# Analysis and Application of the Hierarchical Metadata Framework of the ISO 19115 Standard

Hsiung-Peng Liao
Dept. of Geomatics, National Cheng-Kung University, Ph.D. Candidate
Dept. of Geomatics National Cheng-Kung University #1, University Road, Tainan, Taiwan, China, Taipei
darkness@mail.geomatics.ncku.edu.tw

Jung-Hong Hong
Dept. of Geomatics, National Cheng-Kung University, Associate Professor
Dept. of Geomatics National Cheng-Kung University #1, University Road, Tainan, Taiwan, China, Taipei
junghong@mail.ncku.edu.tw

**Abstract:** The emergence of WebGIS enables GIS users to acquire geographic data from others much easier than before. The rapid growth of OpenGIS technology not only boosts the internet-based exchange of geographic data, but also provides a solid framework for the interoperability of geographic data. It is, however, critical for users to have an unambiguous understanding about data before they can use it in the interested domains. This must depend on the correct modeling and interpretation of metadata and a world-wide metadata standard can certainly reduce the complexity during the data transfer. Being widely accepted by GIS communities, the ISO 19115 metadata standard provides a complete and flexible metadata structure to describe the different natures of geographic data. An interesting feature of the ISO 19115 standard is its hierarchical framework that can be used to describe different levels of geographic data, namely, dataset series-level, dataset-level, feature type-level, feature instance-level, attribute type-level, and attribute instance-level. Though most of the current national standard only concentrate on the dataset-level metadata, we would like to further explore the possible applications of this hierarchical metadata framework in this paper. The analysis shows that the hierarchical framework can effectively reduce the duplicates during metadata creating process, as well as improve the interoperability of geographic features after transferred to clients.

## 1. Introduction

One of the technological impacts for the development of the OpenGIS technology is the direct search, access and transfer of geographic data via internet becomes much easier than before. Meanwhile, it also becomes easier for users to process a variety of data acquired from organizations in their software environment, e.g., map overlay [2, 5]. The internet technology no doubt creates a closer interconnecting relationship among different spatial data and geoservice providers, often know as the National Spatial Data Infrastructure. With such convenience, users must, however, have a way to understand the content about data or services and thus determine its fitness for actually use the data to their interested domains. For example, though simultaneously displaying the contents of different geographic data sets in a map window is no longer a problem, to simply treat the contents as a "map" is a rash and risky decision from the cartographic perspective [9]. A better working environment capable distinguishing the common and differences among the characteristics of acquired geographic data is necessary for future GIS users [12]. After collecting a variety of data from different organizations, the compilation of a traditional paper map at least includes such consideration as scale (influence the degree of the details and position accuracy of a map), time (phenomena may refer to different time periods), coordinate reference system (it is necessary to transfer different coordinate reference system into the same system), the mapping specification (the size of the symbol or attribute) etc. The correct decision will depend on the correct interpretation of the metadata that comes with the spatial data. Metadata is data that is used to describe another data, commonly known as "the data about data." [3]. Nowadays, the majority of the existing metadata files is on the basis of dataset, meaning a metadata file must be created for every geographic data file. The data search mechanism of spatial warehouse and distributed environment also operate on the basis of dataset-level metadata, and the whole dataset is returned if its metadata qualifying the search constraints. The appearance of OpenGIS technology, particularly standards like WFS or simple feature specification for SQL, begins to expand the query and distribution from the whole dataset to individual feature. If only features that qualifying search constraints are returned, whether the metadata designed to describe dataset is correct and sufficient to describe these features becomes an interesting question. On the other hand, the possible duplication in the metadata content may also cause difficulties to the storage and manipulation of metadata files. Metadata administrators may have to repeatedly input the same metadata contents if the corresponding geographic data files sharing similar characteristics. Thru metadata, we can correctly represent and query the described spatial phenomena [4, 6] and further determine its fitness for use [10, 11], but an uncertain future about how to appropriately create and use metadata is right ahead of us.

To allow better search and management, distributed metadata has to be created according to specific metadata standards. Every metadata standard has its own designed set of metadata elements, organization structure, and applicable domains. During the past 20 years, there have been various metadata standards designed for different levels and research domains. Among them, nation-level metadata standards of course had the greatest and full-scale impacts on the distribution mechanism of the whole nation. Nevertheless, since every country has developed her own standard and corresponding software, a metadata-related mechanism cannot be easily transferred to other countries. Furthermore, it neither can execute across domains and countries who adopt different metadata standards. To achieve goals of global geographic data distribution, like Digital Earth or Global Spatial Data Infrastructure (GSDI), this impedance certainly becomes an obstacle, but a metadata standard whose framework is accepted worldwide probably provides the answer to this challenge. In the past two years, the ISO 19115 metadata standard [14], proposed by ISO (the International Standard Organization) Technical Committee 211, has been widely adopted as the basic framework for metadata standard design. Under such circumstances, metadata of different metadata standards can be processed with a single query as long as all of these metadata complying with the ISO19115 metadata standard. The ISO19115 metadata standard will surely bring a huge impact to the future interoperable GIS environment, it is therefore our intension to explore more of its possible applications in this paper, particularly the new design of hierarchical metadata framework [8]. Besides, the encoding of metadata (e.g., XML) [16, 18] and application framework design will also be discussed.

Section 2 briefly discusses the background of the ISO 19115 metadata standard. Section 3 continues to discuss the hierarchical metadata framework and analyze the recorded metadata elements at different levels. The analysis result serves as the basic framework for later metadata recording. In the fourth section, we concentrate on how to record metadata in XML format and further discuss about the determination of metadata elements at different levels. Finally, section 5 summarizes our major contributions and outlines possible future directions.

## 2. The ISO 19115 Metadata Standard

Beginning with early nineties, many countries like the USA, Japan, Australia / New Zealand designed their national metadata standard based on the characteristics of georesource. National metadata standards are designed in a way to describe georesource in the respective country. It is normal that certain differences may exist among these standards and it is difficult to interpret or use metadata elements following different standards due to possible different definitions. To improve geographic data interoperability, reduce the differences among metadata standards, and drive an international transfer standard, ISO/TC211 proposed ISO19115 metadata standard [14] to promote the transfer of geographic data around the world. In year, Japan announced their national metadata standard, JMP 2.0 [15], completely based on the ISO 19115 metadata standard. Though CSDGM [7] and ANZLIC [1] Core Metadata 2.0 were not based on the ISO 19115, both of them have a special program dealing with the transforamation between their national metadata standards and the ISO 19115.

The objectives of the ISO standards are to provide a standard framework for the service of spatial data in internet. The main goal of the ISO 19115 standard is to provide a standardized recorded schema for metadata of geographical data. Based on the conceptual schema, the ISO19115 metadata standard identifies a set of metadata elements and the relations between metadata elements in a organized way. Unified Modeling Language (UML) [17] static structure diagram is used to define the classes of geographic information. To improve the distribution of metadata, all metadata elements are encoded in XML format file.

Illustrated in Figure 1, the comprehensive set of 19115 metadata entity set information consists of 11 entities (UML classes) , with more than 300 either optional or mandatory elements (UML attributes). Besides the entity MD_Metadata, which is mandatory, all the aggregated entities (which are further expanded by some subclasses) are listed as follow:

1. MD_Identification
2. MD_Constraints
3. DQ_DataQuality
4. MD_MaintenanceInformation
5. MD_SpatialRepresentation
6. MD_ReferenceSystem

7. MD_ContentInformation
8. MD_PortrayalCatalogueReference
9. MD_Distribution
10. MD_MetadataExtensionInformation
11. MD_ApplicationSchemaInformation

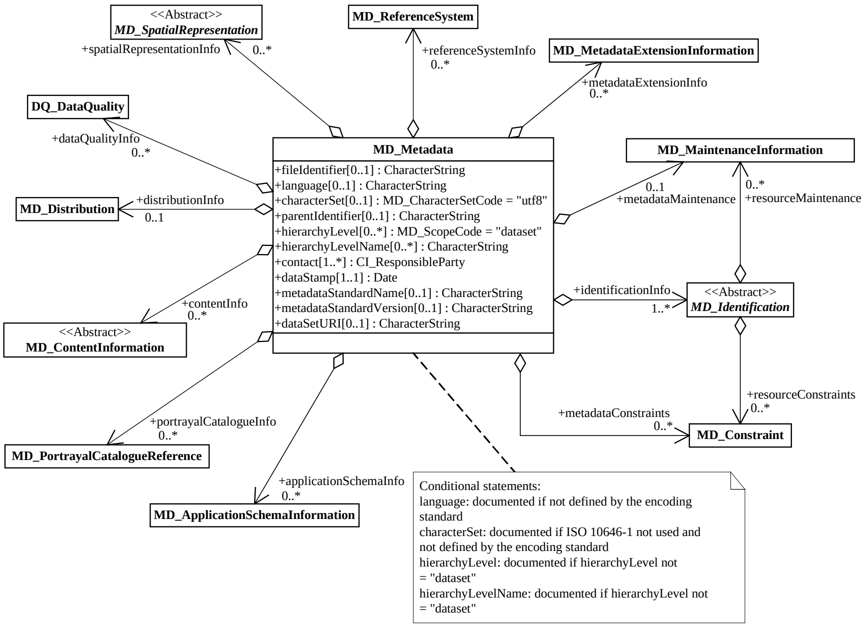


Fig. 1 Metadata Package [from 12 Figure A.1]

In this UML diagram, the core element of the whole metadata is the entity "MD_Metadata", it not only displays the entity's name, but also illustrates the metadata elements and data type of this entity. Except using UML diagrams to illustrate the relations between different entities, the ISO 19115 standard also provides a data dictionary with detailed specification about all the designed metadata elements (e.g., name, short name, definitions, obligation or condition, maximum occurrence, data type and domain). This data dictionary, in conjunction with all the UML diagrams, serves to fully define the total abstract model of metadata. When users try to create their own metadata file, it is convenient to refer this dictionary. According to their needs, users have flexibility to choose appropriate metadata elements and elements may be aggregated by one or more other elements following the predefined structure. Accommodated the XML (recommended, but is not mandatory) or other similar implementation techniques, we can encode metadata in the same international format. With this international format, it becomes easier for uses to manage or interpret metadata and determine its fitness for use.

## 3. Classification of Hierarchical Metadata

To deal with the unnecessary redundancy in the metadata, ISO 1115 proposed a flexible hierarchical metadata framework. The hierarchical framework divides the descriptions of resource into six levels: dataset series, dataset, feature type, feature instance, attribute type and attribute instance. ISO19115 metadata elements can be introduced to describe the appropriate levels of the geographic data nature. One of the challenges is of course the determination of the metadata element set at different levels. This framework works in a way similar to the inheritance relationship in object-oriented programming, such that an individual feature may also use the metadata of the dataset which the feature belongs to. A "Feature" and a "Attribute" level-based metadata are abstracted from feature and attribute, these metadata elements which are often recorded as layers in vector map do not record any phenomena in real world, but can only be regarded as an abstract description.

A “dataset series”, which usually corresponds to the “database” concept of the real data, is a collection of spatial datasets that share similar characteristics of theme, scale or purpose. Dataset series metadata entities are therefore used to describe the common characteristics of the data in the dataset series, for example, digital topographic maps in a dataset series may be produced with the same scale, by the same organization, and following the same mapping specifications. If needed, these metadata elements can all be considered as “dataset series” metadata elements.

A “dataset”, which usually corresponds to the “file” concept of the real data, is used to describe a set of spatial data. Example of dataset may include: a single image, a vector map, etc. A dataset may be composed of a set of identified “feature types”, “feature instances”, “attribute types” and “attribute instances”. A “feature”, which corresponds to the “object” concept of the real data, is used to describe some specific objects in the files or images, like bridge, road, building, parcel, etc. An “attribute” level of metadata which corresponded to the “attribute” concept in tables is used to describe the attributes of the table, for example: parcel number, length, area etc.

Figure 2 illustrates the relations between real data and hierarchical framework metadata. A database may be composed of a set of files and a file may contain a number of objects. For every individual geographic data file, we are supposed to crate a corresponding metadata file. If these geographic data files are in the same database, it is very likely these data files may share a high degree of similarity, e.g., same scale, same mapping specification, same purpose, etc. This is exactly the case between the levels of “dataset” and “dataset series.” Under such circumstances, if the metadata of the “dataset series level” can be found, theoretically its contents can be shared by all the “dataset” and only have to be created once. Furthermore, if the query is related to these metadata elements, we only deal with it and do not need to scan the metadata for all individual dataset one by one. There is a similar relationship between the metadata of “dataset” and “features.”

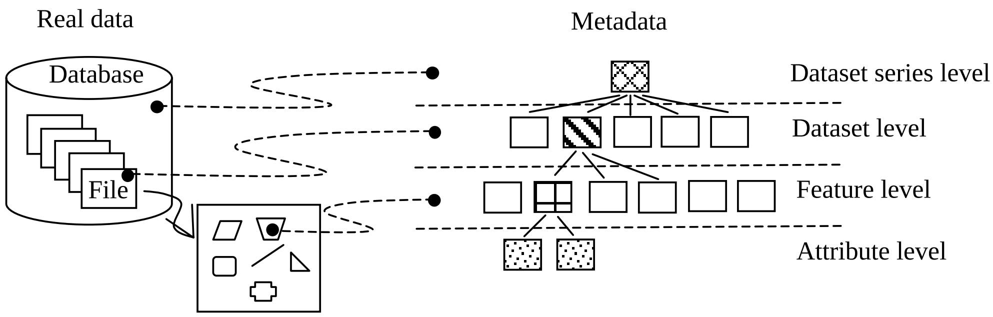


Fig. 2 The concept of hierarchical framework metadata

Up to now, the majority of metadata was created on the basis of individual geographic data file, meaning regardless the metadata elements should be referred to a specific level, they will be recorded and organized “dataset” metadata elements. However, such thinking may need to be adjusted after the service response gradually switched from dataset-based to feature-based. Unlike the whole data file downloading scenario in the past, OpenGIS standards allow geoservice to response only the features satisfying users’ needs.

Figure 3 illustrates the past dataset-based scenario where metadata is created to describe a single geographic data file, regardless the possible different levels. When users only acquire part of the features from databases (one or more databases), the whole set of metadata is acquired as well, rather than the metadata elements that describe those specific features. Suppose the users acquire three features from databases, every feature theoretically will be associated with a metadata file and their contents may be exactly the same. If another feature is acquired from another database, a different metadata file (may refer to a different metadata standard) is associated with the returned feature. With the hierarchical metadata framework, however, users can acquire the feature which satisfied their need and the feature level metadata corresponding to the features, and then the complete metadata file can be aggregated from other levels of metadata. This framework reduces multiple copies of the same metadata information, can efficiently eliminate the volume of the storage quota. In figure 4, users acquire four features (Feature 1, 2, 3 and 4) from the database, thus there are four parts of corresponding feature level metadata elements relevant to the four features, and two parts of dataset level metadata, nevertheless, there is only one parts of dataset series level metadata. To accomplish the complete metadata, what we have to do to is simply aggregating these three levels of metadata. In this complete metadata, not only record the whole database information, but also describe the special metadata elements against the features.

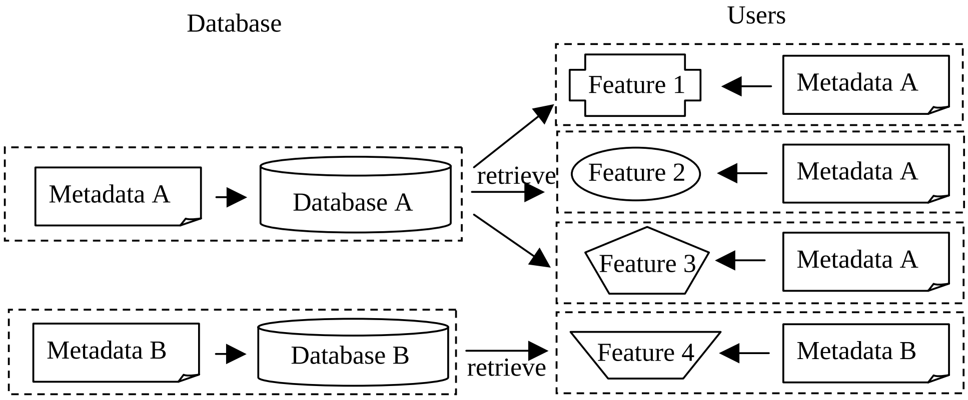


Fig. 3 Retrieve file-based metadata

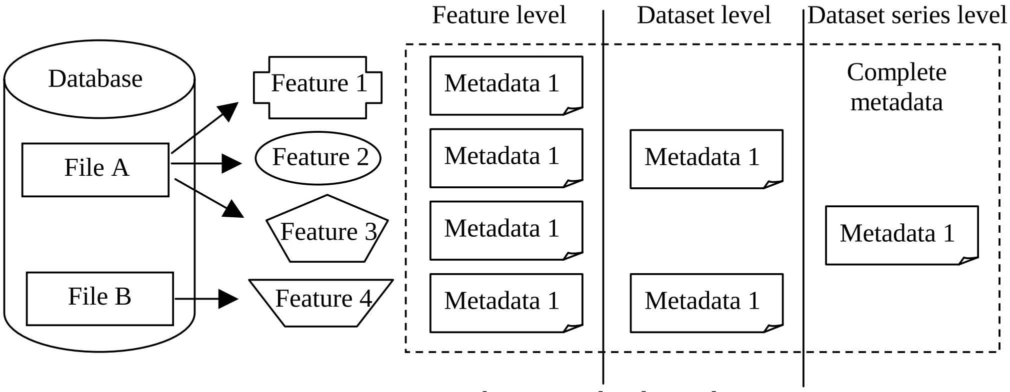


Fig. 4 Retrieve the Feature level metadata

## 4. Creation of Hierarchical Metadata

Theoretically, metadata elements should be classified to their corresponding levels, but the classification result will differ by the characteristics or types of the database actually. Take cadastral map for example, cadastral organization may manage more then one land sections, with a land section aggregated by one to many parcels. Meanwhile, surveyors of the same organization may follow the same mapping specification to produce their cadastral maps. Under such circumstance, it is very logical to classify the entity “data quality” into “dataset series” level. Therefore, in the dataset series level of metadata, we should only include the mapping specification in the data quality information.

However, if the evaluation subjects of data quality aim at the actual evaluating procedure, we have to consider if the change will affect other levels of metadata. And the data quality information of an individual parcel and its associate attributes may be updated and managed separately, e.g. the evaluation method type, description, procedure, date or time and evaluation result. Thus, it seems better to classify the entities “data quality” into “feature” or “attribute” level metadata. Based on the analysis of resource processing procedure, every metadata elements have to be classified into different level first. According to the time of data creation and maintenance, the analysis procedure may work in the following ways:

1. Creation phase:

In this procedure, data providers create their own resource and create the corresponding new metadata file at the same time, for example, surveying of 1/1000 digital topographic map. It is recommended to adopt the top-down approach. First of all, data providers decide which content of metadata elements are the same level, and classify those entities into the “dataset series” level. Remaining examine the lower level (dataset -> feature->attribute) of metadata, we can classify everyone entities of metadata into specific levels. This process continues until all elements of metadata are classified.

2. Maintenance phase:
   In this stage, all data including metadata have been created already, thus, the regular tasks are maintenance and update, like revising the cadastral map. It is recommended to adopt the bottom-up approach. In the beginning, data maintainer has to clarify the revised feature belongs to which level of metadata, for instance the whole file or only one feature. After exam the comparable level of metadata, we continue to check if the corresponding level of metadata has to be changed from the lower to higher level of metadata elements.

Next, we will take cadastral map as an example. In Taiwan, there are two kinds of digital cadastral maps: one kind is analog cadastral maps and the other kind is resurveying maps. The analog cadastral maps were digitized, and the resurveying maps were surveyed by total station or other digital surveying instruments. An organization may simultaneously manage the two types of maps. As time goes by, the maps of parcels and its associated attributes may be changed by surveyors, thus, the content of metadata have to be changed coincidently.

In the following analysis, we first make a preliminary analysis regarding the corresponding metadata elements at different levels for cadastral maps. Since each cadastral organization manages its map database that contains both types of cadastral maps, all datasets in the database, regardless what type it is, are managed by the same organizations. Metadata elements like “point of contact” therefore should belong to “dataset series” level. In the mean time, the file management of cadastral data is often based on individual land section. All common characteristics of a certain type of cadastral maps, e.g., scale, updated time, spatial coverage, etc., should belong to the level of “dataset.” Furthermore, a land section file contains all of the parcels within the land section, metadata elements specifically applied to individual parcel are classified as “feature level”, for example, parcel-based metadata. Beside this analysis, Hong [13] proposed a common framework for geographic data standard in Taiwan. By introducing ISO/TC211 standards, a fundamental class that includes five types of description information, namely, identification, coordinate reference system, temporal, description, and management, is developed. This class contains required metadata elements that are applicable to all the domains of geographic data, such that these attributes can all be regarded as “data series” level following our concepts. Table 1 lists the corresponding metadata class and elements regarding to the five characteristics:

Table 1. “Dataset series” level metadata

| | Metadata class | Metadata element |
|---|---|---|
| Identification | MD_Metadata | fileIdentifier |
| Coordinate reference system | MD_ReferenceSystem | referenceSystemIdentifier |
| Temporal | MD_Metadata | dataStamp |
| Description | MD_DataIdentification | abstract |
| Management | MD_DataIdentification | pointOfContact |

In general, it is sufficiently to collect all the metadata information in dataset series level. However, if one specific land section of cadastre in the organization had been changed, like resurveying or digitized, we have to recollect the dataset level metadata regarding to the changed land section advanced. A “dataset” level metadata is suggested to contain coordinate reference system, data quality, maintenance information and spatial representation information etc. When the owners of different parcels were transferred at different time, it is necessary for us to add additional information, like temporal, in the “feature” level metadata. Depending on the characters and diversity of resource, indeed, we have to exam every metadata element entities belong to which levels situated.

Table 2 lists the preliminary analysis result of metadata entities of different levels, symbol “＊” indicates that corresponding metadata elements are applicable to the specific levels of metadata. Some metadata entities are applicable to all of the levels, while some entities are only applicable to certain levels. For example, because data creators may collect different level of metadata based on the different randomly check procedure or evaluation result. In this table, we can find out that the number of elements in dataset or feature is less than the number of elements in dataset series. It means, in this database, most data share common information of the resource. As a result, when data providers are trying to implement the metadata for each resource, most data can inherit from the dataset series level metadata. The more element entities can derive from the dataset series level metadata, the fewer entities need to be created for other levels of metadata. This can effectively eliminate the difficulty and reduce the time in creating metadata.

Table 2. Different metadata entities corresponding to different level - Cadastre

| | Dataset series level | Dataset level | Feature level |
|---|---|---|---|
| Identification information | * | * | * |
| Constraint information | * | * | |
| Data quality | * | * | * |
| Maintenance information | | * | * |
| Spatial representation information | | * | |
| Reference System | * | * | |
| Content information | | * | |
| Portrayal catalogue information | * | | |
| Distribution information | * | * | |
| Metadata extension information | * | | |
| Application schema information | * | | |

Finally, the designed metadata schema is encoded in XML format. All of the encoding should comply with the related XML schema standard and the data dictionary of the appendix B in the ISO 19115 standard [14]. Consider data provider distributes the metadata and three land sections of different coordinate reference systems to users, and the metadata only describe the contact information and coordinate reference system. Note that all land sections share the same contact information, however, the reference systems of each land section may be different. As we defined above, land section can be regarded as dataset level. So, the metadata could be carried exclusively at dataset series level as XML format is list bellow:

```
Dataset series:
    <MD_DataIdentification>
            <pointOfContact>
                    <organisationName>East-South organization of Tainan</organizationName>
                    <contactInfo>
                            <phone>
                                    <voice>06-2680595</voice>
                            </phone>
                    </contactInfo>
                    <role>006</role>
            </pointOfContact>
</MD_DataIdentification>
```

And the metadata at dataset level should include reference system information, example as XML format is list bellow:

```
Dataset series:
    <MD_DataIdentification>
            <pointOfContact>
                    <organisationName>East-South organization of Tainan</organizationName>
                    <contactInfo>
                            <phone>
                                    <voice>06-2680595</voice>
```

```
                    </phone>
               </contactInfo>
               <role>006</role>
          </pointOfContact>
</MD_DataIdentification>

Dataset:
  <refSysInfo>
       <MdCoRefSys>
            <refSysId>
                 <identCode>TWD97 Coordinate Reference System</identCode>
            </refSysId>
            <projection>
                 <identCode>Transverse Mercator</identCode>
            </projection>
            <ellipsoid>
                 <identCode>GRS80</identCode>
            </ellipsoid>
            <datum>
                 <identCode>TWD97(Taiwan Datum 97)</identCode>
            </datum>
       </MdCoRefSys>
  </refSysInfo>
```


## 5. Conclusions

In this paper, we analyze and explore the hierarchical metadata framework of the ISO 19115 metadata standard. We first provided a preliminary analysis regarding the corresponding metadata elements different levels of metadata should or suggested to be included. The hierarchical framework can reduce the loading of feature-based metadata collection by inherit higher-level (dataset series or dataset) metadata, and eliminate the possible duplicate in metadata. Second, when managing feature or attribute metadata, users can acquire specific level of metadata to fit their need. Therefore, this can provide an efficient searching architecture based on hierarchical framework, especially when users are searching or deciding if some specific data dose fit their need. With the aid of hierarchical based metadata, users may only need to first check if the dataset series level (database) metadata contain the information they need, if the answer is no, it is not necessary for users to search all other level (files) of metadata in this database. Obviously, it is rarely to provide such benefit as file-based metadata framework, otherwise, users have to create index or search every files previously. Finally, talking about the date or time which the creation process was occurred, we proposed two models to implement metadata. Considering vector map, we also try to discuss which metadata elements should be included regarding to the different levels of metadata. We recommend using XML format to encode all metadata elements or entities, cause it is not only benefit for data sharing via internet, but also easily to define and apply additional metadata to better satisfy special user needs. When all the metadata were created comply with the ISO 19115 international standard, we believe that using resource will be an easy thing, and this hierarchical framework will be a basis for the design of future domain metadata standards.

## References


[1] ANZLIC (2001) ANZLIC Metadata Guideline, Australian New Zealand Land Information Council (http://www.anzlic.org.au/asdi/metaelem.html)

[2] Bishr, Y., 1998. “Overcoming the Semantic and Other Barriers to GIS Interoperability” International Journal Geographical Information Science, Vol.12, No. 4, pp. 299-314.

[3] Bretherton, F. P. and Singley, P. T., 1994. ”Metadata: A User’s View”, Proceedings of the Seventh International Working Conference on Scientific and Statistical Database Management, pp. 166-174.

[4] Chou, M. C., 1996, The Discussion About Metadata Design in GIS Database Based on Spatial Fundamental Framework, Master thesis of the Dept. of Surveying Engineering, National Cheng-Kung University, Taiwan, R.O.C..

[5] Devogele, T., C. Parent, Spaccapietra, S., 1998. ”On Spatial Database Integration” International Journal Geographical Information Science, Vol. 12, No. 4, pp.335-352

[6] Dong, S.Y., 1995, The Design of Multi Indexes of Database in GIS, Master thesis of the Dept. of Surveying Engineering, National Cheng-Kung University, Taiwan, R.O.C..
[7] FGDC, 2000. Contents Standards for Digital Geospatial Metadata (http://geology.usgs.gov/tools/metadata/standard/metadata.html)
[8] Hong, J.H., 1997, A Research about the Aggregation of searching and metadata ,Proceedings of the 16th Conference on Surveying and Application, pp. 827-836
[9] Hong, J.H., 1998a: More than Map Overlay – Integration of Geographic Data from the Perspective of Data. Proceedings of the 4th Symposium on Geographic Information System, Taipei (CD-ROM).
[10] Hong, J.H., 1998b, Searching and Browsing of multi Database in Geographic Information System, Journal of Surveying Engineering, Volume 40, Number1, Page 1~24
[11] Hong, J.H., Liao, H.P., 2001. Map Metaphor Consideration toward Overlaying Geographic Data: A Coverage Metadata Perspective, Asia GIS, Tokyo JAPAN, Page 4-5
[12] Hong, J.H., Liao, H.P., Lee, B.L., 2002: The Valid Spatial Coverage Analysis Mechanism for Overlaying Geographic Data from Different Sources. Journal of Surveying Engineering, Volume 44, Number4, Page 97~114
[13] Hong, J.H., 2004, The Research Result of “The Design of Data Standard for Data Sharing in National Information System” Authorized bye the Ministry of the Interior
[14] ISO 19115, 2002. Geographic information – Metadata
[15] JMP 2.0, 2003, http://zgate.gsi.go.jp/ch/jmp20/jmp20.html
[16] Lou, U.C. , 2000, Metadata series report (1), Information Network about XML Paper in Taiwan, (http://www.xml.org.tw/Library/20000805.asp)
[17] OMG, 2003. UML Resource Page，Object Management Group，http://www/omg.org/uml
[18] W3C , 1998. “Extensible Markup Language (XML) 1.0” (http://www.w3.org/TR/1998/REC-xml-19980210)