USING TEMPLATE TO IMPROVE THE EFFICIENCY OF IMAGERY METADATA ENTRY

Yao-Hsien Yeh[1] Jung-Hong Hong[2]
[1]Master student, Department of Geomatics, National Cheng Kung University, No.1, University Rd., Tainan 70101, Taiwan; Tel: +886-6-2370876#851; Fax:+886-6-2375764
E-mail: yaw770602@msn.com

[2] Associate professor, Department of Geomatics, National Cheng Kung University, No.1, University Rd., Tainan 70101, Taiwan; Tel: +886-6-2370876#851; Fax:+886-6-2375764

KEY WORDS: metadata, template, aerial photo

ABSTRACT: As the internet-based geospatial technology rapidly develops, a tremendous volume of geospatial resources are now available on the web. How to enable users to easily access and use these continuously growing resources has become a critical challenge. Although standardized geospatial metadata is widely considered as an effective method for facilitating the discovery of distributed georesources, the correct generation of metadata has long been a difficult task to the geospatial community. This research provides a solution to improve the efficiency of metadata generation via metadata template. Following the standards of ISO19115 and ISO19115-2, the metadata template for aerial photo produced by the Aerial Survey Office is analyzed and generated. The results shows that 90% of the metadata elements of aerial photos produced in a single project can be included in the template. Meanwhile, the majority of metadata elements not included in the template can be automatically collected from the onboard data or production procedures. This template approach shows advantages for reducing the loading of metadata entry and ensuring the quality of metadata. For organizations that routinely produce geospatial datasets following particular specifications, the advantages of using templates are even more obvious.

1. INTRODUCTION

With the rapid progress of Spatial Data Infrastructure (SDI), people began to recognize the importance of data sharing. Metadata is a key element for the implementation of SDI (Nebert, 2004). Geospatial metadata offers the information about the content, coverage, quality and other characteristics of the distributed geospatial data. The various metadata standards used by different countries or organizations, however, have been impeding the cataloguing of metadata and the development of software. To remove such an interoperability barrier, ISO/TC211 published ISO19115 standard as a common reference framework and allows organizations to design their own profiles as long as they are compliant to the ISO standard. As remotely sensed data rapidly develops as an essential resource to the GIS community, the imagery metadata should also receive its deserved attention. A comprehensive comparison about imagery metadata standards can found in Di(2003). To also provide a common metadata framework for imagery and grid data, ISO/TC211 extended the original framework of ISO19115 and officially published ISO19115-2 standard in 2009. Even in its developing stage, there have been efforts for establishing metadata for remote sensing data following ISO standards, e.g., the satellite-observation data from CEOP (Xie et al. (2007). Nowadays, ISO metadata standards are widely adopted by many countries for developing their NSDI. Meanwhile, the software support for standardized metadata has been also quickly growing.

Although the consensus of adopting international metadata standards has been reached, the creation of metadata unfortunately remains to be a world-wide challenge to geospatial community. Such a task would require both the knowledge about geospatial data and metadata. ISO19115 standard contains hundreds of metadata elements organized in a hierarchical structure, some of the elements are repeatedly recorded if necessary. Even a geospatial professional may view the entry of metadata as a tedious and labor intensive work, let alone inexperienced metadata editors. From past experiences, trying to teach everyone the skills for metadata entry is cumbersome and ineffective, so an approach that can reduce the entry cost and ensure the quality of metadata is desperately needed. Manso et al. (2004) suggested that some metadata can be collected from existing data files. Batcheller (2007) proposed a semi-automatic metadata generation approach by broadly including more data sources. Miguel etc. (2007) extended this automatic metadata collection concept to the SDI and demonstrated its usage in cataloguing of internet resources. These works mainly focused on automatically collecting metadata that already exists or hides in current data. Though the loading is reduced, it is not guaranteed all the required metadata can be created. By introducing domain expertise, we will be exploring the efficiency improvement by using domain-specific metadata template in this paper. The following discussion will respectively address the concept of template and how template is created. Aerial photo is used in this paper as an example to demonstrate the impact of using metadata template.

2. METADATA STANDARDS AND TEMPLATE

2.1 ISO Standards

To facilitate a solid and standardized framework for the development of interoperable GIS, ISO/TC211 has developed the so-called ISO19100 series of standards. The ISO19100 series of standards now includes more than 40 standards or technical specifications specifically designed for the standardization of chosen GIS topics. ISO metadata standards are not only used for generating standardized metadata, they are also used to support the definition or implementation of other standards. For generating metadata for remote sensing images, both ISO19115 and ISO19115-2 standards must be considered:

ISO19115 serves as a consensus and general-purpose metadata framework for geospatial data. The standard is composed of 12 packages, each denotes a specific type of property about geospatial data, as shown in figure 1. As a common reference, ISO19115 already includes more than 400 elements, but extension from the original schema is still possible.

ISO19115-2 extends the ISO19115 by additionally defining elements for imagery and grid data. The newly introduced elements can be used to describe the various aspects of an acquisition project for producing RS images. Figure 2 shows the class of MI_AcquisitionInformation and its aggregated set of classes, every class is designed to describe a specific property of imagery or grid data.

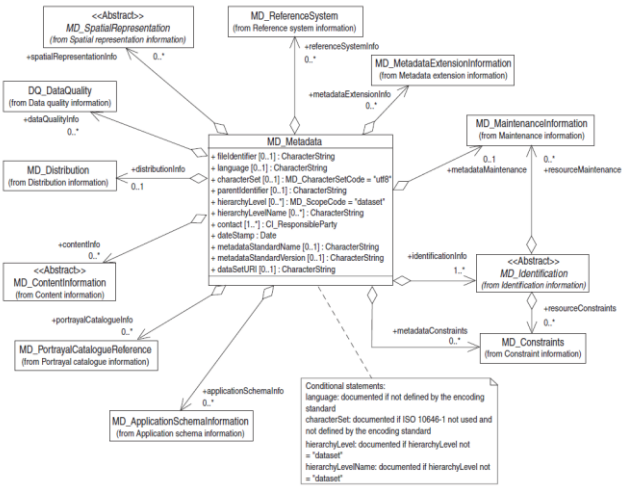

Figure1. The overall structure of ISO19115

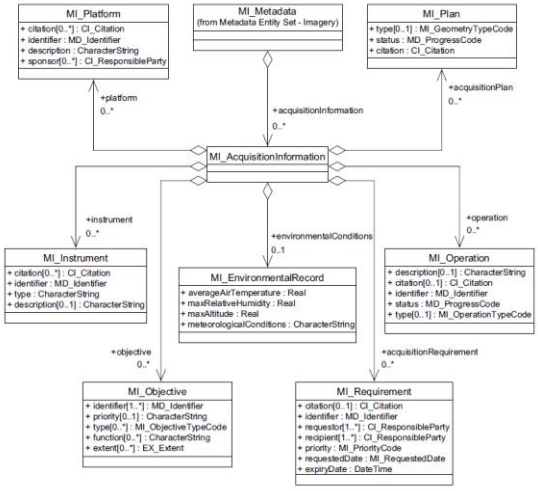

Figure2. The structure of acquisition class in the ISO19115-2

2.2 Metadata Template

The tedious metadata entry loading mainly comes from the huge number of metadata elements and the difficulty for determining the most appropriate content for each element. As the same type of geospatial datasets may share similar properties, the basic idea of metadata template is to first analyze these common characteristics and determine their corresponding metadata elements, then ask domain experts to fill in contents for these elements. After such a template is imported, the editors only need to focus on those elements that are not included in the template. By avoiding repeated entry of the same metadata contents, the efficiency of the whole process is effectively improved. The ideal subject for using metadata template is a collection of datasets that share common characteristics. The more homogeneous this collection of dataset is, the more elements the template can cover. Another advantage is domain-specific templates created by domain experts can improve both the quality and reliability of metadata. Because a series of imagery data is often produced by the same organization following the same specifications, the use of template should be able to effectively reduce the loading of metadata entry.

3. 2-STEP METADATA TEMPLATE GENERATION

Because the content of metadata template is totally dependent on the chosen collection of datasets, template design requires the designer to simultaneously have in-depth knowledge about metadata and domain data. Without a thorough understanding about the metadata standards, it may take a long time for domain experts to complete the design of a template. As its UML diagrams suggest, the classes and metadata elements in ISO19115 has its design logic. We argued the template design can take full advantages of this logic and largely reduce the difficulty for template design.

3.1 Step1: Exclude inapplicable metadata elements.

As a general-purpose framework, some metadata elements from ISO19115 may be not applicable to a specific type of data and absolutely do not need to be included in its metadata entry process. Following the designing logic of ISO19115, this step helps users to identify and exclude inapplicable metadata elements. Users are guided to answer a list of questions about the property of datasets. If the dataset does not have a particular property, the pre-determined corresponding group of metadata elements will be excluded. For example, the elements used to describe vector data are not applicable to raster. With this Q/A design, designers are not required to deal with individual elements and inapplicable elements can be excluded by logically grouping related elements. Altogether 30 questions were formulated from ISO19115 and ISO19115-2, only portion of them are listed in Table 1 due to page limits.

Table1. Enumeration of the applicable questions.

Question Type	Class	Question
Yes/No	MD_Identification	Does the dataset have a scale? Is the unit of coordinate meaningful?
		Does the dataset include time extent information?
		Does the dataset include vertical extent information?
		Does the dataset have a browsing diagram?
		Is the dataset specifically used for some applications?
		Is the dataset aggregated by other datasets?
	DQ_Ouality	Is the information of data processing history available?
		Has the dataset evaluated by quantitative-based measures?
	MD_Spatial Representaiton	Does the dataset include parameters for georectification?
		Does the dataset include information of control points?
	MD_Content	Is the dataset created following specific feature catalogue?
		Does the dataset offer information about exposure condition or image quality?
		Is calibrated information of radiometric, camera and lens distortion available ?
		Is band information available?
		Is the dataset the product of a specified remote-sensing image acquisition plan?

3.2 Step2: Determining elements in metadata template

After step 1, the remaining elements are further analyzed to isolate elements that should be included in the template. Similar to step 1, we again design a list of questions for designers to answer. Their answers determine whether a particular group of metadata elements are added into the template or not. After this logical correspondence relationship is established, the loading of template design can be largely reduced. Table2 lists the questions and its corresponding classes or metadata elements. Some of the questions are from ISO19115-2 and are only meaningful to the template design for imagery data.

Table2. The template applicable questions.

No.	Question	Class	
1	Is metadata created following the same standard?	MD_Metadata	metadataStandardName, version, characterSet, language
2	Is metadata managed by the same organization?		contact
3	Are all datasets belong to the same database?		parentIdentifier, hierarchyLevelName
4	Are the datasets the products from the same project?	MD_Identification	abstract, purpose, status, language, characterSet, topicCategory, credit, supplementalInformation, environmentDescription
5	Are the datasets managed by the same organization?		pointOfcontact
6	Are the spatial representations of all datasets the same?		spatrialRepresetationType
7	Are the resolution of all datasets the same?		spatialResolution
8	Can the same keywords be applied to all datasets?		MD_Keyword
9	Are the intended applications of the datasets the same?		MD_Usage
10	Is there common access constraint for all the datasets?	MD_Constraint	useLimitation, accessConstraints
11	Is there common data use constraint for all the datasets?		useConstraints, othConstraints
12	Are the level of security of all datasets the same?		MD_SecurityConstraints

13	Is the production procedure following the same specification?	DQ_Ouality	LI_Lineage
14	Is the quality of all datasets evaluated according to the same specification?		nameOfMeasure, measureIdentification, measureDescription, evaluationMethodType, evaluationMethodDescription, evaluationProcedure
15	Is quality represented using the conformance method?		DQ_ConformanceResult
16	Is the grid definition of all datasets the same?	MD_Spatial Representaiton	numberOfDimensions, axisDimensionsProperties, cellGeometry、transformationParameterAvailability
17	Are all grid datasets referenced to the same collection of control points?		MI_GCPCollection
18	Are all datasets distributed in the same way (online/offline/both)?	MD_Distributio n	online→CI_OnlineResource offline→MD_Medium
19	Are all datasets referred to the same coordinate system?	MD_Reference System	MD_ReferenceSystem
20	Is the maintenance frequency for all datasets the same?	MD_Maintenan ce Information	maintenanceAndUpdateFrequency
21	Are all datasets created following the same feature catalogue?	MD_Content	MD_FeatureCatalogueDescription
22	Are the image process levels or image quality the same?		imageQualitycode, processingLevelCode
23	Are RS images produced by the same instruments and platform?	MD_Content, MI_Acquisition Information	MD)CoverageDescription, MD_RangeDimesion, MD_Band, MI_Band,MI_Platform, MI_Instrument
24	Are RS images the products from the same project?	MI_Acquisition Information	MI_Plan, MI_Operation, MI_Requirement, MI_Operation, MI_Objective, MI_Event,MI_PlatformPass
25	Are the datasets portrayed following the same specification?	MD_Portrayal CataloguReferen nce	MD_PortrayalCatalogue Reference

4. ANALYSIS RESULTS

4.1 Characterization of experimental data

The Arial Survey Office, Bureau of Forestry is the major government organization in Taiwan responsible for the routine and urgent production of aerial photos and ortho-images. Aerial photo is chosen as the target for template design in this paper. After reviewing its measurement specification and interviewing responsible personnel, typical properties of aerial photos from ASO are as follows:

- Aerial photos are produced, managed and maintained by specific parties in ASO.
- Aerial photos are produced according to pre-designed plan.
- Aerial photos are produced following rigorous operation specification to ensure their quality
- ASO follow its SOP to produce aerial photos and parameters in all processes have been collected and recorded.

The following analyzes the characteristics of aerial photos according to the structure of ISO19115 standards:

- **MD_Metadata:** Metadata is created following ISO standards by the same responsible party.
- **MD_Identificaiton:** Aerial photos are produced on a plan basis. ASO has a specific way for automatically naming each produced photo. Photos produced from the same projects are often managed in the same archive. The extent information and the browsing graphic for individual images are different and should be processed individually.
- **MD_Constraint:** The data access and use constraint of aerial photo follows the constrained order from the Ministry of Interior.
- **DQ_DataQuality:** The SOP of AOS ensures the overall quality of the produced aerial photo.

- **MD_SpatialRepresentation** and **MD_Content:** Aerial photos of the same plan are normally produced by the same technology and stored in fixed configuration of data formats.
- **MD_Distribution:** The distribution mechanism is operated by AOS.
- **MD_ReferenceSystem:** All aerial photos are referred to the same coordinate system.
- **MD_MaintenanceInformation:** Every aerial photo is regarded as a new version. Ortho-image is normally updated once a year.
- **MI_AcquisitionInformation:** The platform and instruments for photos of the same plan are normally the same .
- **MD_PortrayalCataloguReference** and **MD_ApplicationSchemaInformation:** These two classes are not applicable for the aerial photos.

4.2 Implementation of the template analysis

This section will detail the efficiency improvement by using domain-specific template. In the first step, elements related to vector data and class of MD_PortrayalCataloguReference, MD_ApplicationSchemaInformation and MD_AggrgationInformation are excluded. We thoroughly assess the remaining metadata elements and use the questions from the step 2 to generate the metadata template for aerial photos. Table 3 shows the overall analyzed result categorized by the questions in step 2. The column of "T No." denotes the number of elements that can added to template after answering questions. The column of "I No." denotes the number of elements that must be processed individually for each aerial photo. Question No.9, 21, 22 and 25 are not applicable to the aerial photo, so these questions will be skipped.

Table 3 The template design for aerial photo produced by AOS.

Main Class	Q No.	Applied to the Aerial photo(航空相片適用情況)	T No.	I No.
MD_Metadata	Individual item：fileIdentifier(1), datestamp(1)			2
	1	The creation of metadata is following ISO metadata standard.	4	
	2	Aerial photos are from the same plan and have the same parent identifier.	3	
	3	The responsible party for metadata is the Arial Survey Office	18	
MD_Identification	Individual item：Citation-title(1), EX_boundingPolygon(1), EX_TemporalExtent(2), MD_BrowseGraphic(3)			7
	4	Aerial photos are produced according to the same plan.	28	
	5	The responsible party for aerial photo is the Arial Survey Office.	18	
	6	The spatial representation type is raster.	1	
	7	The ground resolution for individual images is not always the same.		1
	8	Theme keywords for individual images can be regarded the same, but the spatial extent and temporal extent information for individual images are different.	6	6
MD_Constraint	10	The use and access constraints must follow the constrained order from the Ministry of the Interior.	2	
	11		2	
	12	Some photos are classified as secret photo.		3
LI_Lineage	13	The production of aerial photo follows the content of the plan. AOS has a SOP to produce aerial photos, only the time for photo acquisition may be different from one photo to another. Every process steps and resources in the SOP must be recorded. (The LI_Linaeage class is composed of LI_ProcessStep, LI_Source subclasses and "statement" item. The "dateTime item in the LI_ProcessStep is individual.)	1	
			20	1
			29	
DQ_Element	Individual item：datetime(1)			1
	14	The verification process follows the specification from the Ministry of the Interior	33	
	15	All distributed aerial photos must pass the verification criterion	22	
MD_Spatial Representaiton	Individual item：MD_Gerectified(7)			7
	16	Since aerial photos are produced according to the same plan, the definition of the grids is normally the same. The AOS uses the same datasets of ground control point. Unless requested, the AOS does not voluntarily provide information of control points.	9	
	17		15	
MD_Distribution	18	The AOS is responsible for the distribution mechanism, data format and ordering process.	33	

Main Class	Q No.	Applied to the Aerial photo(航空相片適用情況)	T No.	I No.
MD_Reference System	19	All photos are geo-referenced to the same coordinate system.(2-degree TM projected coordinate (TWD97))	13	
MD_Maintenance Information	20	Though not necessarily refers to the same spatial coverage, the same place is normally updated with new photos once a year..	21	
MD_Content	Individual item: illumination information(2), cloudcoverPercentage(1)			3
	Boolean type: calibrationinformation(2), trangulationIndicator(1)		3	
	23	The instruments used by Arial Survey Office are ADS40 and Zeiss Intergraph DMC. Instruments are always calibrated before operation.	70	
MI_Acquisition Information	24	The acquisition process is restricted by the aerial photo acquisition plan.	139	
Total number			490	31
Percentage			94%	6%

Table 3 indicates that altogether 521 metadata elements must be filled in for aerial photo. This number may be somewhat surprise and overwhelming to some people. We must stress that table 3 also indicates that 491 out of the 521 elements, 94% of the total elements, are included in the template. The even better advantage is these elements only need to be created once by domain experts. As for the 31 elements that need to be processed individually for every photo, most of them are elements about the identifier, date, extent and keywords of the aerial photo. As these types of information can normally be created automatically or acquired during photo processing, the loading of manual entry of metadata can be further reduced.

5. CONCLUSION

The increasing awareness of the SDI demands a complete and comprehensive approach to produce high quality metadata. A 2-step metadata template design approach is discussed in this paper. It guides domain experts to easily complete the design of metadata template by answering carefully designed questions. Since the fill-in content is determined by experts, this approach can improve both the quality of metadata and reduce the unnecessary loading of metadata entry. Using aerial photo as an example, we demonstrated only a small percentage of metadata elements must be processed individually and these elements can be automatically collected from routine production procedures. Many current practice for metadata creation focuses on the training to metadata editors. We instead suggest the metadata should be created using a top-down approach. After template is designed by domain experts, not only the required time and cost for processing the metadata for an individual dataset can be remarkably reduced, the quality of metadata can also be largely improved because of the domain expertise being introduced.

6. Reference

Batcheller, J. K, 2007, Automating geospatial metadata generation—An integrated data management and documentation approach, Computers & Geosciences, Vol. 34, pp. 387-398

Di., L., 2003, The Development of Remote-Sensing Related Standards at FGDC, OGC, and ISO TC211, IEEE International, vol.1, pp. 643-647

Diaz L., Martin C., Gould M., Granell C., Manso M.A., 2007, Semi-Automatic Metadata Extraction from Imagery and Cartographic Data. IGARRS'07, 3051-3052

ISO, 2003, ISO19115: Geographic Information-Metadata

ISO, 2009, ISO19115-2: Geographic Information-Metadata-part2: Extension for imagery and gridded data

Manso, M.A., Nogueras-Iso, J., Bernabé, M.A., Zarazaga-Soria, F.J., 2004, Automatic Metadata Extraction from Geographic Information, 7 AGILE Conference on Geographic Information Science, University of Crete, Greece, pp. 379-385.

Nebert, D. 2004, Developing Spatial Data Infrastructure: The SDI Cookbook, v.2, Global Spatial Data Infrastructure.

Xie, R., Shibasaki, R., Ono, M., 2007, Metadata Development for the Integration of CEOP Satellite-Observation Data, Journal of the Meteorological Society of Japan, Vol. 85A, pp. 487-517