The state-of-the-art digital geologic mapping in Korea towards UML-based data modeling approach

Kiwon Lee[1)], No-Wook Park[2)], and Kwang-Hoon Chi[2)]

1) Department of Information System, Information and Computer Science Division, Hansung University, 389 Samsun-dong 2-ga, Sungbuk-gu, Seoul 136-792, Korea
e-mail : kilee@hansung.ac.kr

2) National Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources, 30 Kajung-dong, Yusung-ku, Daejon 305-350, Korea
e-mail : {nwpark, khchi}@rock25t.kigam.re.kr

Abstract

In Korea, national digital geologic mapping has been carried out since the late 1990s. This approach is a part of so-called NGIS (National GIS) of Korea, and geologically mappable database information is one of nation-covering thematic mapping. Several further researches after the first year project for Korean digital geologic mapping will be followed. Especially, to provide basic information from a variety of sources in a consistent, applicable, and searchable structure, digital data modeling approach is needed. Among geo-data modeling approach, the data modeling in UML can provide: sufficient semantics and notation to address a wide variety of contemporary modeling issues, extensibility mechanisms for the future modeling approaches. In this paper, first, we introduce general review of the state-of-the-art Korea national digital geologic mapping and concepts of geo-data model with data modeling aspects are addressed, and then case histories for geo-data modeling and several approaches for data modeling in GIS application fields are discussed. Lastly, tentative conceptual geologic data modeling by using UML(Unified Modeling Language) of OO(Object-oriented) concepts is attempted. Through this approach, the main benefits for standardization and implementation lineage with conceptual model in consideration to reusability are expected.

Introduction

A digital geologic map can be defined as any geologic map whose geographic details and explanatory data are recorded in a digital format that is readable by computer, with fundamentally different conceptual perspectives: cartography for digital mapping and analysis for geo-data processing. Generally, the purpose of a data modeling for digital geologic maps is to provide a structure for the organization, storage, and use of geologic map data in a computer

(Bain and Giles, 1997; Johnson, et al., 1998). In other words, the geologic data model formally defines the grammar of the digital geologic maps. A data model is formally defined as a set of fundamental conceptual objects and mathematical and logical rules that govern their behavior. The rules are usually expressed in terms of how and why objects may exist, and what interactions are permitted. Therefore, the data modeling is a complex task that attempts to capture the intricacies of real-world situations, including the characteristics of real-world objects, events, and object-event interrelations. Thus, the modeling process occurs at many levels of abstraction.

Geologic maps have, as a fundamental characteristic, line and polygon attributes that are interrelated. Thus, a fault may separate two polygons and continue internally into a third polygon. Such lines need to be included with the polygon data in order to do structural analyses, for example, to select individual polygons on the upper plate of a thrust. At practical implementation within GIS, this requires that dangling lines are included in the topological definition of the polygon coverage. Mechanisms are needed to document the source of each individual geologic object. For example, the source would include the full bibliographic reference for the object.

In order to represent geological data for use in spatial information systems a description of the data types, especially geometric data types underlying the description of spatial structure, must be made. If a single data model is to be used across a wide variety of application areas and software systems, it must be capable of supporting the representation of the data and relationships between data items required by those applications(Burrough; 1992; Frank, 1992).

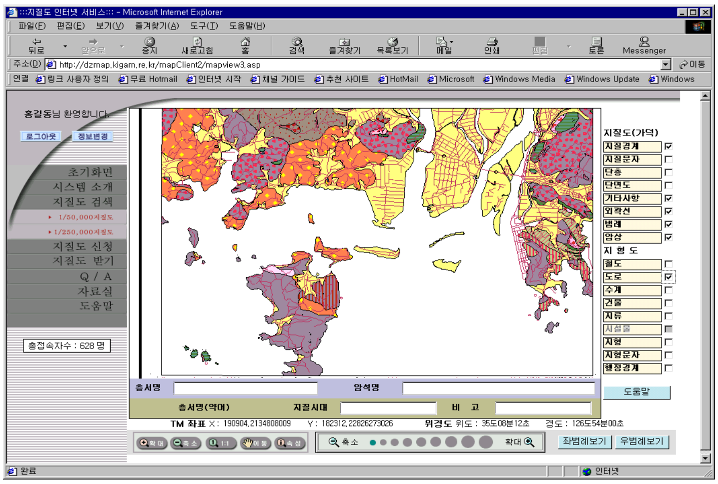
As for general or specific data modeling, numerous researches or open reports have been performed (Laxton and Beckon, 1996; Allen, 1997; Bain and Giles, 1997; Giles, et al., 1997). However, currently data models at geo-science field mainly aims at geologic database building in nation-wide; therefore, geo-data model in consideration to actual applications is somewhat rare.

In this paper, the first attempt for digital geologic mapping in Korea is addressed. Afterthat, basic concepts for geo-application data modeling are overviewed. And practical necessity of geo-data model is also explained. Next, UML(Unified Modeling Language) approach including OOAD(Object-oriented Analysis and Design) and its usefulness are briefly overviewed towards Korean geologic digital mapping.

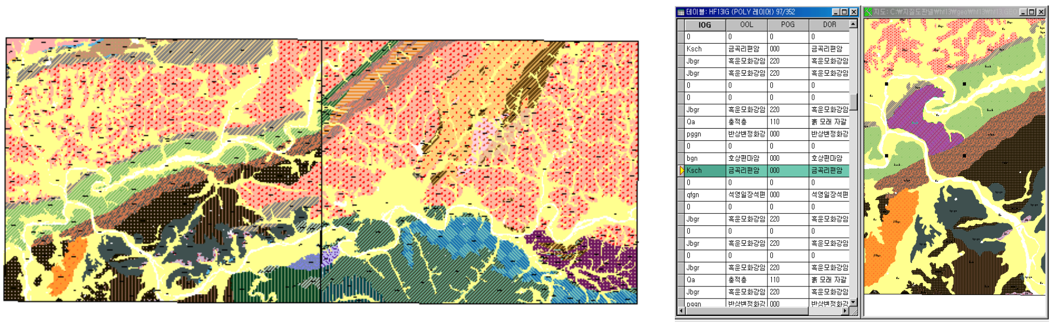
An overview of digital geologic mapping in Korea

Since 1999, the digital geologic mapping has been carried out in Korea, as a part of the National Geographical Information Systems (NGIS) policy. With geological paper maps (220 geological paper map-sheets) of the scale of 1:50,000, geology, geological structure, bedding, geological time and location of mines area are input for areas covering the whole southern parts of Korean peninsula. Also, GIS-based digital geological mapping system and web-based service system were developed for in-future digitally public services. This digital geological mapping system provides the digital information viewer, searching for each layer, on-line editing, information extraction such as area, distribution statistics, and display of more than two maps. The data formats which can be supplied are DXF, SHP(ESRI shape format), and tiff image. Figure 1 and Figure 2 show the web-based digital geological mapping system developed in this study and examples of display of two maps and SQL implementation.

After the first project for digital geologic mapping, standardization including symbol, annotation, pattern and digital code, is in progress.



<Fig. 1> Web-based geological mapping system



(a) (b)

<Fig. 2> Examples of (a) simple digital join two maps, (b) geo-processing implementation

Geo-data model and its meaning

General progress steps or levels of data models can be categorized as conceptual, logical, and physical. Conceptual model means a business representation independent of specific processing requirements and specific database management systems, and logical model is defined as a complete representation able to support actual processing needs, being independent of any specific database management system. A full logical data model includes detailed attribute characteristics such as instance identification, field lengths, data types, validity, bounds, etc. Finally, physical model means a representation of the implementation of a data model for a specific database management system capable of supporting actual application software.

Furthermore, a number of design criteria have been reported, which guide the development of the data model. Those criteria regarding geologic model are the followings. (1) The data model should be easy to implement and should place minimal requirements on the person or organization creating a digital geologic map. (2) The data model should be easily extended to include new features. These extensions should be additional tables or objects that attach additional information to the geologic map. (3) There should be a set of minimal requirements that are necessary for all geologic maps. The minimal requirements are indicated as required tables. The model should avoid explicit use of code dictionaries for translation of geologic vocabulary. (4) The data model does not address standard vocabulary but provides the capability to incorporate vocabulary standards.

In the domain of geological mapping, the real world objects modeled typically range widely from the details of individual observations to their inter-connections, and to their synthesis into explanatory structures. A number of methodologies are used to define data models. Many include diagrams and definitions. As well as these aspects, it is considered a mechanism to identify individual geologic occurrences, provide for such things as outcrop mapping, describing the lithology of a specific area within a larger map unit, or a specific segment of a fault zone, or geophysical activities.

While, geo-data model can be classified into three model, according to viewpoints: earth model, geo-spatial model, and system model. The earth Model, a model of geo-data themselves, is driven by the underlying model of the earth and the entities and phenomena upon it. It defines what we see as common earth features, such as continents and landforms, elements of the natural and human landscape, such as forests and cities, and variable characteristics of the environment, such as topography and air quality. While, geo-spatial models manifest themselves in different systematic approaches to managing collections of geo-data. These systems have evolved partly in response to their respective geo-spatial modeling requirements, and also track the continued advancements in information management technologies generally. The geographic information is collected and managed for numerous purposes, each of which has its own requirements for how data are most efficiently organized, what comprises features of interest, what degree of precision and accuracy is necessary, how information is analyzed and displayed, and so on. That is basic concept of geo-spatial model; through this model, there are now many geo-data representations, whose incompatibilities limit their utility for a community of users. Finally, system models are realized as several systems: feature-based, geo-relational, and object-oriented. An object-oriented paradigm was chosen for designing the data model. This choice ensures maximal flexibility for clients of the software, who may choose to use or modify our work at different levels of abstraction.

In summary, a geo-data model provides conceptual framework for a complete, consistent, coherent description of geo-science and adheres to both a logical view and a mathematical description.

Applicability of UML to Geo-data modeling process

The Unified Modeling Language (UML) is a language for specifying, visualizing, constructing, and documenting the artifacts of software systems, as well as for business modeling and other non-software systems (OMG, 1997). The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a well-defined and widely accepted response to that need. It is the visual modeling language of choice for building object-oriented and component-based systems.

The objectives of UML are as followings: (1) Provide sufficient semantics and notation to address a wide variety of contemporary modeling issues in a direct and economical fashion. (2) Provide sufficient semantics to address certain expected future modeling issues, specifically related to component technology, distributed computing, frameworks, and executability. (3) Provide extensibility mechanisms so individual projects can extend the meta-model for their application at low cost. (3) Provide extensibility mechanisms so that future modeling approaches could be grown on top of the UML. (4) Provide sufficient semantics to facilitate model interchange among a variety of kinds of tools. (5) Provide sufficient semantics to specify the interface to repositories for the sharing and storage of model artifacts. In terms of the views of a model, the UML defines the following graphical diagrams, use case diagram, class diagram, behavior diagrams of state diagram and/or activity diagram, interaction diagrams of sequence diagram and/or collaboration diagram, or implementation diagrams of component diagram and/or deployment diagram. The choice of what models and diagrams one creates has a profound influence upon how a problem is attacked and how a corresponding solution is shaped.

Some Cases of Geo-data Model

CSIRO model (Lamb et al. 1996)

This is the data model for such a description prepared as a guide to implementing storage, access and data exchange systems for software, developed by the CSIRO Division of Exploration and Mining and the Division of Information Technology.

An object-oriented model is used for the definition of the geometric and geological models. Primary goal of design of this model is the data classes encapsulating. It is composed of (1) Geological entities such as faults, folds, formations, contacts, histories, (2) a Quality of Information hierarchy, (3) an Audit Trail such as tracking references, sources of information, development of multiple hypotheses, (4) Multi-dimensional Geometrical Descriptions such as points, curves, surfaces, volumes, spatio-temporal regions, and (5) Geometrical Coordinate System conversions.

POSC model (1997-present)

POSC, a not-for-profit corporation, is dedicated to facilitating integrated business processes and computing technology for the exploration and production segment of the international petroleum industry. The focus of POSC model(1997) is to develop and deliver specifications for exploration and production technical applications and data management systems. These specifications describe a set of standard interfaces between software applications, database management systems, workstations and the users.

They provide a common set of specifications that allow data to flow smoothly between application programs and allow users' workflow to move smoothly from one application program to another. Finally, POSC model is one of important Logical model based on well-known E(Entity)-R(Relation) diagram.

USGS/AASG model (Johnson et al. 1998, 1999)

US National Geologic Map Data Model (NGMDM) by USGS provides a structure for a national data library that is designed to accept data from many disparate geologic maps. Therefore, this model leads to much complexity that is not needed for individual geologic maps, as well as difficulty problem at the implementation step.

The major concept is to separate symbolization from data description. These concepts lead to a large number of tables with complex linkages. The complexity of such data structures is managed through user interface tools. Richard(1998) used this model at actual database building with classified views: classification domain, relationship domain, visualization domain, and symbolization domain. Furthermore, this conceptual model has been developed as practical system model by Brodaric et al.(1999) and Gautier et al.(1999).

BGS model (Laxton and Beckon, 1996; Bain and Giles, 1997)

BGS(British Geological Survey) has developed the logical model including storage and cartographical press of geological information. Also Allen(1997) suggested the standardization of mapping practices and Giles et al.(1997) proposed geological dictionaries related to extension of system. In short, BGS model provides the conceptual and logical model directly applicable to digital geologic map system in GIS environment.

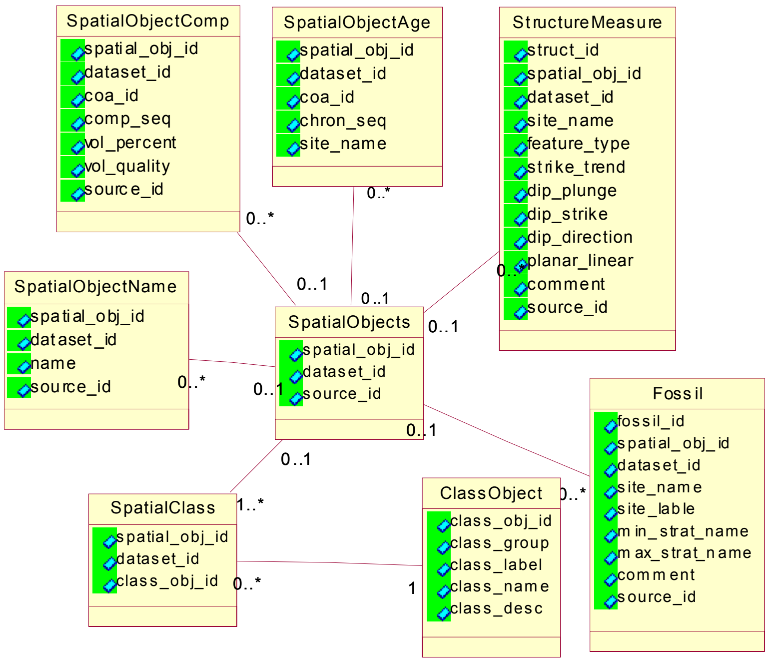
Data modeling Approach using UML

It is known that geo-scientific application using GIS is on the stage of data management, not reached geo-analysis such as geologic field works and geochemical or geophysical data processing, though a few experimental researches of geo-data integration in GIS are exists (not cited references in this paper). The key concepts of this research focused on adaptation of OOAD paradigm practically referred to USGS/AASG model for conceptual level and logical model and CISRO model for geometric process in 3D model and topology, with benefit expectation of compatibility, extensibility, and standardization.

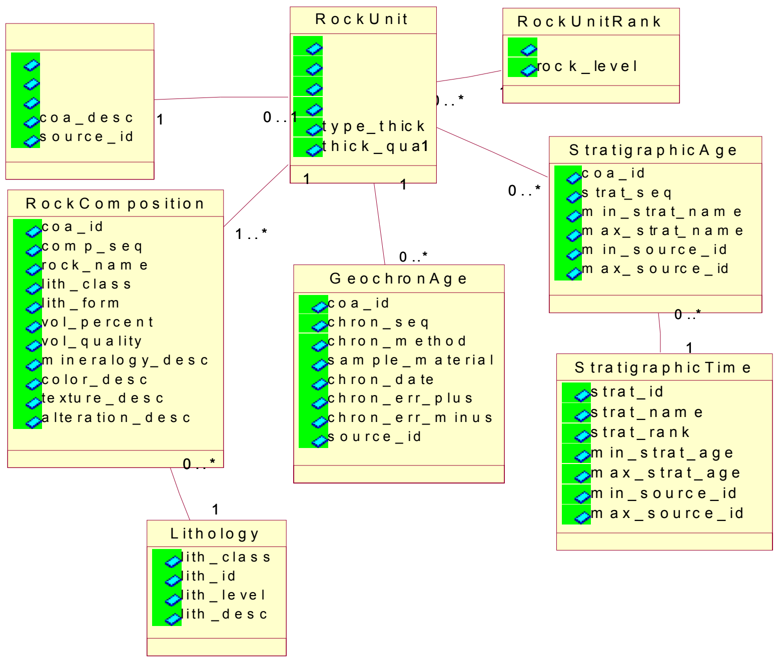
In this study, we attempted tentative conceptual geological data modeling by using UML of object oriented concepts with respect to USGS/ASSG geo-data model as an example model. As for CASE(Computer-Aided Software Engineering) tool, Rational ROSE tool is used, and this is followed by OGC modeling standard(OGC, 2000).

Figure 3 shows the UML representation of legend and metadata in USGS/ASSG model. In this, Geosopurce represents the basic paper geological map used in digital geologic mapping and ClassScheme criteria of classifying the objects. Especially, information of symbol, pattern, and color needed for cartographic processing, is designed as an extra index table.

Figure 4 shows the UML representation of rock portion of compound objects. COA represents a kind of abstract class of compound object archive, and plays a role in adding functionalities, such as overlay, inclusion, extraction of adjacency, to object. Also, RockUnitRank is the class representing a class posed in geological map. GeochronAge and StratigraphicTime represent the connection of geological time and geological object and the sequence of sedimentary beddings, respectively.



<Fig. 3> Simple UML translation of USGS/AASG geo-data model: legend and metadata portion



<Fig. 4> Simple UML translation of USGS/AASG geo-data model : rock unit portion of compound objects

Discussion and concluding remarks

This research is actually motivated from recent change of most commercialized GIS application system architecture such as objects processing in GIS engines or component-based application. However, this geo-data model can be referred to the conceptual view of a set of wide and various information, for example map themes, discrete features and objects, observations, or numeric or algorithmic descriptions. Rather, it is just initiatives for the final goal of general geo-data model for geological applications, in consideration to local and/or regional geologic condition and setting.

In OOAD view, this research is in the middle of some diagrams: use-case, class, and object, not full UML diagrams. The reason is that these diagrams are enough to represent conceptual model with logical concepts. However, other diagrams, in fact, are interoperable between each diagram so that diagram translation or conversion can be easily carried out, if key diagrams are as possible as well fixed.

As for geological applications using this model, several pre-requisites are needed: what application goal is, who uses it, how/which database uses, which software uses, and so forth. That can be directly compared with conventional GIS application development strategy using proprietary or monolithic system for limited user or special user group. In other words, this model or UML diagram provides base architecture with purpose of common uses of geo-scientific applications whatever just geologic field works, geophysical database, geochemical processing and mineral exploration and so forth. Therefore, it fits CBD(Component-Based Development) of one of major trends in recent IT(Information Technology).

Currently, the prototype model suggested in this study is under developing stage in consideration of UML approach, with national standard guideline for general digital map production. Finally, it is expected that this research is one of motives towards with geo-data model building for geological information system.

References

Allen, P. M., 1997, Standardization of mapping practices in the British Geological Survey, Computers and Geosciences, 23, p. 609-612.

Bain K. A. and Giles, J. R. A., 1997, A standard model for storage of geological map data, Computers and Geosciences, 23, p. 613-620.

Brodaric, B., E. Boisvert, and K. Lauziere, 1999, GeoMatter: A Map-Oriented Software Tool for Attributing Geological Map Information According to the Proposed U. S. National Digital Geologic Map Data Model, in Soller, D. R (ed), Digital Mapping Techniques '99 Workshop Proceedings: US Geological Survey Open File Report 99-386, p. 101-106.

Burrough, P. A., 1992, Are GIS data structures too simple-minded ?, Computers and Geosciences, 18(4), p. 395-400.

Gautier, D. L., 1999, Data Model for Single Geologic Maps: An Application of the National Geologic Map Data Model, in Soller, D. R (ed), Digital Mapping Techniques '99 Workshop Proceedings: US Geological Survey Open File Report 99-386, p. 77-81.

Giles, R. A. and Lowe, D. J. and Bain, K. A., 1997, Geological Dictionaries-Critical elements of every geological database, Computers & Geosciences, 23, p. 621-626.

Frank, A. U., 1992, Spatial concepts, geometric data models, and geometric data structures: Computers and Geosciences, 18(4), p. 409-417.

Johnson, B. R., Brodaric, B., and Raines, G. L., 1998, Digital Geologic Maps Data Model, ver. 4.2, USGS Open File Report.

Johnson, B. R., Brodaric, B., Raines, G. L., Hastings J. T. and Wahl, R., 1999, Digital Geologic Maps Data Model, ver. 4.3, USGS Open File Report, 69p.

Lamb, Peter, Horowitz, Frank, and Schmidt, H.W., 1996, A Data Model for Geoscientific Spatial Information Systems, version 1.2, Draft Nov. 27, 1996: CSIRO Division of Information Technology, 36p.

Laxton J. L. And Beckon, K., 1996, The design and implementation of a spatial database for production of geological maps, Computers and Geosciences, 22, p. 723-733.

NCA, 2000, Report on 1999 Information employment project: Digitalization of national geological map, 145p.

OMG(Object Management Group), 1997, UML Notation Guide, ver 1.1, 142p.

Open GIS Consortium Inc., 2000, Open GIS Grid Coverage Specification, Revision 0.05, OpenGIS Project Document 00-019r1, 62p.

POSC, 1997, POSC Specifications, ver 2.2, Petroleum Open Software Cooperation, 2 sets of CD-ROM.

Richard, S. M., 1998, Digital Geologic Database Model ver 1.0, Arizona Geological Survey, 21p.