation
 - During publication process to GDRS and Commons

Note that many elements are difficult or impossible to automatically test, for example, “Mandatory if applicable” elements. Automatic tests can ensure that elements have text in them, but not that elements such as Abstract or Lineage are well-written or complete.

- Monitor user feedback on the Commons to ensure that publishers respond to the feedback and improve metadata if needed.
- Periodically remind publishers to check their metadata for currency.
- If resources are available, improve metadata tools.
- Last resort: If education, training, feedback and persuasion don’t work, pursue governance options to warn publishers and then suspend them from publishing until they comply with requirements.