Basic Elements and Implication of Software Metadata in the Intelligent Geospatial Web

Kiwon Lee†
Dept. of Information System Engineering, Hansung University

Abstract: During over decades, metadata on spatial data have been developed, and they have been widely applied at the national and international metadata standards such as file structure, format, and data model. However, in the web 2.0 paradigm toward user participation and openness, sources and contents of geospatial products are also diversified, not being limited to well-organized and structured data sets or databases. Especially, software products in both open source software and commercially packaged software are considered into important resources in the geospatial domain. But there are no reports or studies regarding software metadata from the side of software engineering or information technology, till now. The motivation of this study is based on practical needs to build search engine in the intelligent geospatial web. Brief review on current metadata standards is presented, and necessity for software metadata is discussed as well as related works. Basic elements, initially considered, of software metadata are presented. This work is the first attempt for software metadata, although it just covers geospatial software products. Further practical works to meet industrial demands need to actual applications of software metadata.

Key Words: Geospatial Web, Metadata, Standards, Software.

1. Introduction

At present, new concepts or new computing paradigm in the web are emerged; in the geospatial domain as a kind of the mainstream in information technology communities, those have been adopted and adapted in the viewpoint of standardization and industrial needs. Furthermore, this trend is converged in the paradigm of web 2.0 and more, under the basic slogan of openness, sharing, and participation. While, in the other words, applications in web 2.0 emphasizes on interactivity by mean of social computing and networking. Technically, this aspect in the geospatial web or geo web meets web 2.0 computing patterns: service-oriented architecture, software as a service, mashup, rich user experience, structured information and so forth. Fig. 1 shows comparative summary of web paradigm.

The geospatial web or geoweb is a term that implies the merging of geographical information with the abstract information that currently dominates the Internet. As well, it is extended that the geospatial web has a profound impact on managing knowledge, structuring workflows within and across organizations,
and communicating with like-minded individuals in virtual communities, being characterized by the self synchronization of network addressing, time and location. It is conceived that the geospatial web is regarded as a visual medium and geospatial platform for data self-organization, discovery and use. The geo-browsers which revolutionize the production and consumption of media products are the important and enabling technologies for the geospatial web (Scharl and Tochtermann, 2007).

As the main subject in this study, metadata is important to aid in clarifying and finding the actual data or contents. An item of metadata may describe an individual datum, or content item, or a collection of data including multiple content items and hierarchical levels, such as a database schema. In data processing, metadata provides information about, or documentation of, other data managed within an application or environment. This commonly defines the structure or schema of the primary data. Often data providers will provide users access to a variety of metadata fields, which can be used individually or in combinations, and applied by different users to achieve different goals. In general views, metadata helps data developers, data users, and organizations. For data developers, it is known that metadata helps: avoidance duplication, sharing reliable information, publicizing efforts, and reducing workload. For data users, metadata gives advantages: facilitating understanding, focusing on key elements, and discovery inside and outside organizations. As well, organizations use metadata for these reasons: protecting investments, countering personnel changes, creating institutional memory, sharing data with other agencies, reducing costs, and limited potential liability.

However, in the web 2.0 paradigm toward user participation and openness, sources and contents of geospatial products are also diversified, not being limited to well-organized and structured data sets or databases. Especially, software products in both open source software and commercially packaged software
are considered into important resources in the geospatial domain. But there are not reports or studies regarding software metadata from the side of software engineering or information technology, till now. The motivation of this study is based on practical needs to build search engine in the intelligent geospatial web.

2. Overview of Metadata Standard

Metadata is made up of a number of elements which can be categorised into the different functions they support. A metadata standard will normally support a number of defined functions, and will specify elements which make these possible (Bretherton and Singley, 1994; Stephens, 2003; Moellering, 2005).

Metadata in information and technology area is classified to three views in Fig. 2: content, directory, and dictionary. Most metadata standards in geospatial domain are inclined to reflect three aspects. Content metadata is for characteristics of metadata of language, coding system, and reference. Directory metadata is identification of the dataset, as name, origin, and description of the extents of the datasets. Dictionary metadata is semantic, geometric, and temporal data definition, quality information, and description of parameter. A metadata standard may support some or all of the following functions. Descriptive metadata enables identification, location and retrieval of information resources by users, often including the use of controlled vocabularies for classification and indexing and links to related resources. Technical metadata describes the technical processes used to produce, or required to use a digital object. Administrative metadata is used to manage administrative aspects of the digital object such as intellectual property rights and acquisition. Preservation metadata, amongst other things, documents actions which have been undertaken to preserve a digital resource such as migrations and checksum calculations. Structural metadata is used to describe the structure of computer systems such as tables, columns and indexes. Guide metadata is used to help humans find specific items and is usually expressed as a set of keywords in a natural language. Often metadata is split into three categories: Internal metadata which is relevant for system administrators and external metadata which is relevant for end-users.

![Metadata Diagram](image_url)

Fig. 2. Category of metadata (Bretherton and Singley, 1994; Stephens, 2003; Moellering, 2005).
Technical metadata correspond to internal metadata, business metadata to external metadata. In additions, a third category named process metadata.

As the concept of geospatial data documentation is introduced to organizations, most efforts focus on the documentation of existing geospatial data resources. As a result, metadata are captured after the data development process has been complete (NOAA, 2000). Since the mid 1990s, lots of researches and standardization about metadata in the geospatial field have been carried out. At the initial stage of those works, the efficient data management and retrieval for spatial DBMS running on internet or client/server environment is the primary aim; as the core parts, geo-based data sets were the main target for metadata specification. Therefore, during over decade of metadata works concentrated on data, national and international metadata standards such as file structure, format, and data model are being developed and released for public uses.

As actual metadata standards with the those bases, the Dublin Core Metadata Element Set (ISO Standard 15836) in 1995 is a basic standard which can be easily understood and implemented and as such is one of the best known metadata standards. It was originally developed as a core set of elements for describing the content of web pages and enabling their search and retrieval. It consists of 15 elements which address the most basic descriptive, administrative and technical elements required to uniquely identify a digital resource.

ISO 19115 in 2003 with the basis of a metadata profile of 22 core metadata for geographic datasets, was developed by the geospatial community to address specific issues relating to both the description and the management of spatial data: resource title (M), resource reference date (M), resource responsible party (O), geographic location of the dataset (C), resource language (M), resource character set (C), resource topic category (M), spatial resolution of the resource (O), abstract describing the resource (M), distribution format (O), additional extent information for the resource (vertical and temporal) (O), spatial representation type (O),

Fig. 3. Metadata standard: Basic scheme (Kresse and Fadaie, 2004).
reference system (O), lineage (O), on-line access to the resource (O), metadata file identifier (O), metadata standard name (O), metadata standard version (O), metadata language (C), metadata character set (C), metadata point of contact (M), metadata date stamp (M). In notations above, M, C, and O represent mandatory, conditional, and optional, respectively. Fig. 3 shows that other standards are based on ISO 19115.

There are five types of data standards: data classification, data content, data symbology or presentation, data transfer, and data usability. Among them, data usability standards describe how to express the applicability or essence of a data set or data element and include data quality, assessment, accuracy, and reporting or documentation standards. The FGDC Content Standard for Digital Geospatial Metadata Standard (CSDMS) is an example of a data usability standard, which has ten kinds of contents on metadata, identification information, data quality information, spatial data organization information, spatial reference information, entity and attribute information, distribution information, metadata reference information, citation information, time period information, contact information.

CEN 287 is the technical committee of CEN concerned with geographic information to define, describe, and transfer geographic data and services among European governments (http://geostandards.geonovum.nl/index.php/1.3.2_CEN/TC_287).

3. Related Works for Software Metadata

In web 2.0 or web 3.0 toward better functionalities and performance, the geospatial web handling multi-styled resources and complex data sets is an important implementation case. Originally, data sources are composed of multiple types or formats: OGC-based standard services, geoweb service, and non geoweb service. But these are mostly limited to data-centric resources. Therefore, non data-centric resources are considered in this work. Software or software components are included in these new categories.

As the first related works, ISO standards are demonstrated with the example of INSPIRE (2008) in Fig. 4. ISO 19100 standards are categorized to
infrastructure standards such as 19103 for conceptual scheme language or 19107 for spatial scheme base, basic standards such as 19111 for spatial reference by coordinate, 19112 for spatial reference by geographic identifiers or 19113 for quality, imagery standards such as 19130 for sensors, catalogue standards such as 19135 for registration procedure, 19110 for feature catalogue, or 19126 for data dictionary, and implementation standards such as 19131 for data product, 19128 for web map server interface, 19139 for metadata implementation.

As the second related things to establish non data-centric metadata, Fig. 5 represents that spatial data system for enterprise applications requires many standards: hardware, software, communication network and web service, and data. Among them, data standards in hierarchical level are well established. In the case of software standards, standards regarding DBMS and operating system are needed, and DBMS dichotomizes application and development environment, and application needs query and interface. However, this scheme is not fully practical one, because all standards are not developed yet, on the contrary data standards.

While, Fig. 6 shows timeline for GIS open source software with OGC standards released time.

In information technology metadata, it is dealt with all physical data and knowledge from inside and outside an organization, including information about the physical data, technical and business processes, rules and constraints of the data, and structures of the data used by a corporation, beyond data-centric scope (David, 2006).

As for relational database metadata, each relational database system has its own mechanisms for storing metadata. Examples of relational-database metadata include: Tables of all tables in a database, their names, sizes and number of rows in each table. But it is known that not all databases implement it, even if they implement other aspects of the database query language standard.

While, data warehouse metadata helps to design to manage and store the data, which focuses on the usage of data to facilitate reporting and analysis, which data warehouse is a repository of electronically stored data. An essential component of a data

---

**Fig. 5.** Software standards in spatial database systems standards (Yeung and Hall, 2007).
A warehouse or business intelligence system is the metadata and tools to manage and retrieve metadata. As data and information increases metadata will probably be very important in relation to search for the relevant already existing information to be used in a specific decision process (David, 2000).

In the viewpoint of software engineering, metadata is casually used to describe the controlling data used in software architectures that are more abstract or configurable. Most executable file formats include what may be termed metadata that specifies certain, usually configurable, behavioral runtime characteristics. However, it is difficult if not impossible to precisely distinguish program of metadata from general aspects of stored-program computing architecture; if the machine reads it and acts upon it, it is a computational instruction, and the prefix “meta” has little significance. For these aspects, Object Management Group (OMG) has defined metadata format for representing entire existing applications for the purposes of software mining, software modernization and software assurance. OMG developed Knowledge Discovery Metamodel (KDM), which is a common language-independent intermediate representation that provides an integrated view of an entire enterprise application, including its behavior, data, and structure (OMG, 2008). The goal of KDM is to ensure interoperability between tools for maintenance, evolution, assessment and modernization, and it is defined as a metamodel that can be also viewed as an ontology for describing the key aspects of knowledge related to the various facets of enterprise software. For this purpose, KDM is designed to be a common intermediate representation for existing software systems and their operating environments, which defines common metadata required for deep semantic integration of application lifecycle management tools (Fig. 7). While, KDM uses an XMI (extensible Metadata Interchange) format between tools that work with existing software as well as an abstract interface (API). KDM supports
the ecosystem which means a growing open-standard based cohesive community of tool vendors, service providers, and commercial components. KDM represents entire enterprise software systems, not just code.

4. Proposed Scheme for Software Metadata

In wide extension of web 3.0 or intelligent web, an importance of geospatial contents is increasingly accentuated. But current main interests of geospatial web operated by web portal are web mapping of geo-based data including high-resolution images and 3D and open API for mashup application development.

Furthermore, numerous software have been developed in the geospatial communities, and the activity of open source project, being progressed mainly in OSGEO (Open Source Geospatial foundation), is regarded as one of important research trends. As well, open APIs provided by search engine such as Google or geospatial engine software companies are being used to web mapping mashup applications by general users and experts. This means that the primary achievement in the geo spatial area is extended, not limited to geo-based data sets.

Once some one who queries in order to get specific information accesses an intelligent geospatial web, an example of query literally is “How do I develop or implement an urban planning system with natural landscape modelling functions ?”. The answers in the current web portal are brief documents about books or papers related to subjects such as landscape, landscape model, landscape development, urban planning, urban development, and so on. And at the next step, he/she finds a geospatial web site providing a catalogue service to get data sets or metadata of those. If that person who obtains data sets has an experience dealt with special geospatial software or tools, there is a little problem. But if not, that person wants to get more information about software or tools in many ways.

In scenario above, web-accessible information about documents and data sets is already established in the metadata, but information regarding software is not typically structured. In other words, there are so many answer types.

Table 1 is initial items to be considered as software metadata proposed. Basic idea is followed by ISO 19115 metadata standard for spatial data, but most items are different from that. As the core mandatory elements, eleven items can be chosen: software name, keywords of software functionalities, reference date, software category, software version, software release type, development language, software license, point of contact, operating system, software I/O file formats. The remaining items are optional: reference system, lineage software, metadata file identifier, software responsible party, maintenance information, distribution Information, software size, documentation, supporting standards, software design diagram. In this proposed elements, KDM metamodel may be also added. But at the initial stage for software metadata, it is too developer-oriented, not for the public users. It can be taken into account at the detailed level for each item of initial model. Fig. 8 shows metadata processing and its procedure for software, demonstrated with those of spatial data.

In general views, metadata about data helps data developers, data users, and organizations. This shows almost same effect for software developers, software users, and organizations, as long as provided software metadata. Software metadata helps: avoidance duplication, sharing reliable information, publicizing efforts, and reducing workload. Software users take advantages of facilitating understanding, focusing on key elements, and discovery inside and outside
<table>
<thead>
<tr>
<th>Suggested metadata elements</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software name (M)</td>
<td>Software name or software development project name</td>
</tr>
<tr>
<td>Keywords of software functionalities (M)</td>
<td>Core functions for software users</td>
</tr>
<tr>
<td>Reference date (M)</td>
<td>Release date, Update date, Patch date ...</td>
</tr>
<tr>
<td>Software category (M)</td>
<td>Application areas, Development usability</td>
</tr>
<tr>
<td>Software version (M)</td>
<td>Version information, Patch number</td>
</tr>
<tr>
<td>Software release type (M)</td>
<td>Open source code, Shareware, API-level, Commercial program ...</td>
</tr>
<tr>
<td>Development language (M)</td>
<td>C, C++, Java, C#, Python, ...</td>
</tr>
<tr>
<td>Software license (M)</td>
<td>Copyright, ...</td>
</tr>
<tr>
<td>Point of contact (M)</td>
<td>Name of contact person</td>
</tr>
<tr>
<td>Operating system (M)</td>
<td>Any versions of windows, Linux ...</td>
</tr>
<tr>
<td>Software I/O file formats (M)</td>
<td>Multiple exchange formats</td>
</tr>
<tr>
<td>Reference system (O)</td>
<td>Supporting geo-reference functions</td>
</tr>
<tr>
<td>Lineage software (O)</td>
<td>Linked library, Required software ...</td>
</tr>
<tr>
<td>Metadata file identifier (O)</td>
<td>Metadata itself</td>
</tr>
<tr>
<td>Software responsible party (O)</td>
<td>Individual programmer, Project group, Company ...</td>
</tr>
<tr>
<td>Maintenance information (O)</td>
<td>If necessary</td>
</tr>
<tr>
<td>Distribution information (O)</td>
<td>If necessary</td>
</tr>
<tr>
<td>Software size (O)</td>
<td>Number of Source code lines, Executable size in MB</td>
</tr>
<tr>
<td>Documentation (O)</td>
<td>Quick guidebook, Cookbook, ...</td>
</tr>
<tr>
<td>Supporting standards (O)</td>
<td>ISO-TC, OGC ...</td>
</tr>
<tr>
<td>Software design diagram (O)</td>
<td>UML, Other software engineering design methodology</td>
</tr>
</tbody>
</table>

Fig. 8. Geospatial software metadata processing and procedure, mainly dealing with existing geospatial open sources, along with spatial data.
organizations. As well, organizations use metadata so as to protect investments, counter personnel changes, create institutional memory, and reduce costs. As well as these general advantages, software metadata could give several benefits with regards to its applications.

For contents providers and users in the geospatial web, information about software description and its applicability data metadata are built and accessed in their web services, respectively. This information will be used as guidance or reference sites through intelligent web search. For open source project team or developer’s group, software metadata is a standard template which describes their works and products. For demander of software standards, it can be useful as initial requirement for the system architecture design, especially considering both open sources and multi-vendor products.

5. Concluding Remarks

Web 2.0 paradigm is towards user participation and openness, and the next stage of web 2.0 is expected as intelligent web under cloud computing environment. The geospatial web is also regarded as one of important trends in web 2.0 and more. Still, main contents in the geospatial web are data sets, and an importance of metadata concerned has been emphasized. However, sources and contents of geospatial products are also being diversified, not being limited to well-organized and structured data sets or databases. Open source software is one of core components in web 2.0 paradigm, but publishing or managing for it is negligible, compared to those of geo-based data sets. Furthermore, open source is closely related to standardizations within geospatial domain, such as OGC, ISO-TC, and other organizations or communities. Also in the software engineering, OMG is an important group, and KDM is crucial one for software development. But there is no approach for software metadata or metadata for software, till now.

From the previously learned experiences about metadata, several aspects point out. First, necessary metadata is more varied and complex than first appears, so that producers of data or content and software products require metadata model to be extendable to accommodate specific operational parameters. Secondly, it is that metadata is both population and validity. In other words, metadata must be reliably populated to be useful. And it is difficult to correct metadata after it is entered into inventory.

This work is the first attempt for software metadata, although covering geospatial software products. Basic idea in the proposed software metadata is followed by ISO 19115 metadata standard for spatial data, but most items are different from that. As the core mandatory elements, eleven items can be chosen: software name, keywords of software functionalities, reference date, software category, software version, software release type, development language, software license, point of contact, operating system, software I/O file formats. The remaining items are optional: reference system, lineage software, metadata file identifier, software responsible party, maintenance information, distribution information, software size, documentation, supporting standards, software design diagram. Of course, further concrete and deliberate specifications for software metadata are needed in the viewpoints from software engineering and geospatial communities.

Finally, software metadata will provide useful information for both software developers in open or commercial purposes and consumers in general or specific uses under the intelligent geospatial web.
Acknowledgements

This research was financially supported by Hansung University in the year of 2009.

References


