ONTOLOGY SPECTRUM
FOR GEOLOGICAL DATA INTEROPERABILITY

Xiaogang Ma

Examining committee:

Prof. dr. ir. M. Molenaar	University of Twente
Prof. dr. M.J. Kraak	University of Twente
Prof. dr. ir. P.J.M. van Oosterom	Delft University of Technology
Prof. dr. ir. A.K. Bregt	Wageningen University

ITC dissertation number 197
ITC, P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN 978-90-6164-323-4

Cover picture: Fuzzified Taiji with a seal of Xiaogang Ma (Outline of seal is border of his hometown Mayang in Tianmen, China)
Spine pictures: Top: Visualization of geological time scale and a geological field photo taken in Hejin, Shanxi, China; Bottom: 'System' in ancient Chinese (c.1000BCE) and year of birth in tombstone of Benedict de Spinoza

Cover designed by Xiaogang Ma
Printed by ITC Printing Department

ITC
UNIVERSITY OF TWENTE.
FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

ONTOLOGY SPECTRUM FOR GEOLOGICAL DATA INTEROPERABILITY

DISSERTATION

To obtain
the degree of doctor at the University of Twente,
on the authority of the Rector Magnificus,
Prof. dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on November 30, 2011 at 12:45 hrs

by

Xiaogang Ma

born on December 30, 1980

in Tianmen, China

This thesis is approved by

Prof. dr. F.D. van der Meer, promoter
Prof. C. Wu, promoter
Dr. E.J.M. Carranza, assistant promoter

To Kun

.--- . ..--- ----- ..--- ----.   --... ..--- .---- --... ...-- ...--

> “ *Man thinks.* ”
>
> Benedict de Spinoza,
> *Ethics: Part II (Axiom II)*

Acknowledgements

The time for conducting this PhD research at ITC has been inspiring and enjoyable. The dissertation could not have been finished without the support and help of many people.

I am indebted to my daily supervisor Dr. John Carranza for his step by step and word by word challenging and guiding. John's supervision is sharp, which often causes sparks around my head like a boxer suffering hits in the face and followed by sleepless nights. Well, those stressed moments only occurred in the very early stages. John's supervision is effective for my research and now I can turn these sparks into flames lighting up the way in my work.

I sincerely appreciate my promoter Prof. dr. Freek van der Meer for offering me the opportunity to do this research, allowing me the freedom to do interested works and providing me insightful comments. Talking with Freek always recharges my mind with calm and confidence. I also thank my co-promoter Prof. Chonglong Wu for supporting me in scholarship application, providing me pilot data and keeping me updated on GIS studies in China.

Dr. Ernst Schetselaar paved the way for me to come to ITC, but he left for Canada three months prior to my arrival. Ernst still offered help on my research proposal and the first ISI journal paper. Ernst also introduced me to Dr. Boyan Brodaric and organized a face-to-face meeting among us three. The meeting was short but it immediately led to the revision of my work plan. Outputs of my research proved its value. Thank you, Ernst and Boyan.

I am grateful to researchers in the Commission for the Management and Application of Geoscience Information of the International Union of Geological Sciences (CGI-IUGS). In particular, I want to thank Dr. Kristine Asch, Dr. Simon Cox, Dr. Steven Richard, Dr. John Laxton, Mr. Jan Jellema, Dr. Guillaume Duclaux, Mr. Bruce Simons, Mr. Ian Jackson, Dr. Koji Wakita for discussing techniques of geoscience vocabularies and online map services. The discussion of multilingual vocabularies with Mr. Jan Jellema initiated an ISI journal paper and inspired me to continue the work of Web-based geo-spatial data services. Jan also introduced me to Dr. Jeroen Schokker and Mr. Jan Kooijman, who offered help in pilot data and an overview of Dutch geological information services. Through CGI-IUGS I also met Prof. dr. Peter Fox, Prof. dr. Stanley Finney, Dr. Marcus Ebner and Mr. Bernd Ritschel, who gave insightful comments on different parts of my work.

My affiliation with the International Association for Mathematical Geosciences (IAMG) should be appreciated. In 2005, I began to translate the book *Geoscience after IT* authored by Dr. Thomas V. Loudon, one of the founding members of IAMG. The book seeded a strong willing in my mind to take up a study in a different country. The meeting with Ernst at Liège during the IAMG 2006 conference initiated my application for ITC. After coming to Enschede, I proposed the idea of setting up the IAMG Student Chapter at ITC (ISCI) and with supports and participations of a group of enthusiasts we realized it in 2010. Through IAMG and ISCI I met so many excellent researchers of geomathematics and geoinformatics. In particular, I want to thank Prof. dr. Frits Agterberg, Dr. Harald Poelchau and Prof. dr. Gang Liu for discussing multilingual terms of geological time scale when we met at Budapest during the IAMG 2010 conference.

I express my appreciations to friends and colleagues at ITC for showing me the way in research. I thank David for teaching me research skills and Bart, Wim, Janneke, Mark, Frank, Boudewijn, Harald, Chris, Tsehaie, Marleen, Abbas, Dhruba, Cees, Victor, Norman, Dinand, Menno and Rob (Hennemann) for offering comments and suggestions. I thank Martien for offering comments and suggestions on mathematical symbols. I thank Rob (Lemmens), Dongpo and Imran for discussing ontologies and Barend, Ulanbek, Javier, Corné, Rolf, Ivana, Raul and Diego for discussing XML and online map services. Barend also translated into Duth the summary of the dissertation. I also thank Martin, Paul, Loes, Christie, Rebanna, Theresa, Bettine and Marie for organizational supports, Marga, Carla and Nina for library supports, Job and Benno for publication supports, and Ard, Harold and Aiko for IT supports. In addition, the work in the ITC faculty council (2010) allowed me learn a lot from Tom (Veldkamp), Erna, Sjef, Tom (Loran), Barend, Wan, Petra and Lieke.

I am heartily grateful to my PhD friends at ITC. Pablo, Sanaz and Waew, my officemates, thank you for offering help in work and life. It is a great pleasure to share the PhD journey with Juan (Francisco), Mila, Sekhar, Nicky, Wiebke, Armindo, Monica, Coco, Mobushir (Khan), Sabrina, Nugroho, Sharon, Frederick, Saibal, Tolga, Pankaj, Zack, Thi Hai Van, Byron, Paresh, Khamarrul, Sumbal, Shafique, Shruthi, Andre, Getachew, Muhammad (Yaseen), Fekerte, Nasrullah, Juan (Pablo), Tanmoy, Clarisse, Irma, Gaurav, Rishiraj, Anthony, Yaseen (Mustafa), Adam, Amjad, Anas, Atkilt, Babak, Aidin, Claudia, Abel, Ullah, Maitreyi, Rafael, Mariela, Flavia, Sukhad, Divyani, Ngoc Quang, Alphonse, Sejal, Christine, Arif, Joris, Enrico, Syarif, Laura, Alain, Mustafa (Gokmen), Tanvir, Fouad, Jahanzeb, Jeniffer, Leonardo and Kitsiri. I wish you successful in your study and career.

My Chinese friends at ITC deserve much more than a special paragraph for receiving my thanks. Yijian, Tiejun, Shi Pu, Xia Li, Zhenshan, Ningrui, Xin Tian, Changbo, Longhui, Ouyang, Lei Zhong, Qiuju, Jing Xiao, Xiang Zhang, Xi Zhao, Fangfang, Pu Hao, Yali, Xuanmei, Liang Zhou, Meng Bian, Teng Fei, Xuelong, Xiaojing, Yixiang, Biao Xiong, Sudan, Xinping, Bob Su, Linlin, Tina, Lichun, Ye Du, Zhiling, Siqi, Yawen and Jun Wang, my best wishes for all of you.

I will never forget the running experience with the Run4Fun group at ITC, expecially the Batavierenrace, which defined another me. Thank you, Wan, Simon, Roelof, Emile, Gabriel and many other friends.

I appreciate my parents for allowing me stay so far away from them (感谢父母允许我远游而久不归), and my elder brother Xiaohu and elder sister Xiaofang for their help in the past thirty years.

Last but not least, I express my wholeheartedly thanks to my wife Kun for her understanding, support and devotion during the time of this PhD research. I also want to thank my son Kewei for bringing so much happiness to the family. I owe both of you too much and I definitely will spend more time with you.

Xiaogang Ma
2011, Enschede

Table of Contents

List of figures

List of tables

List of abbreviations

1G	OneGeology
1G-E	OneGeology-Europe
5W1H	Who, When, Where, Why, What and How
AGROVOC	Multilingual agricultural vocabulary coordinated by Food and Agriculture Organization of the United Nations
Ajax	Asynchronous JavaScript and XML
AMTG	Asian Multilingual Thesaurus of Geosciences
ANSI	American National Standards Institute
AQSIQ	General Administration of Quality Supervision, Inspection and Quarantine of P.R. China
AuScope	Organization for a national earth science infrastructure program (Australia)
BGS	British Geological Survey
CCOP	Coordinating Committee for Geoscience Programmes in East and Southeast Asia
CGI-IUGS	Commission for the Management and Application of Geoscience Information of the International Union of Geological Sciences
CGMW	Commission for the Geological Map of the World
CGS	China Geological Survey
CIFEG	Centre International pour la Formation et les Echanges en Géosciences (International Center for Training and Exchanges in the Geosciences)
CPRM	Companhia de Pesquisa de Recursos Minerais (Geological Survey of Brazil)
CSW	Catalog Service for the Web
DFM	Data-Flow Model
DL	Description Logic
DNPM	Departamento Nacional de Produção Mineral (National Department of Mineral Production, Brazil)
FAO	Food and Agriculture Organization of the United Nations
FL	Fuzzy Logic
GEON	Opening collaborative project developing cyberinfrastructure for integration of 3 and 4 dimensional earth science data
GIN-RIES	Groundwater Information Network-Réseau d'Information sur les Eaux Souterraines (Canada)
GSJ-AIST	Geological Survey of Japan-National Institute of Advanced Industrial Science and Technology
GSSP	Global Boundary Stratotype Section and Point
GTS	Geological Time Scale
HTML	HyperText Markup Language
ICS	International Commission on Stratigraphy
ICS chart	International Stratigraphic Chart coordinated by International Commission on Stratigraphy
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
ITC	Faculty of Geo-Information Science and Earth Observation

	(ITC), University of Twente
KML	Keyhole Markup Language
LBLE	Letter-by-Letter Equality
MLTGTS	Multilingual Thesaurus of Geological Time Scale
MTG	Multilingual Thesaurus of Geosciences
NADM	North American Geologic Map Data Model
NGMDB	National Geologic Map Database (United States)
NISO	National Information Standards Organization (United States)
NMRA	National Mineral Resources Assessment (China)
OGC	Open Geospatial Consortium
OOM	Object-Oriented Model
OWL	Web Ontology Language
RDF	Resource Description Framework
RGB	Red-Green-Blue
SKOS	Simple Knowledge Organization System
SLD	Styled Layer Descriptor
SOL	Second-Order Logic
SVG	Scalable Vector Graphics
UML	Unified Modeling Language
UNECE	United Nations Economic Commission for Europe
UNFC	United Nations Framework Classification for Energy and Mineral Resources
UNPSC	United Nations Standard Products and Services Code
USGIN	United States Geoscience Information Network
W3C	World Wide Web Consortium
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WPS	Web Processing Service
XML	eXtensible Markup Language

Chapter 1

Introduction

1.1 Background and motivation

Geological data – records of physical structures and substances of the earth, their history, and the processes associated with them – are not only essential for studying our mother planet but also for addressing key societal challenges. Examples can be seen in resources exploration and management (Agterberg, 1989; Bonham-Carter, 1994; Carranza, 2009), urban development (Dai et al., 2001; Culshaw et al., 2009), climate change (Anandakrishnan et al., 1998; Gerhard et al., 2001), water quality (Sharpe et al., 1987; Roy et al., 2001; Pipkin et al., 2008), and hazard mitigation (Michael and Eberhart-Phillips, 1991; Bell, 2003), etc. In the present Digital Age (Kleppner and Sharp, 2009), computer-based hardware and software are being widely used in the capture, update, integration, analysis, evaluation and publication of geological data. Compared to the ongoing deluge of digital geological data, approaches for promoting effective geological data interoperability are currently underdeveloped. Interoperability of geological data, thus, has long been a topic of concern in scientific works.

Interoperability is essential for efficient information retrieval and knowledge discovery in studies and applications using geological data (cf. Loudon, 2000; Richard et al., 2003; Carranza et al., 2004; Asch, 2005; Brodaric and Gahegan, 2006; Gahegan et al., 2009). Challenges of data interoperability can arise at different levels, such as systems (i.e., network and services), syntax (i.e., language and encoding), schemas (i.e., modeling and structure), semantics (i.e., content and meaning), and pragmatics (i.e., use and effect) (Bishr, 1998; Harvey et al., 1999; Sheth, 1999; Ludäscher et al., 2003; Brodaric, 2007). In this dissertation, ***geological data interoperability is defined as the ability of geological data provided by a data source to be accessed, decoded, understood and appropriately used by users***. Among the various finished and/or ongoing studies addressing geological data interoperability, ontology-based approaches have attracted increasing attentions in recent years to address geological data interoperability.

Ontologies in computer science are defined as shared conceptualizations of domain knowledge (Gruber, 1995; Guarino, 1997b), which originate from *the* *study of being* in philosophy. Ontologies have been extensively studied to address data interoperability issues in different knowledge or scientific domains, such as genetics (Ashburner, 2000), geographical information (Frank, 2001), soil classification (Rossiter, 2007), and solar-terrestrial physics (Fox et al., 2009), etc. It was increasingly discussed in the literature (e.g., Welty, 2002; McGuinness, 2003; Obrst, 2003; Uschold and Gruninger, 2004; Borgo et al., 2005) that, in building and using ontologies, it is worth to keep in mind an ontology spectrum, which covers ontology types with varying semantic richness (Fig. 1.1).

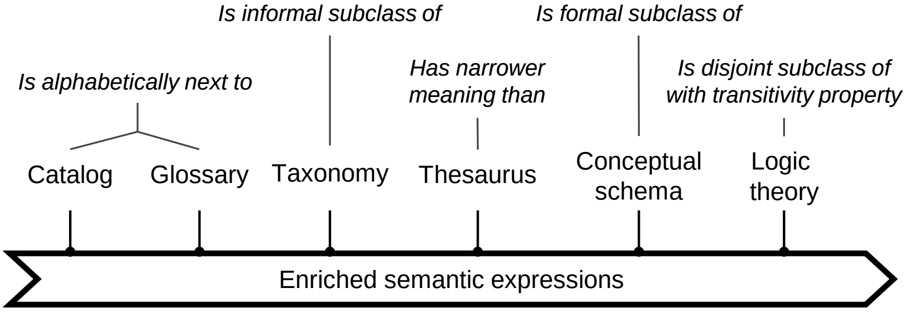


Fig. 1.1. Ontology spectrum (adapted from Welty, 2002; McGuinness, 2003; Obrst, 2003; Uschold and Gruninger, 2004; Borgo et al., 2005). Texts in italics explain a typical relationship in each ontology type.

In the field of geological ontologies, there are examples of controlled vocabularies (e.g., Bibby, 2006; Richard and Soller, 2008; Ma et al., 2010b), conceptual schemas (e.g., Brodaric, 2004; NADM Steering Committee, 2004; Richard, 2006) and logical language-based formal ontologies (e.g., Ludäscher et al., 2003; Raskin and Pan, 2005; Tripathi and Babaie, 2008), etc. In several recent projects, different types of ontologies have been applied to provide featured functions in national, regional and global geological data infrastructures, thereby promoting geological data interoperability and facilitating information retrieval and knowledge discovery in applications. The AuScope[1] project built vocabulary-based services for querying geological maps to overcome differences in geoscience terms due to language, spelling, synonyms and local variations and, thus, help users to find desired geological information of Australia (Woodcock et al., 2010). The NADM model/schema (NADM Steering Committee, 2004) was proposed and implemented in the NGMDB[2] project (Soller and Berg, 2005) to promote collaborations among

---

[1] http://www.auscope.org [Accessed March 21, 2011].

[2] http://ngmdb.usgs.gov [Accessed March 21, 2011]

geological map databases in the United States. The OneGeology (1G)[3] project adopted the GeoSciML (Sen and Duffy, 2005) as a common conceptual schema and online exchange format to improve the exchange/integration of online geological maps distributed globally (Jackson, 2007). GeoSciML was also applied in the OneGeology-Europe (1G-E)[4] project, and the 1G-E further extended vocabulary-based services to enable multilingual annotation and translation of geological map contents among 18 European languages (Asch et al., 2010; Laxton et al., 2010). Strategies similar to 1G (i.e., applying common conceptual schemas among distributed data sources) were also applied in the USGIN[5] project in the United States (Allison et al., 2008) and the GIN-RIES[6] project in Canada (Brodaric et al., 2009) to address interoperability of geoscience and groundwater information, respectively. In the GEON[7] project, formal ontologies were used to mediate conceptual schemas of heterogeneous geological maps and enable semantic integration among them (Ludäscher et al., 2003; Baru et al., 2009).

In the aforementioned studies and application projects, substantial progress has been made in developing geological ontologies and in using them to mediate heterogeneous geological data, in which the capability of ontologies for promoting geological data interoperability is commonly acknowledged. A technical trend in those projects is deploying works in the environment of the Semantic Web (Berners-Lee et al., 2001; Hendler, 2003) and developing ontologies with Web-compatible global standards (e.g., eXtensible Markup Language (XML) or sub-languages of XML, such as W3C® proposed Simple Knowledge Organization System (SKOS), Resource Description Framework (RDF) and Web Ontology Language (OWL), etc.).

Despite the progress in building and using different types of geological ontologies, the application of an ontology spectrum to promote geological data interoperability still faces vast challenges, among which are the following key challenges addressed in this dissertation:

(1) Modeling and encoding of ontologies – modeling transforms humans' tacit knowledge of a domain into concepts and relationships, whereas encoding implements the modeling with symbols/languages in a context (cf. Kuhn, 2010). Modeling can generate varied semantic richness of ontologies, whereas encoding is related to the environment in which ontologies are used. Differences and relationships between modeling and

---

[3] http://www.onegeology.org [Accessed March 21, 2011].
[4] http://www.onegeology-europe.org [Accessed March 21, 2011].
[5] http://www.usgin.org [Accessed March 21, 2011].
[6] http://www.gw-info.net [Accessed March 21, 2011].
[7] http://www.geongrid.org [Accessed March 21, 2011].

encoding of ontologies are less discussed for applications in the field of geology;

(2) Multilinguality of geological data and ontologies – geological units are naturally independent of language borders, but geological data are not, whereas commonly agreed multilingual ontologies are limited in many subjects in geology (cf. Asch and Jackson, 2006), and applications of multilingual geological ontologies with online geological data are underdeveloped;

(3) Flexibility and usefulness of ontology-based applications – incorporating ontologies into state-of-the-art technologies in geo-information science, such as OGC® web service standards, algorithms of information retrieval (e.g., Baeza-Yates and Ribeiro-Neto, 2011), conceptual mapping (e.g., Noy, 2009) and data visualization (e.g., Fox and Hendler, 2011), etc., allow exploration and evaluation of the potential of ontologies for promoting interoperability of geological data; and

(4) Mediation and evolution of geological data and ontologies – heterogeneous geological data can be mediated in a short-term period, but data are continuously flowing and updated in a long-term perspective and, thus, paradigms are needed to address the interoperability of geological data underpinned by ontologies in an evolving environment.

1.2 Study objectives

The research leading to this dissertation aimed to explore approaches to address the aforementioned key challenges, and thus to provide a route map for applying an ontology spectrum to promote geological data interoperability at local, regional and global levels. The dissertation answers the following research questions.

(1) How can ontologies be modeled and encoded, so that the resulting ontologies are not only efficient for harmonizing local geological data but also function to improve the interoperability of local or internal geological data with extramural or external projects?

(2) In a regional/global environment, how can linguistic barriers of online geological data be alleviated by building and using multilingual ontologies?

(3) How can different methods of conceptual analysis be integrated to develop thematic conceptual schemas that are efficient for problem-solving and are compatible with commonly used standards in the field of geology?

(4) How can ontology-based tools be developed to improve the interoperability of online geological data, so as to help both geologists

and non-geologists to retrieve geological information and discover geological knowledge?

(5) What are the context-caused challenges for geological data interoperability, and how can these challenges be addressed in a long-term perspective?

To provide insights into the above-stated research questions, results of research case studies of geological data interoperability at local, regional and global levels are described in this dissertation. Several types of ontologies such as taxonomies, thesauri, conceptual schemas and RDF/OWL-based ontologies were developed and deployed, respectively, according to the context of each research case study. Based on the results of these research case studies, this dissertation discusses answers to each of the above-stated research questions and, as a whole, presents strategies and methods for properly deploying an ontology spectrum in practices to promote geological data interoperability.

1.3 Dissertation outline

The dissertation consists of seven chapters. The five core chapters (2–6) focus on the aforementioned five research questions, respectively. These chapters have either been published or are submitted for publication as peer-reviewed papers in ISI-indexed journals.

Chapter 1 describes the research background and the key challenges, and then specifies research questions in the research objectives and outlines the structure of the dissertation.

Chapter 2 describes methods developed for organizing, encoding and building concepts in a taxonomical controlled vocabulary for local geological data in mining projects. A strategy of “global thoughts and local actions” is deployed in the work to promote both the harmonization of geological data within a local context and the interoperability of local geological data with the external environment.

Chapter 3 describes a SKOS-based multilingual thesaurus of geological time scale developed for alleviating linguistic barriers of geological time scale records among online geological maps, and discusses methods to obtain satisfactory semantic expressions of concepts in a thesaurus.

Chapter 4 describes construction and application of conceptual schemas/models for geological data in the compositing of borehole metal-grade intervals, and discusses both data-flow models and object-oriented models for developing a computer program. Concepts in these two groups of models are compatible with commonly used standards in the field of mineral resources estimation.

Chapter 5 describes a RDF/OWL-based ontology of geological time scale developed to support annotation, visualization, filtration and generalization of geological time scale information from online geological map services, and evaluates the usefulness of the developed works with a user-survey.

Chapter 6 demonstrates a model representing contexts of geological data sources, and proposes a procedure of semantic negotiations for approaching pragmatic interoperability of distributed geological data in an evolving environment.

Chapter 7 summarizes the results discussed in Chapters 2–6, presents answers to the research questions, describes the main contributions of this study, and provides recommendations for further studies.

Chapter 2

A controlled vocabulary for interoperability of local geological data

This chapter is based on: Ma, X., Wu, C., Carranza, E.J.M., Schetselaar, E.M., van der Meer, F.D., Liu, G., Wang, X., Zhang, X., 2010. Development of a controlled vocabulary for semantic interoperability of mineral exploration geodata for mining projects. Computers & Geosciences 36 (12), 1512–1522.

2.1 Introduction

Many geological data (geodata) are captured and used within local contexts, such as mineral exploration geodata, whereas the interoperability of these geodata is of less concern. Mineral exploration is a continuous process involving integration and re-use of multi-source, multi-disciplinary and multi-temporal geodata. Geodata accumulated in preceding and ongoing mineral exploration projects should be structured orderly and re-used as necessary, in order to advance the understanding of geological assurance, economic viability and exploitation feasibility of mineral deposits for mining. However, inconsistent conceptual schemas and heterogeneous terms among diverse mineral exploration geodata sources may hinder their efficient use and/or re-use in mining projects, as well as for sharing of geodata for further applications in the same or related knowledge domains, such as for estimation of mineral resources and confirmation of estimates (Carranza et al., 2004; Ma et al., 2007). A possible solution to this problem is a controlled vocabulary-driven database scheme derived from studies on ontology-based information systems (Guarino, 1998; Sugumaran and Storey, 2006).

In general, a controlled vocabulary is a set of consistent terms used within a specific knowledge domain (Smith and Kumar, 2004; Soller and Berg, 2005; Richard and Soller, 2008). In a controlled vocabulary, the same concept (i.e., notions, ideas or principles) is represented by the same term (or group of terms). In this regard, a controlled vocabulary-driven database scheme is often used in applications (e.g., cross-database queries (Jaiswal et al., 2005; McGuinness et al., 2006) and integration of heterogeneous databases (Linnarsson, 1989; Ludäscher et al., 2003) which need a common representation and understanding of concepts within a knowledge domain. Therefore, if the scheme of a controlled vocabulary-driven database is followed for different applications within a knowledge domain, diverse local schemas can be mapped to unified schemas, while inconsistent terms from each application can be mapped to standard terms provided by a controlled vocabulary. Heterogeneous geodata sources in a mining project can thereby be transformed to a consistent form in a mineral exploration geoscience database.

Since mineral exploration for mining applications is a multi-disciplinary synthesis of numerous concepts, a proper representation of concepts and their inter-relationships is essentially needed in the controlled vocabulary (i.e., internal aspects of a controlled vocabulary). Moreover, in order to improve the interoperability of mineral exploration geodata for mining projects, the controlled vocabulary underpinning them should be interoperable with concepts in related applications in the mineral exploration domain (i.e., external aspects of a controlled vocabulary). Thus, the purpose of the study described in this chapter is to develop methods for organizing, encoding and building concepts in a controlled vocabulary for mining applications of mineral exploration geodata, so as to make such a controlled vocabulary not only efficient for reconciling heterogeneous geodata in various mining projects, but also consistent and coherent with other concepts in the mineral exploration domain.

2.2 Methods for building a controlled vocabulary

A controlled vocabulary is necessary basis for the ontology of a knowledge domain (Gruber, 1995; Guarino, 1997b). An effective way to build ontology is to start using current professional standards and dictionaries, and then modify and/or extend it (McGuinness, 2003; Bibby, 2006). In the same way, in developing the controlled vocabulary discussed here, a Chinese national standard (AQSIQ, 1988) and several other standards derived from it were referred for geoscience taxonomies and terms, because these standards are widely accepted and used in mineral exploration in China.

Several adaptations were made to transform the original national standards into a desired controlled vocabulary. The domain of mineral exploration for mining applications was classified into subjects, subclasses and concepts, which were embedded into the hierarchical (i.e., taxonomical) organization structure of the controlled vocabulary. In accordance with this organization structure, a coding method was applied to provide a unique code for each subject, subclass or concept. In order to support applications in databases, a metadata schema was also developed for the definition of terms in the controlled vocabulary. The developed controlled vocabulary provides an extensible structure so that new subjects, subclasses or concepts evolving from mineral exploration in a mining project can be added. The following sections describe in detail the guidelines and procedures for developing the controlled vocabulary.

2.2.1 Representation and organization of concepts

The study of an ontology spectrum (Welty, 2002; McGuinness, 2003; Obrst, 2003; Uschold and Gruninger, 2004; Borgo et al., 2005) reveals the relationship between a controlled vocabulary and an ontology (Fig. 2.1). A catalog and a glossary are both regarded as a simple controlled vocabulary, because they are both often only an alphabetical list of terms; whereas a taxonomy and a thesaurus are both regarded as a complex controlled vocabulary, because they both enrich definitions of concepts and relationships between concepts (ANSI/NISO, 2005; Coleman and Bracke, 2006). However, relationships between concepts in a taxonomy or thesaurus are often informal (e.g., a subclass may not inherit all the properties of its superclass). A conceptual schema (e.g., an object-oriented conceptual schema) formalizes the relationships between concepts (e.g., a subclass inherits properties of a superclass) (McGuinness, 2003). Nevertheless, catalogs, glossaries, taxonomies, thesauri and conceptual schemas are all machine-processable, but they are not machine-interpretable and thus cannot be used to make valid inferences (Obrst, 2003). In order to improve or attain machine-interpretability, a formal ontology is described by a logic theory (e.g., the Description Logic).

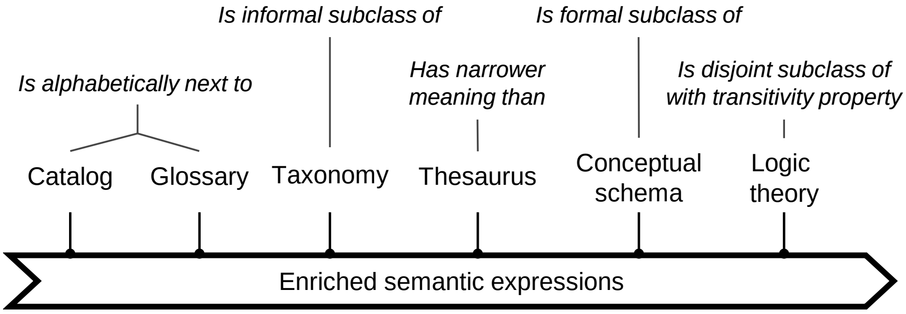


Fig. 2.1. Ontology spectrum (adapted from Welty, 2002; McGuinness, 2003; Obrst, 2003; Uschold and Gruninger, 2004; Borgo et al., 2005). Texts in italics explain a typical relationship in each ontology type.

Defining concepts and their inter-relationships involves both semantics and syntax (Guarino, 1997a; Raskin and Pan, 2005; McGuinness et al., 2007;

Durbha et al., 2009). Semantics deals with the meanings of concepts and syntax deals with the structure of expressions in a language. Generally, concepts in a controlled vocabulary are represented by terms in a natural language (Bronowski and Bellugi, 1970; Boguraev and Kennedy, 1997; Helbig, 2006), which is a human language that has evolved naturally in a community and is typically used for communication. The natural language-based representation of a concept should have a clear and distinct form, so as to reveal the intended meaning of this concept within a domain (Babaie et al., 2006). This is important and necessary as users can only access the meaning of a concept in a form that they can understand and use (Sinha et al., 2007). Therefore, terms used for representing concepts within a controlled vocabulary should be restricted and organized according to certain semantic and syntactic guidelines.

For implementation of semantics and syntax in a controlled vocabulary, ISO/IEC 11179-5 (ISO, 2005) recommends that the name of a concept may consist of four terms (Fig. 2.2): object class term, qualifier term, property term and representation term. An object class term represents a genus or category to which a concept belongs. A qualifier term represents a differentia that distinguishes a concept from other concepts within the same object class. A property term represents a common characteristic of all concepts belonging to the same object class. A representation term describes the form of a set of valid values of a concept. A representation term may be overlapped with part of the property term and is often eliminated.

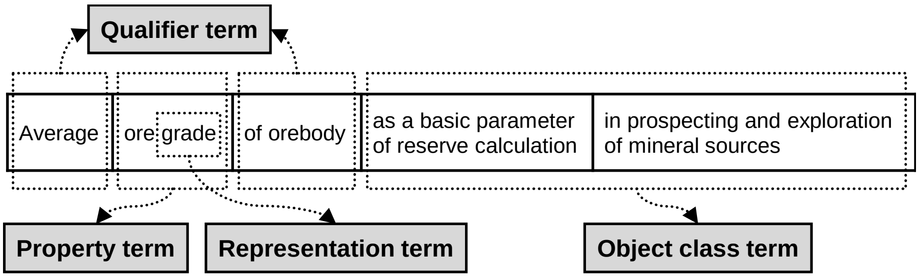


Fig. 2.2. Semantic and syntactic compositions of terms representing a concept based on guidelines recommended by ISO/IEC 11179-5 (ISO, 2005).

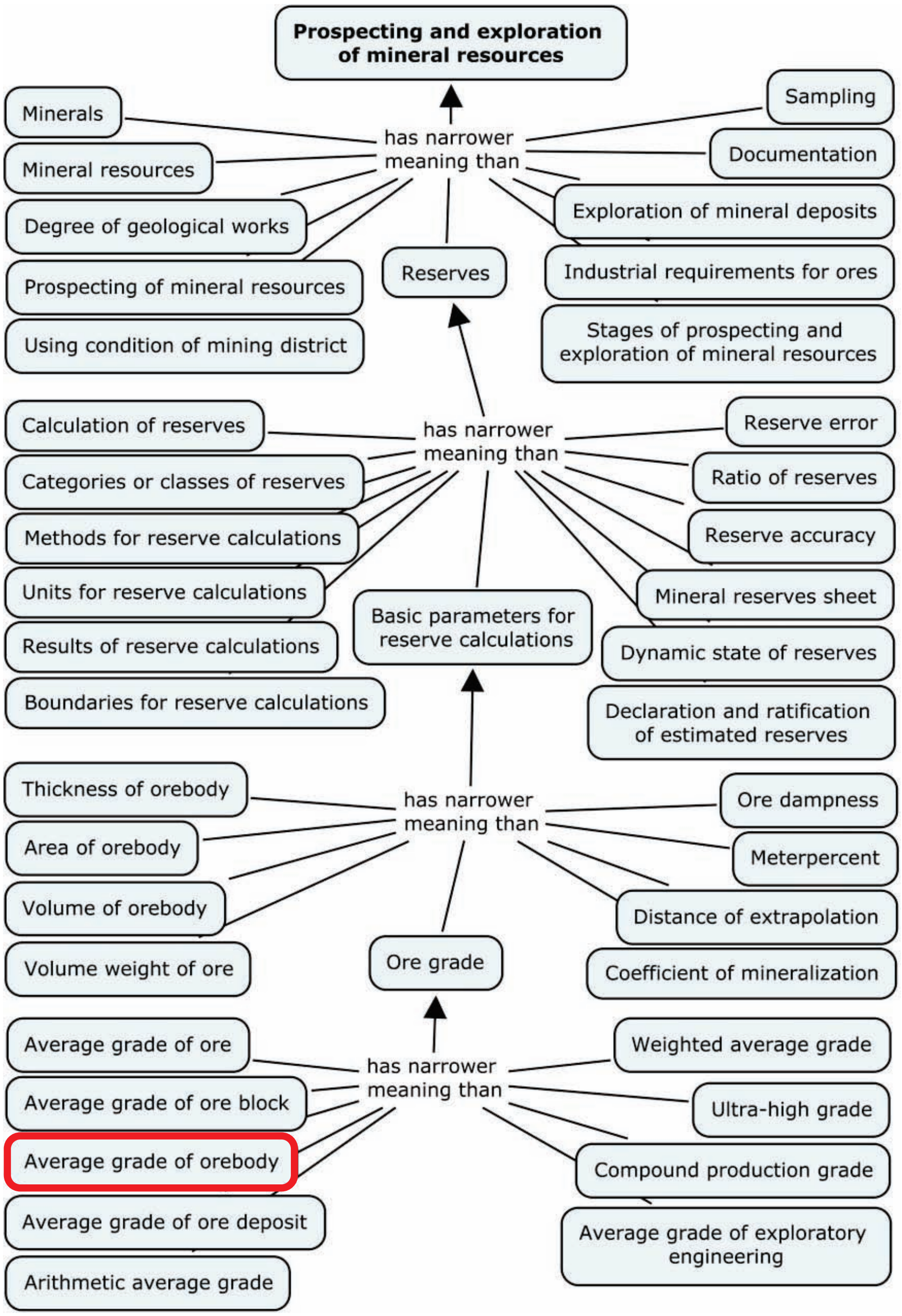


Fig. 2.3. Hierarchical structure revealed by object class of a concept. A four-level hierarchy of subject and subclasses is derived from a concept name shown in Fig. 2.2. Terms in this diagram are derived from a Chinese national standard (AQSIQ, 1988).

For a group of concepts within the same object class, they share the same object class term. Generally, an object class is a subclass within a subject (i.e., a branch of knowledge), and there may be a hierarchical structure of subdivisions from a subject to a subclass. Explanation of a hierarchical structure needs a group of terms. However, putting these terms into the name of every concept within an object class causes huge redundancy. Instead, previous studies (Gillespie and Styles, 1999; Brodaric et al., 2002; Huber et al., 2003) propose that the hierarchical structure of subject and subclasses can be represented and implemented in the organization structure of a controlled vocabulary. Such an organization structure sets up a context for concepts within an object class and helps to ascertain/explain the meaning of each concept.

In the controlled vocabulary discussed here for mining applications of mineral exploration geodata, the aforementioned guidelines were followed to set up hierarchical/taxonomical organization structures and concise names for concepts. For example, Fig. 2.2 shows the name of a concept that contains a long explanation of its object class. This object class was divided into three levels: the subject “Prospecting and exploration of mineral resources” and two subclasses “Reserves” and “Basic parameters for reserve calculations”. Meanwhile, another subclass “Ore grade” was taken out as it is a common property of a group of concepts. Thus, a four-level hierarchy of subject and subclasses was set up as an organization structure; and as a result, names of concepts within the same object class were simplified (Fig. 2.3).

2.2.2 Encoding and definition of concepts

Compared to hierarchically organized concepts (or terms) expressed in a natural language, a coding system is a simpler representation of concepts in a knowledge domain (Loudon, 2000; Deissenboeck and Pizka, 2006; Sinha et al., 2007). Codes use short abbreviations to represent information defined by concepts or terms (Mori, 1995; Cimino, 1998). A unique code can be assigned to each concept or term in order to reveal their hierarchical levels in a controlled vocabulary. A typical example is the United Nations Standard Products and Services Code (UNSPSC). It provides classification codes that clearly reveal hierarchical levels of terms (UNSPSC, 2004). For example, the category “11-Mineral and textile and inedible plant and animal materials” has a subclass “111-Minerals and ores and metals”, which in turn has a subclass “111015-Minerals”, and the subclass “111015-Minerals” contains different mineral names, such as “11101501-Mica”, “11101502-Emery” and “11101503-Quartz”, etc.

In order to encode concepts in the controlled vocabulary discussed here, the following guidelines were followed (Wang et al., 1999): (1) each subject, subclass or concept has a unique code; (2) the code of a subject or subclass is included in the code of its subclasses and concepts; and (3) pure (i.e., alphabetic) codes are adopted for subjects, subclasses and concepts that can be used as fields in a database, whereas mixed (i.e., alphanumeric) codes are adopted only for concepts that can be used as records in a table column. For example, the pure code “PKCDDC” represents the concept “Average grade of orebody”. The subject and subclasses related to codes “PK”, “C”, “D” and “D” are listed in Table 2.1. If a subclass contains a number of concepts in an enumeration scheme (i.e., an exact listing of all concepts within a subclass), it is clearer and easier to use mix codes to encode these concepts. For example, in Table 2.2, the code “YSEB14801” represents the concept “Granite”, in which “YS”, “E” and “B” are respectively related to the upper subject and subclasses of the concept.

Table 2.1 Examples of pure codes and hierarchical levels they represent

Level	**Code**	**English name**
Subject	PK	Prospecting and exploration of mineral resources
|–Subclass	PKC	Reserves
| |–Subclass	PKCD	Basic parameters for reserve calculations
| | |–Subclass	PKCDD	Ore grade
| | | |–Concept	PKCDDC	Average grade of orebody

Table 2.2 Example of mixed codes and hierarchical levels they represent

Level	**Code**	**English name**
Subject	YS	Petrology
|–Subclass	YSE	Classification and name of rocks
| |–Subclass	YSEB	Rock name
| | |–Concept	YSEB14801	Granite

Another purpose of assigning unique codes to every concept, subclass or subject is to use those codes as field names in a database, in order to avoid errors caused by Chinese field names and to speed up database queries. Using codes instead of Chinese letters as field names depends on the multi-lingual compatibility of the database system adopted (Chen et al., 2003), although using Chinese letters as field names may cause errors when transferring records between two different database systems (Zhang, 2002). The controlled vocabulary discussed here provides terms in both English and

Chinese versions. Names of subjects, subclasses and concepts shown in Table 2.2 and Table 2.3 are retrieved from the English version, whereas in the actual works of data integration in several mining projects in China only the Chinese version has been used. The codes were used as field names (i.e., to be used in physical databases and SQL scripts) and the names of related concepts or subclasses as field captions (i.e., to be shown in user interfaces).

Table 2.3 Metadata elements defining concepts, subclasses or subjects as fields in databases

Element name	Description	Example
Code	An abbreviation representing a concept, subclass or subject uniquely	PKCDDC
Chinese name	A Chinese string representing a subject, subclass or concept	矿体平均品位
English name	A English string representing a subject, subclass or concept	Average grade of orebody
Data type	Data type of a field (e.g., Text, Number, Date, etc.)	Number
Specified data type	Data type defined in computer language	Float
Decimal place	Number of decimal places of numerical records	4
Unit	A division of quantity accepted as a standard of measurement or exchange	% or g/t
Required	Indicates whether a record is required in a table column	TBD* in application
Null	Indicates whether a null record is valid in a table column	TBD in application
Default value	A value automatically being assigned to a new record input to a table column	TBD in application
Max value	Maximum value of a record stored in a table column	TBD in application
Min value	Minimum value of a record stored in a table column	TBD in application
Restriction	Other requirements for records stored in a table column	Only applicable for solid minerals
Remarks	Additional descriptions and restrictions	Unit “%” for base metals; and “g/t” for precious metals

*TBD = to be determined.

For the controlled vocabulary discussed here, one of its primary functions is modeling conceptual schemas (cf. Bermudez and Piasecki, 2006; Batanov and Vongdoiwang, 2007) for mineral exploration geodatabases in mining projects. In this regard, a metadata schema was developed (Table 2.3), by which concepts, subclasses and subjects in the controlled vocabulary can be defined as fields in databases.

2.2.3 Extensible structure for adding new concepts

With the aforementioned guidelines and methods, the controlled vocabulary for mineral exploration geodatabases of mining projects was developed, in which the subjects cover almost all the topics of geology and mineral resources (Fig. 2.4). Nevertheless, it is necessary in further studies to create new subclasses, concepts as well as subjects in the controlled vocabulary, because new terms for objects and properties may evolve from actual works in different mining projects. This requires that the controlled vocabulary discussed here has an extensible structure. The methods for organizing, coding and defining concepts set up an “umbrella” structure that supports extensions of the controlled vocabulary.

A new concept that evolves from actual mining works can be compared with existing concepts and subclasses in the hierarchy, in order to check whether the new concept can be included in an existing subclass; or else, a new subclass can be created and into which this new concept can be included. For example, in an old version of the controlled vocabulary, there were no concepts corresponding to borehole inclination data. Since inclination record is a part of borehole loggings that belong to the subject “Geophysical exploration”, a new subclass “Borehole inclination record” and its relevant concepts in this subject were created (Table 2.4). Meanwhile, the coding method was used to assign a unique code to each newly added subclass or concept.

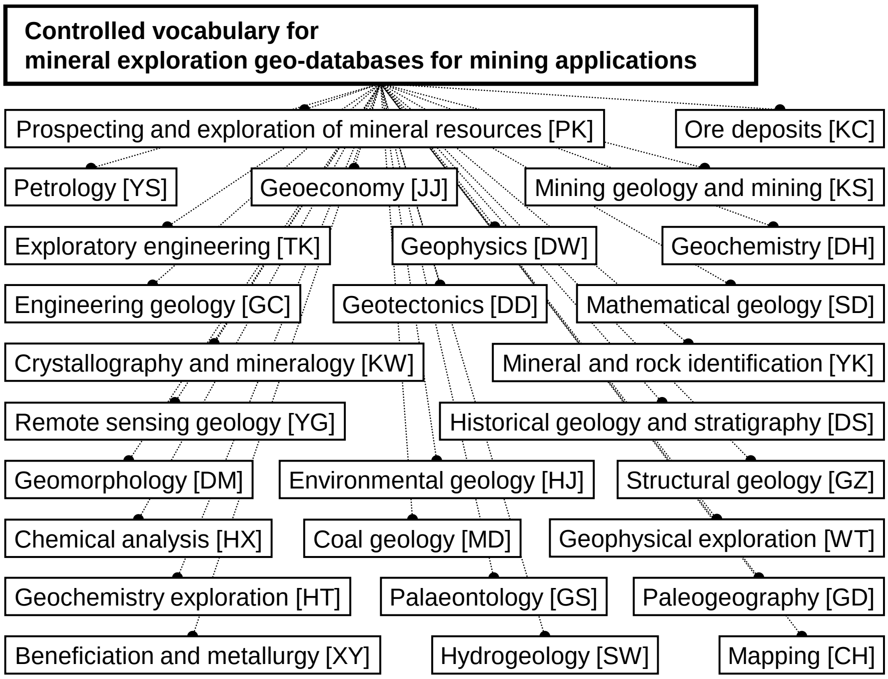


Fig. 2.4. Subjects in studied controlled vocabulary for mineral exploration geodata in mining projects. Codes of subjects are shown in square brackets next to names of subjects.

Table 2.4 Naming and coding of newly added subclass “Borehole inclination record” and relevant concepts

Level	Code	English name
Subject	WT	Geophysical exploration
|–Subclass	WTH	Well logging terminologies
| |–Subclass	WTHG	Recording in site
| | |–Subclass	WTHGF	Borehole inclination record
| | | |–Concept	WTHGFA	Measured deviation angle
| | | |–Concept	WTHGFB	Examined deviation angle
| | | |–Concept	WTHGFC	Adopted deviation angle
| | | |–Concept	WTHGFD	Measured azimuth angle
| | | |–Concept	WTHGFE	Examined azimuth angle
| | | |–Concept	WTHGFF	Adopted azimuth angle

2.3 Case study to standardize and integrate multi-source borehole databases

The controlled vocabulary discussed here has been used for reconciling heterogeneous geodata and setting up integrated databases for various mining projects of the Zijin Mining Group in China. In this section, a case study using multi-source borehole data is described to demonstrate applications of the controlled vocabulary for standardization and integration of mineral exploration geodata for mining.

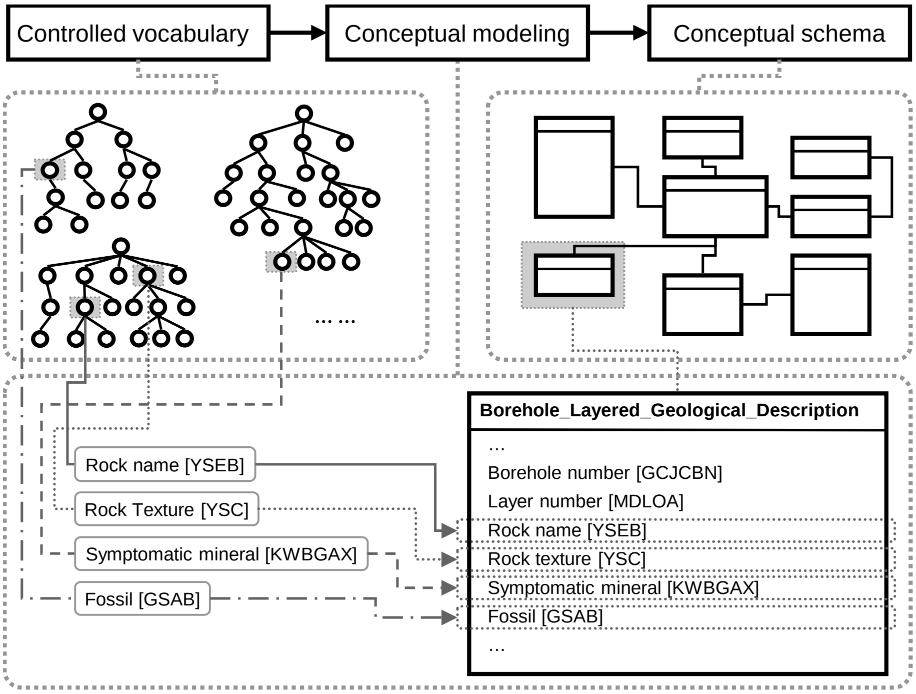


Fig. 2.5. Applying studied controlled vocabulary to build conceptual schemas of databases. Relationships between terms in the controlled vocabulary are different from those in the conceptual schemas of databases. A step called conceptual modeling is applied between the controlled vocabulary and the resulting conceptual schemas. Terms, codes and their definitions in the controlled vocabulary are used as building blocks to set up conceptual schemas for databases.

Due to the long history of mineral exploration conducted by the Zijin Mining Group, borehole data in the studied mining projects were stored in heterogeneous databases. For example, in one of these mining projects there are three borehole databases. In order to reconcile these three databases into an integrated geoscience database, a unified conceptual schema of borehole data was first set up by using the controlled vocabulary (Fig. 2.5).

Then entities and properties in each conceptual schema of the three databases were mapped to relevant entities and properties in the unified conceptual schema (Fig. 2.6).

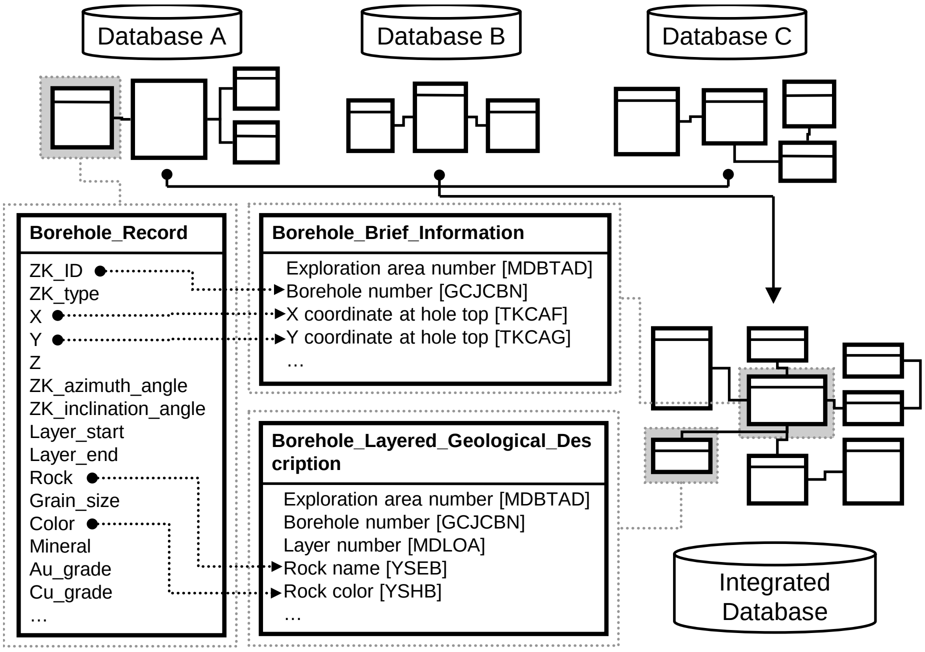
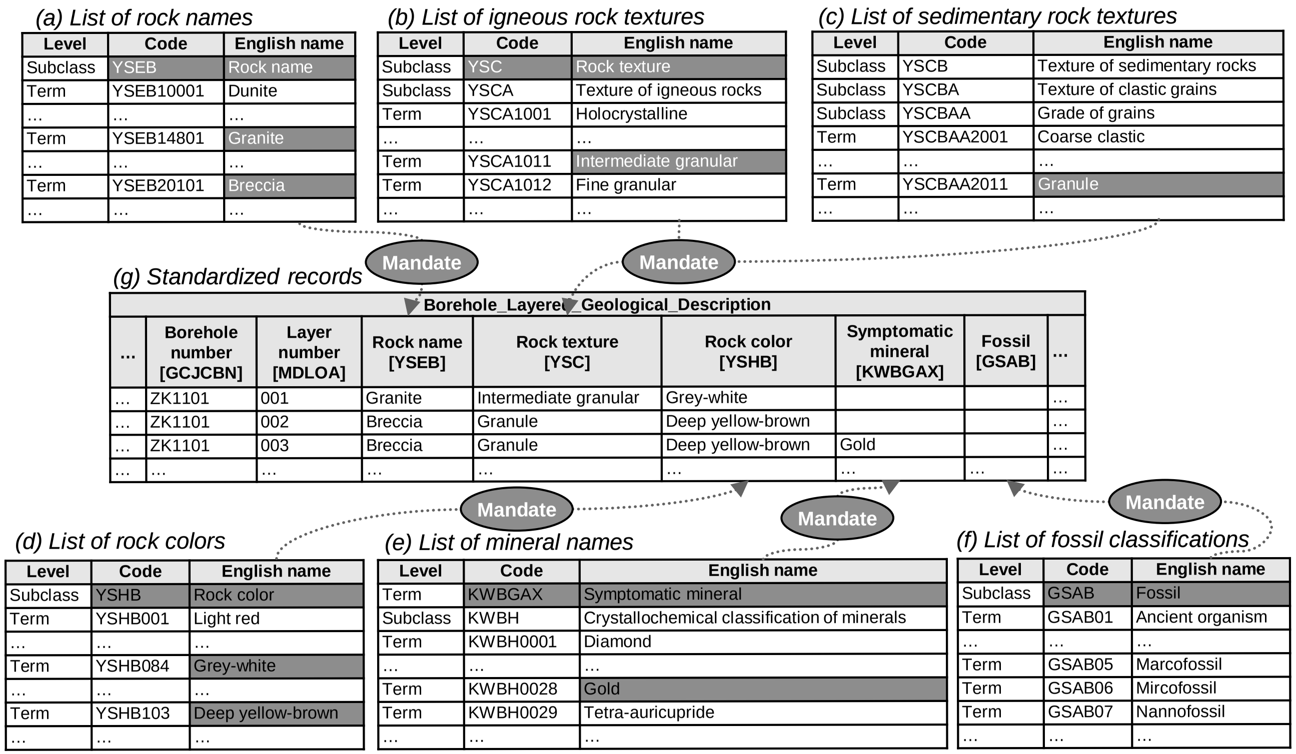


Fig. 2.6. Mapping diverse conceptual schemas of borehole data to a unified schema. Mapping between conceptual schema of database A and a unified schema of an integrated database is shown in partial detail.

Professional terms provided by the controlled vocabulary were also used as mandatory terms for borehole data in the integrated geoscience database (Fig. 2.7). Several computer programs were developed, by using C++ and SQL (Structured Query Language) languages, to support the transformation from heterogeneous records to standard terms. Most of these programs are based on systemically organized terms in the controlled vocabulary, such as terms of geological time scale, rock names and rock colors, etc. For example, a program was applied to transform records in column “Color” of table “Borehole_Record” in the original database A (Fig. 2.8a). This column was first connected to subclass “Rock color [YSHB]” of subject “Petrology [YS]” in the controlled vocabulary (Fig. 2.7d). Then, records in this column were respectively compared with standard terms in the subclass “Rock color [YSHB]” in order to find “abnormal” records (i.e., cannot find a same term in the controlled vocabulary). Once such a record (e.g., “Dark yellow-brown”) was found, a dialog box popped up indicating further operations, one of which is replacing this record with a standard term (e.g., “Deep yellow-brown”) chosen from the subclass “Rock color [YSHB]” in the controlled vocabulary. After confirmation, all other records of “Dark yellow-brown” in the column “Color” of table “Borehole_Record” were replaced by “Deep yellow-brown” (Fig. 2.8d).

Standardized borehole data underpinned by the controlled vocabulary improved applications that perform comprehensive processing of most or all borehole records in a mining project, such as mapping of borehole logs, modeling of ore bodies and estimation of mineral resources. A significantly improved application in the study presented in this chapter is the automatic borehole log mapping. A computer program was developed whereby contents, legends and layouts of borehole log maps can be edited as libraries and scripts by parametric methods (Auerbach and Schaeben, 1990; Liu et al., 1999). For example, a library of map symbols has been developed for different rocks. When a record in column “Rock name” was retrieved, the computer program found the relevant symbol in the library. Then that symbol was used to fill a cell in the map (Fig. 2.9). Standardized rock names are, therefore, essential for the automatic process of borehole log mapping.



(a) List of rock names

Level	Code	English name
Subclass	YSEB	Rock name
Term	YSEB10001	Dunite
…	…	…
Term	YSEB14801	Granite
…	…	…
Term	YSEB20101	Breccia
…	…	…

(b) List of igneous rock textures

Level	Code	English name
Subclass	YSC	Rock texture
Subclass	YSCA	Texture of igneous rocks
Term	YSCA1001	Holocrystalline
…	…	…
Term	YSCA1011	Intermediate granular
Term	YSCA1012	Fine granular
…	…	…

(c) List of sedimentary rock textures

Level	Code	English name
Subclass	YSCB	Texture of sedimentary rocks
Subclass	YSCBA	Texture of clastic grains
Subclass	YSCBAA	Grade of grains
Term	YSCBAA2001	Coarse clastic
…	…	…
Term	YSCBAA2011	Granule
…	…	…

(g) Standardized records

Borehole_Layered_Geological_Description								
…	Borehole number [GCJCBN]	Layer number [MDLOA]	Rock name [YSEB]	Rock texture [YSC]	Rock color [YSHB]	Symptomatic mineral [KWBGAX]	Fossil [GSAB]	…
…	ZK1101	001	Granite	Intermediate granular	Grey-white			…
…	ZK1101	002	Breccia	Granule	Deep yellow-brown			…
…	ZK1101	003	Breccia	Granule	Deep yellow-brown	Gold		…
…	…	…	…	…	…	…	…	…

(d) List of rock colors

Level	Code	English name
Subclass	YSHB	Rock color
Term	YSHB001	Light red
…	…	…
Term	YSHB084	Grey-white
…	…	…
Term	YSHB103	Deep yellow-brown
…	…	…

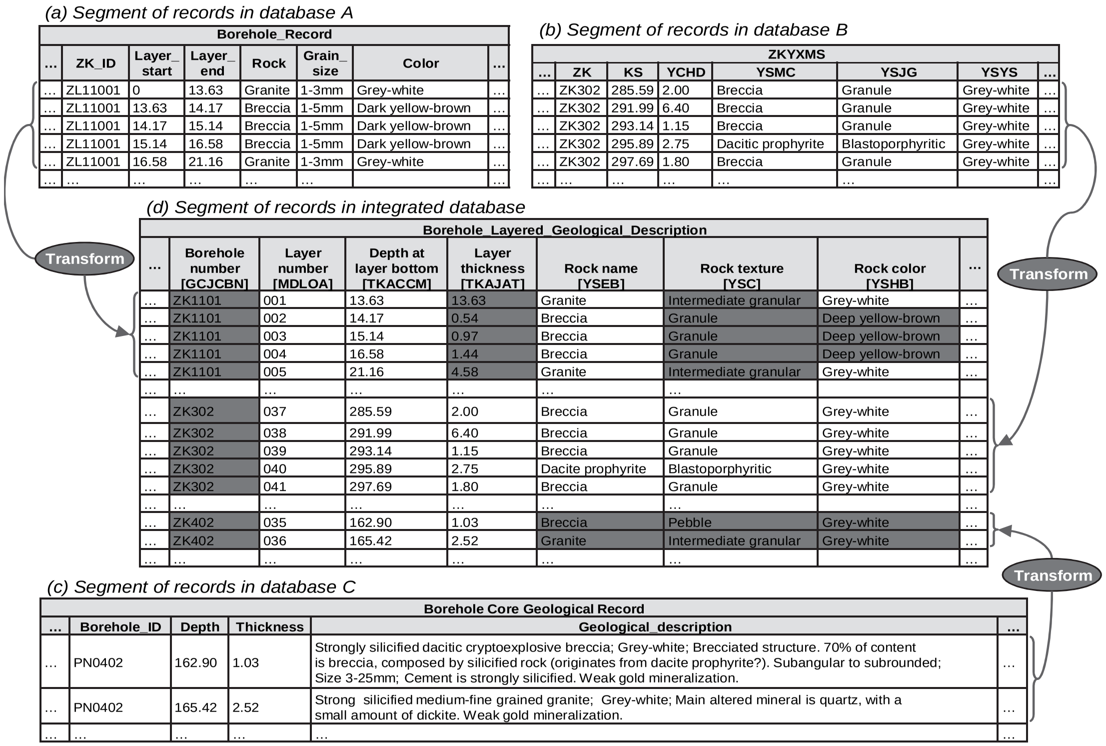
(e) List of mineral names

Level	Code	English name
Term	KWBGAX	Symptomatic mineral
Subclass	KWBH	Crystallochemical classification of minerals
Term	KWBH0001	Diamond
…	…	…
Term	KWBH0028	Gold
Term	KWBH0029	Tetra-auricupride
…	…	…

(f) List of fossil classifications

Level	Code	English name
Subclass	GSAB	Fossil
Term	GSAB01	Ancient organism
…	…	…
Term	GSAB05	Marcofossil
Term	GSAB06	Mircofossil
Term	GSAB07	Nannofossil
…	…	…

Fig. 2.7. Using studied controlled vocabulary to mandate standard terms and codes. Tables (a)–(f) are segments of terms in studied controlled vocabulary. Table (g) is a segment of standardized borehole data in an integrated mine-scale database. In (a)–(f), cells filled with dark color indicate codes and names of subclasses and concepts used as column headings or records in (g).



(a) Segment of records in database A

Borehole_Record							
…	ZK_ID	Layer_ start	Layer_ end	Rock	Grain_ size	Color	…
…	ZL11001	0	13.63	Granite	1-3mm	Grey-white	…
…	ZL11001	13.63	14.17	Breccia	1-5mm	Dark yellow-brown	…
…	ZL11001	14.17	15.14	Breccia	1-5mm	Dark yellow-brown	…
…	ZL11001	15.14	16.58	Breccia	1-5mm	Dark yellow-brown	…
…	ZL11001	16.58	21.16	Granite	1-3mm	Grey-white	…
…	…	…	…	…	…		…

(b) Segment of records in database B

ZKYXMS							
…	ZK	KS	YCHD	YSMC	YSJG	YSYS	…
…	ZK302	285.59	2.00	Breccia	Granule	Grey-white	…
…	ZK302	291.99	6.40	Breccia	Granule	Grey-white	…
…	ZK302	293.14	1.15	Breccia	Granule	Grey-white	…
…	ZK302	295.89	2.75	Dacitic prophyrite	Blastoporphyritic	Grey-white	…
…	ZK302	297.69	1.80	Breccia	Granule	Grey-white	…
…	…	…	…	…	…		…

(d) Segment of records in integrated database

Borehole_Layered_Geological_Description								
…	Borehole number [GCJCBN]	Layer number [MDLOA]	Depth at layer bottom [TKACCM]	Layer thickness [TKAJAT]	Rock name [YSEB]	Rock texture [YSC]	Rock color [YSHB]	…
…	ZK1101	001	13.63	13.63	Granite	Intermediate granular	Grey-white	…
…	ZK1101	002	14.17	0.54	Breccia	Granule	Deep yellow-brown	…
…	ZK1101	003	15.14	0.97	Breccia	Granule	Deep yellow-brown	…
…	ZK1101	004	16.58	1.44	Breccia	Granule	Deep yellow-brown	…
…	ZK1101	005	21.16	4.58	Granite	Intermediate granular	Grey-white	…
…	…	…	…	…	…	…		
…	ZK302	037	285.59	2.00	Breccia	Granule	Grey-white	…
…	ZK302	038	291.99	6.40	Breccia	Granule	Grey-white	…
…	ZK302	039	293.14	1.15	Breccia	Granule	Grey-white	…
…	ZK302	040	295.89	2.75	Dacite prophyrite	Blastoporphyritic	Grey-white	…
…	ZK302	041	297.69	1.80	Breccia	Granule	Grey-white	…
…	…	…	…	…	…	…		…
…	ZK402	035	162.90	1.03	Breccia	Pebble	Grey-white	…
…	ZK402	036	165.42	2.52	Granite	Intermediate granular	Grey-white	…
…	…	…	…	…	…	…		…

(c) Segment of records in database C

Borehole Core Geological Record					
…	Borehole_ID	Depth	Thickness	Geological_description	…
…	PN0402	162.90	1.03	Strongly silicified dacitic cryptoexplosive breccia; Grey-white; Brecciated structure. 70% of content is breccia, composed by silicified rock (originates from dacite prophyrite?). Subangular to subrounded; Size 3-25mm; Cement is strongly silicified. Weak gold mineralization.	…
…	PN0402	165.42	2.52	Strong silicified medium-fine grained granite; Grey-white; Main altered mineral is quartz, with a small amount of dickite. Weak gold mineralization.	…
…	…	…	…	…	…

Fig. 2.8. Transforming heterogeneous terms to standard terms provided by studied controlled vocabulary. Tables (a), (b) and (c) are segments of records retrieved from original databases A, B and C, respectively. Table (d) is a segment of records in an integrated mine-scale database. In (d), cells filled with dark color indicate standardized records after a mandated transformation process.

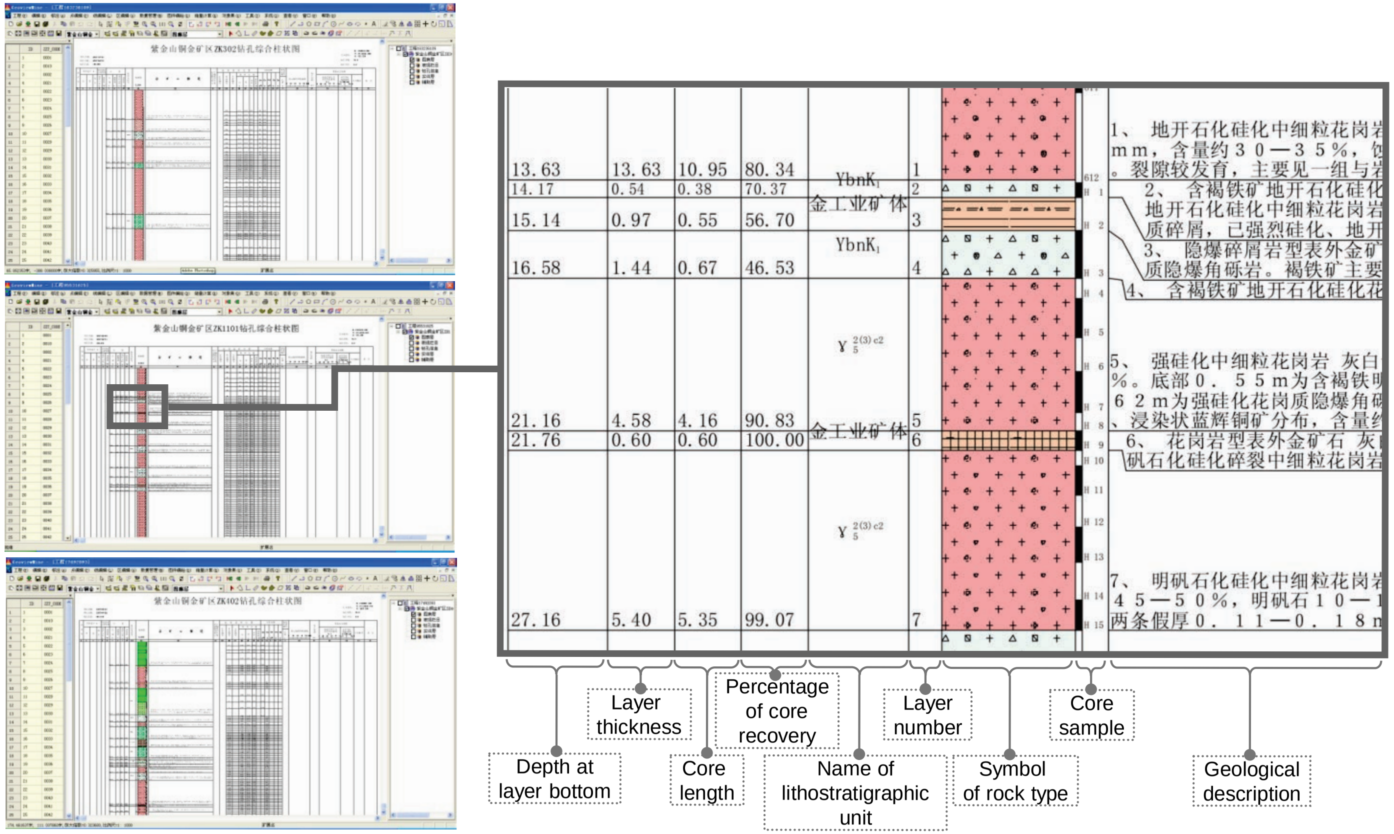


Fig. 2.9. Automatically generated borehole log maps with consistent symbols and terms based on standardized borehole data. Map of borehole “ZK1101” is shown in partial detail.

Following the aforementioned procedure of standardization, integration and application for mining project geodata, three mining projects of the Zijin Mining Group in different locations have reconciled their heterogeneous borehole data into integrated geoscience databases. As a part of the database used for estimation of mineral resources, borehole data of each mining project were forwarded to the institute of geological and mineral exploration of the Zijin Mining Group, which estimates mineral resources for its individual mining projects. The controlled vocabulary-driven borehole data were welcomed by geologists at this institute. Data and mineral resources estimates in any mining project are subsequently forwarded to the mineral resources and reserves evaluation center of the Ministry of Land and Resources of China for checking and confirmation. The standardized borehole data from these mining projects of Zijin Mining Group also obtained positive comments at this center.

2.4 Discussion

The focus of the study presented in this chapter is on the interoperability of mineral exploration geodata of local contexts (i.e., mining projects of a mining group). The methods applied here for developing a taxonomical controlled vocabulary for the knowledge domain of mineral exploration for mining applications resulted in or improved the interoperability of multi-source mining project geodata. The study presented in this chapter shows that a properly organized controlled vocabulary is not only efficient for reconciling heterogeneous geodata sources within a mining project, but is also helpful in making geodata of individual mining projects interoperable with other applications in the knowledge domain of mineral exploration. Promotions of the results of this study by the headquarters of the mining group helped in convincing managers of its mining projects to adopt the controlled vocabulary as a common platform for building integrated geoscience databases. However in a general and practical sense, it is hard to convince different institutions to adopt a unified controlled vocabulary and replace their customary ones. Therefore, besides methodological works (e.g., concise organization structure, hierarchical encoding and extensible structure, etc.), negotiation and consensus among various institutions are also necessary to promote the wider acceptability and interoperability of a controlled vocabulary in practical works. A primary reason for adopting and adapting professional standards in the controlled vocabulary described here is that these standards are results of negotiations and collaborations on certain topics in the larger knowledge domain of geology and mineral resources, and thus they have been widely accepted and used.

The knowledge domain of mineral exploration for mining applications is a synthesis of diverse subjects and concepts. The controlled vocabulary discussed here does not represent precisely relationships between these subjects and concepts as what a formal ontology does. However, the presented hierarchical organization structure of subjects, subclasses and concepts provides a simple but concise representation for this knowledge domain. In the aforementioned case study of mining project geoscience databases, researchers could easily retrieve terms from the controlled vocabulary in order to reconcile various geodata and build new databases for their studies in mining applications. The coding method provides an even simpler representation of concepts and their inter-relationships. Codes are also a link between multi-lingual terms in the controlled vocabulary, and have been used as field names in mining databases in order to avoid errors caused by Chinese field names and speed up database queries. The definition schema defines a concept, subclass or subject in the controlled vocabulary as a field in a database. In the case study, the definition schema has been proved useful for modeling conceptual schemas of databases. Thus, the organization structure, coding method and definition schema make the controlled vocabulary efficient in reconciling various geodata sources in a mining project (Fig. 2.10).

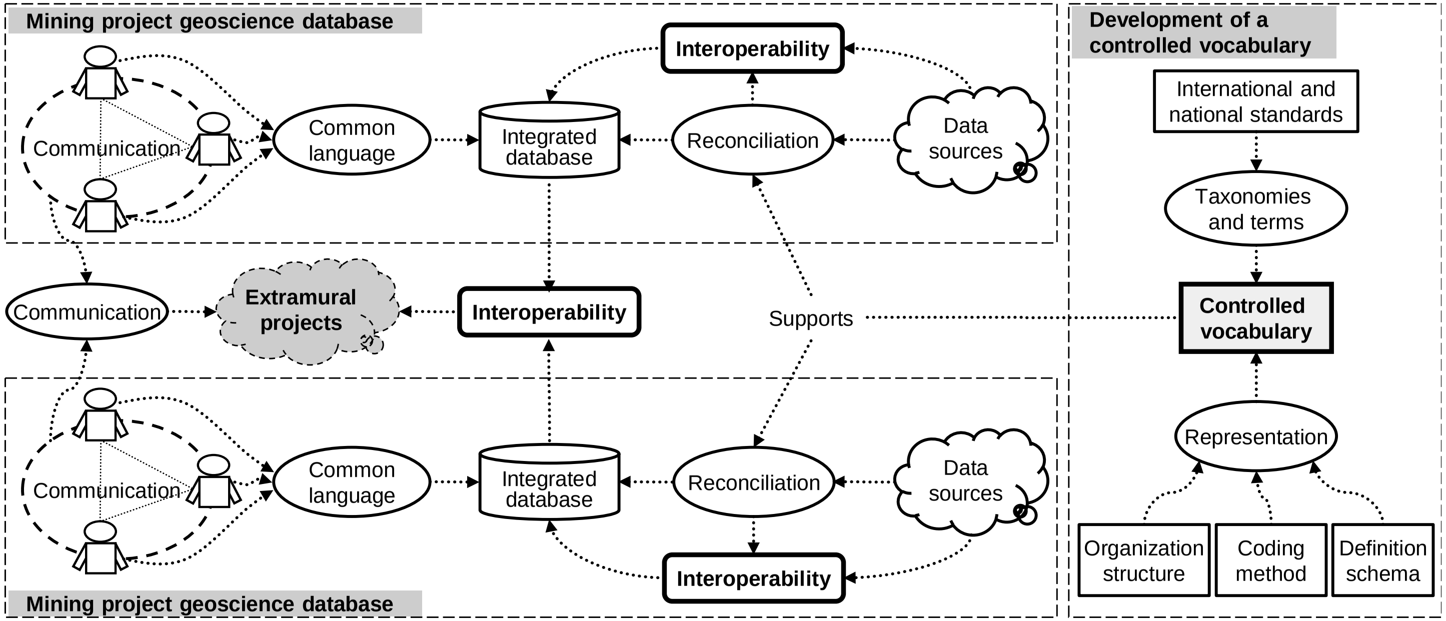


Fig. 2.10. Using a controlled vocabulary to reconcile/standardize multi-source geodata at mining projects and to improve their interoperability with extramural projects.

The extensible structure and use of professional standards in the controlled vocabulary improve the interoperability of reconciled geodata for mining purposes (Fig. 2.10). The extensible structure makes the controlled vocabulary continuously growing by addition of new terms that evolve from different mining projects. In this way, the controlled vocabulary became a common platform of professional terms among different mining projects in the aforementioned case study. National standards were adopted and adapted in the controlled vocabulary, because these standards have high credibility for general geological work in China (Wu et al., 2005). By using taxonomies and terms from standards, the study presented in this chapter did not actually “re-invent the wheel”, but developed a controlled vocabulary that is more credible and acceptable for databases of mining projects. Geodata standardized by a controlled vocabulary enable concerned researchers in a mining project to access these data with a “common language”, thus improving their communication with each other (Fig. 2.10). Reports and maps of mineral resources estimates based on standardized mining project geodata were also welcomed by researchers outside the mining projects, because they can easily read and understand professional terms derived from standards and use the standardized data according to their needs. In this way, data consumers and providers do not need to negotiate the concepts and the contents of data models, while they are communicating with each other (cf. Richard and Soller, 2008).

Controlled vocabularies have been studied in many different research projects related to sharing and integration of geodata, such as GEON (Ludäscher et al., 2003), NGMDB (Richard et al., 2003; Soller and Berg, 2005) and OneGeology (Jackson, 2007; CGI-IUGS, 2008). A similarity between these studies and the study described in this chapter is the objective of using controlled vocabularies or ontologies to improve the standardization and semantic interoperability of geodata. Nevertheless, there are also differences in knowledge domains, organization structures and coding methods, etc. One significant difference is that the controlled vocabulary in this study does not involve XML-based encoding, whereas controlled vocabularies in many other studies are edited with XML-based languages, in order to support web-based services and applications (cf. Soergel et al., 2004; Richard and Soller, 2008). Considering that XML-based controlled vocabularies are more compatible with web-based applications and can be more easily transferred to description logic-based formal ontologies (cf. Raskin and Pan, 2005; Brodaric and Probst, 2009), one of the further studies is to re-edit the controlled vocabulary with an XML-based language, such as SKOS (van Assem et al., 2006), so as to enable the controlled vocabulary to support web-based services of mining project geodata.

Hierarchical organization and encoding of terms have been studied in other research projects, such as the AGROVOC (FAO, 2010) and the UNSPSC (UNSPSC, 2004). A similarity between the study described in this chapter and those two studies is that relationships between terms are initially defined as “has broader meaning than”, “has narrower meaning than”, or “is related to” (Fig. 2.3). Whereas AGROVOC does not provide codes in the thesaurus, UNSPSC provides hierarchical codes for every term. Whereas only numeric codes are used for terms in the UNSPSC thesaurus, both alphabetic and alphanumeric codes are used for terms in the controlled vocabulary developed in this study. A notable similarity between the UNSPSC thesaurus and the controlled vocabulary in this study is that, although every concept, subclass or subject can have different names in different languages, every concept, subclass or subject is given a single code. This means that a code is a potentially unique representation of a concept expressed in different languages. For example, by using an encoded multi-lingual controlled vocabulary, codes can be stored in columns (e.g., “Rock name [YSEB]” and “Rock color [YSHB]” in Fig. 2.8d) of a database, and records can be shown in a preferred language in user interfaces. Consequently, geodata underpinned by a multi-lingual controlled vocabulary can be translated from one language to another. This is a useful function when data sharing is carried out internationally (Gravesteijn and Rassam, 1990). Nevertheless, a previous study on AGROVOC (Soergel et al., 2004) shows that, for well-defined semantics, relationships between terms should be enriched in order to transform a thesaurus into an ontology. This is another direction to improve the controlled vocabulary in further studies.

Mapping and transforming of geodata from a local terminological system to the controlled vocabulary inevitably causes information loss. For example, a local term may be transferred to a term in the controlled vocabulary, but the meaning of both terms can be slightly different. Or, only a term with a broader meaning can be found in the controlled vocabulary for a local term. In order to reduce such information loss, researchers in mining projects in the case study received trainings of both terminological systems before they operated the mapping and transformation of mining project geodata. Another way reducing information loss was to keep many local terms unchanged in the transformed geodata, while adding these terms into the controlled vocabulary. This was possible when no terms with equivalent or similar meanings could be found in the controlled vocabulary.

2.5 Conclusions

By extrapolation, it can be concluded that a properly organized controlled vocabulary allows for efficient reconciliation of heterogeneous and multi-source geological data in local contexts and makes those geological data interoperable with extramural applications in the same knowledge domain. In order to achieve this purpose, it is necessary that a controlled vocabulary provides a concise structure for representing and organizing concepts and their inter-relationships in a knowledge domain. Moreover, negotiations and collaborations among stakeholders in the same or related knowledge domains can be helpful in promoting wider acceptability and interoperability of a controlled vocabulary as well as the geological data underpinned by it. Nevertheless, when the geological data interoperability issues are discussed on regional/global scales, an often-encountered challenge is the multilinguality of geological data provided by different data sources. Chapter 3 will elaborate on this issue and describe approaches for alleviating linguistic barriers of online geological data.

Chapter 3

A multilingual thesaurus for interoperability of online geological data

This chapter is based on: Ma, X., Carranza, E.J.M., Wu, C., van der Meer, F.D., Liu, G., 2011. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps. Computers & Geosciences 37 (10), 1602–1615.

3.1 Introduction

Regional/global cooperation is a trend of scientific works in the field of geology, whereas challenges caused by multilinguality of geological data (geodata) arise in many of these works. Linguistic barrier is a long-term challenge for the interoperability of geodata and retrieval of geoinformation (Lloyd, 1973; Gravesteijn and Rassam, 1990; Asch and Jackson, 2006; Laxton et al., 2010). Users of geological maps have been facing that challenge since this type of geodata has evolved. With increasing internationalization and globalization of geological scientific and technological works (e.g., de Mulder et al., 2006; Jackson, 2007), overcoming that challenge has become an important issue in sharing of geodata and/or geoinformation. Most geological maps are produced by governmental organizations; thus, they are encoded in official languages of their producers. If users cannot read the languages of a geological map, then it is hard for them either to understand the meaning of that map or to use that map efficiently. Recently, some digital geological maps have been published in bilingual formats (e.g., the 1: 200,000 Geological Map of Japan published in Japanese and English (GSJ-AIST, 2009)) and multilingual formats (e.g., the 1:5,500,000 Geological Map of South America published in Spanish, Portuguese and English (CGMW et al., 2003)) to alleviate linguistic barriers for international users. However, the number of languages used in these maps is still limited and several other geological maps remain in monolingual formats. Consequently, the interoperability of most geological maps is precluded or hindered.

Since the past decades, researchers coordinated by the CGI-IUGS[8] and its predecessors have been attempting to alleviate linguistic barriers of geological maps by developing multilingual geoscience thesauri. Earlier outputs of their works include the published 1st and 2nd editions of Multilingual Thesaurus of Geosciences (or MTG) (Rassam et al., 1988; Gravesteijn et al., 1995). The 2nd edition includes 5823 terms in English (as the basic reference), French, German, Italian, Russian and Spanish[9]. Another recently published output is the Asian Multilingual Thesaurus of Geosciences (or AMTG) (CCOP and CIFEG, 2006), which includes 5867 terms in English (as the basic reference), Khmer, Chinese, French, Indonesian, Japanese, Korean, Lao, Malaysian, Thai and Vietnamese[10]. These thesauri help users understand and use geological maps in foreign languages. However, the MTG contains some geoscience terms that are "inconsistent, incomplete and inaccurate" (Asch and Jackson, 2006), and so its applications are limited. The newer AMTG also contains some "inconsistent, incomplete and inaccurate" terms, and its applications have not been fully demonstrated yet. Thus, it is worthwhile to further study how to make more useful multilingual geoscience thesauri.

Rapidly evolving web technologies pave the way for development of platforms for sharing geological maps to the international community, and for developing and applying multilingual geoscience thesauri. The OGC® web service standards (e.g., WMS[11], WFS[12] and WCS[13], etc.) enable the flow of geodata more open and faster through the World Wide Web (Peng and Tsou, 2003). By using these web services, organizations or individuals can publish geological maps online easily. For example, through the OneGeology project (Jackson, 2007; Jackson and Wyborn, 2008), 116 countries have agreed to share geological maps by the middle of 2010, and 50 of them have already provided WMS or WFS of their national or regional geological maps[14].

---

[8] Commission for the Management and Application of Geoscience Information of the International Union of Geological Sciences. http://www.cgi-iugs.org [Accessed February 07, 2011].

[9] The online version of MTG also includes Finnish and Swedish. http://en.gtk.fi/Geoinfo/Library/multhes.html [Accessed February 07, 2011].

[10] http://www.ccop.or.th/download/pub/AMTG_2006.pdf [Accessed February 07, 2011].

[11] Web Map Service. http://www.opengeospatial.org/standards/wms [Accessed February 07, 2011].

[12] Web Feature Service. http://www.opengeospatial.org/standards/wfs [Accessed February 07, 2011].

[13] Web Coverage Service. http://www.opengeospatial.org/standards/wcs [Accessed February 07, 2011].

[14] http://www.onegeology.org/participants/app/1gCountries.cfc?method=viewCountryStatus [Accessed February 07, 2011].

Meanwhile, extensive studies related to the W3C®-proposed Semantic Web[15] have been addressing the essentiality of ontologies for formal and common representations of subject domain knowledge (e.g., Davies et al., 2003; Antoniou et al., 2005; Garcia-Sanchez et al., 2009). Developments of geoscience thesauri, as basic elements for building geoscience ontologies and representing geoscience knowledge, have increasingly become one of the foci of studies in the context of the Semantic Web (e.g., Raskin and Pan, 2005; Deliiska, 2007; Smits and Friis-Christensen, 2007; Buccella et al., 2009).

Researchers in the Geoscience Concept Definitions Task Group[16] of the CGI-IUGS are currently working on multilingual geoscience thesauri with the Simple Knowledge Organization System (SKOS)[17], a standard recommended by W3C®. This effort is consistent with works of MTG and AMTG, and it aims to make significant improvements to the online applications of geoscience thesauri. The MTG and AMTG classify geoscience terms by subject domains, but terms classified into each subject domain are arranged alphabetically without definitions. The current work of the Geoscience Concept Definitions Task Group is compatible with the Semantic Web and has great potential in applications with online geological maps, such as those in the OneGeology project. Although impressive progress has been made by now, the work on SKOS-based geoscience thesauri by the Geoscience Concept Definitions Task Group of the CGI-IUGS is still ongoing, and methods for developing SKOS-based geoscience thesauri still require further practical testing and discussion. Meanwhile, online services and/or applications based on SKOS-based geoscience thesauri are still rare.

The study presented in this chapter aims to develop a SKOS-based multilingual thesaurus of geological time scale (MLTGTS) for alleviating linguistic barriers of geological time scale (GTS) records among online geological maps. The contributions of this study are three-fold. First, to extend the SKOS model to build a more semantically-expressive structure for the subject domain of GTS. This would motivate building thesauri of other subject domains in geosciences. Second, to maintain a MLTGTS and use it with developed JavaScript programs to recognize and translate GTS terms in online geological maps. The approach of characteristic-oriented term retrieval implemented in the JavaScript programs is effective for recognizing GTS terms from records in geological maps. Third, to package the first and second contributions into a novel methodology for improving the interoperability of

---

[15] http://www.w3.org/standards/semanticweb [Accessed February 07, 2011].

[16] https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/ ConceptDefinitionsTG [Accessed February 07, 2011].

[17] http://www.w3.org/2004/02/skos [Accessed February 07, 2011].

online geological maps in the context of the Semantic Web. With functions for online recognition and translation of geoscience terms, massive monolingual geological maps can be published online directly and users can access and use them although they cannot read their original languages.

3.2 SKOS-based multilingual thesaurus of geological time scale

3.2.1 Addressing the insufficiency of SKOS in the context of the Semantic Web

By using ontologies in the Semantic Web, meanings of concepts and relationships between concepts are made accessible as the material in which certain concepts appear (Berners-Lee et al., 2001). This paradigm is also supported by recent studies related to the Geospatial Semantic Web[18] (Bishr, 2006; Yue et al., 2009; Zhang et al., 2010). Ontologies in computer science are valuable functions because they are derived from shared conceptualizations of domain knowledge (Gruber, 1995). Thesauri are regarded as a necessary foundation for building ontologies in computer science (Gruber, 1995; Guarino, 1997b, 1998). Professional (e.g., geoscience) terms in a thesaurus may refer to the same real-world features in a subject domain, as what an ontology does. However, unlike a precise conceptualization (i.e., detailed semantics) in an ontology, a thesaurus is simpler in definitions of meanings and relationships of terms (i.e., concise semantics) and, thus, it leads to a simple organizational structure (Gilchrist, 2003). For example, the MTG and AMTG arrange geoscience terms alphabetically, and each term is tagged with a label indicating its subject domain in geosciences.

To promote functions for indexing and navigating resources on the Web, it would be useful to encode thesauri in Web-compatible formats. Similar to OWL's[19] role in editing ontologies, the SKOS can be used for encoding thesauri in the context of the Semantic Web. SKOS is a common data model based on the RDF[20], which in turn is a standard recommended by W3C®. Compared to the flexible uses of RDF or OWL, SKOS provides a pre-defined concise structure for conceptualize a domain of discourse, which are

---

[18] http://www.opengeospatial.org/projects/initiatives/gswie [Accessed February 07, 2011].

[19] Web Ontology Language. http://www.w3.org/TR/owl-guide [Accessed February 07, 2011].

[20] Resource Description Framework. http://www.w3.org/RDF [Accessed February 07, 2011].

specifically defined for building thesauri. In the SKOS model (Table 3.1) [21], there are pre-defined object properties for defining relationships between concepts and datatype properties for defining differentiating attributes (or qualities) of concepts. These properties let users set up hierarchical and associative relationships between terms within a thesaurus, and assign essential attributes (e.g., multilingual labels) to each term (e.g., Pastor-Sanchez et al., 2009). For example, Fig. 3.1 shows a GTS concept "Lower_Triassic" defined with the pure SKOS model.

Table 3.1 Object and datatype properties in the SKOS model

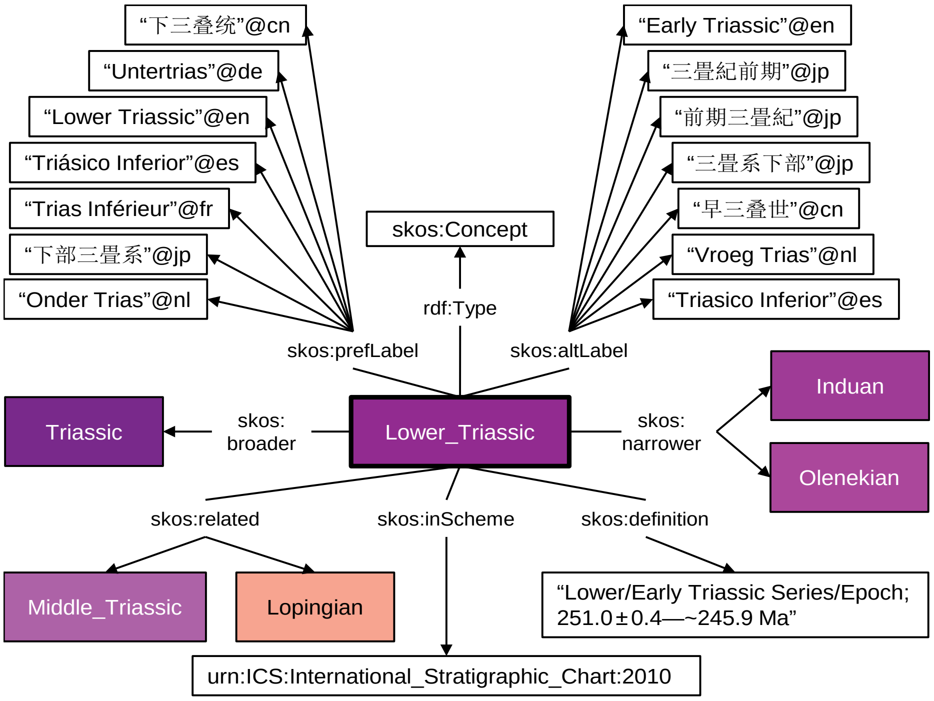
Object property	**Meaning**	**Datatype property**	**Meaning**
`skos:broadMatch`	has broader match	`skos:altLabel`	alternative label
`skos:broader`	has broader	`skos:changeNote`	change note
`skos:broaderTransitive`	has broader transitive	`skos:definition`	definition
`skos:closeMatch`	has close match	`skos:editorialNote`	editorial note
`skos:exactMatch`	has exact match	`skos:example`	example
`skos:hasTopConcept`	has top concept	`skos:hiddenLabel`	hidden label
`skos:inScheme`	is in scheme	`skos:historyNote`	history note
`skos:mappingRelation`	is in mapping relation with	`skos:notation`	notation
`skos:member`	has member	`skos:note`	note
`skos:memberList`	has member list	`skos:prefLabel`	preferred label
`skos:narrowMatch`	has narrower match	`skos:scopeNote`	scope note
`skos:narrower`	has narrower		
`skos:narrowerTransitive`	has narrower transitive		
`skos:related`	has related		
`skos:relatedMatch`	has related match		
`skos:semanticRelation`	is in semantic relation with		
`skos:topConceptOf`	is top concept in scheme		

---

[21] SKOS Reference. W3C Recommendation 18 August 2009. http://www.w3.org/TR/2009/REC-skos-reference-20090818 [Accessed February 07, 2011].

```
<skos:Concept rdf:ID="Lower_Triassic">
    <skos:definition>Lower/Early Triassic Series/Epoch; 251.0±0.4—~245.9 Ma</skos:definition>
    <skos:prefLabel xml:lang="cn">下三叠统</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Untertrias</skos:prefLabel>
    <skos:prefLabel xml:lang="en">Lower Triassic</skos:prefLabel>
    <skos:prefLabel xml:lang="es">Triásico Inferior</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">Trias Inférieur</skos:prefLabel>
    <skos:prefLabel xml:lang="jp">下部三畳系</skos:prefLabel>
    <skos:prefLabel xml:lang="nl">Onder Trias</skos:prefLabel>
    <skos:altLabel xml:lang="en">Early Triassic</skos:altLabel>
    <skos:altLabel xml:lang="jp">三畳紀前期</skos:altLabel>
    <skos:altLabel xml:lang="jp">前期三畳紀</skos:altLabel>
    <skos:altLabel xml:lang="jp">三畳系下部</skos:altLabel>
    <skos:altLabel xml:lang="cn">早三叠世</skos:altLabel>
    <skos:altLabel xml:lang="nl">Vroeg Trias</skos:altLabel>
    <skos:altLabel xml:lang="es">Triasico Inferior</skos:altLabel>
    <skos:broader rdf:resource="#Triassic"/>
    <skos:narrower rdf:resource="#Olenekian"/>
    <skos:narrower rdf:resource="#Induan"/>
    <skos:related rdf:resource="#Middle_Triassic"/>
    <skos:related rdf:resource="#Lopingian"/>
    <skos:inScheme rdf:resource="urn:ICS:International_Stratigraphic_Chart:2010"/>
</skos:Concept>
```

(a) Source code.



(b) Graphic view of (a).

Fig. 3.1. Definition of "Lower_Triassic" as a GTS concept with object and datatype properties of SKOS model.

Despite those features, the SKOS model is insufficient for encoding certain semantics in a thesaurus (cf. Tennis and Sutton, 2008). Because GTS is an ordinal hierarchical scheme divided by time boundaries (Cox and Richard, 2005), a pure SKOS model cannot properly represent this core feature of GTS. For example, "Lopingian" rocks are older than "Lower_Triassic" rocks, which are, in turn, older than "Middle_Triassic" rocks. These ordinal relationships cannot be represented properly with the "skos:related" property (Fig. 3.1). To address this problem, the SKOS model was extended by adding some user-defined properties and some pre-defined properties in RDF (cf. Rector et al., 2004; Pan and Horrocks, 2007; Jupp et al., 2008). Multilingual GTS terms were collected, and then encoded with this extended SKOS model.

3.2.2 Addressing semantics and syntax / lexicon in multilingual GTS terms

Like studies on interoperability of multisource geodata and/or geoinformation (Bishr, 1998; Brodaric and Gahegan, 2006; Ludäscher et al., 2006), collecting multilingual terms of a subject domain in geosciences also involves semantic and syntactic/lexical issues. Semantics deals with meanings of terms whereas syntax/lexicon deals with words and structures of expressions in each language.

The first challenge is addressing semantics. If the meanings of several terms in different languages are the same, then they are semantically matched and they can be registered as entries in a multilingual thesaurus. For a certain geoscience concept, if there are no semantically matched terms in different languages (i.e., one cannot find multilingual terms describing exactly the same thing or falling exactly into the same inter-relationships), then it is difficult to register a full entry in a multilingual thesaurus for this concept. The meanings of geoscience concepts, in general, are defined by international commissions of different subject domains in geosciences. In the study of the MLTGTS, global boundaries of geological time are defined by the ICS[22] and the International Stratigraphic Chart compiled by ICS is globally accepted and used by the international geoscience community (cf. Ogg, 2009; Walker et al., 2009; U.S. Geological Survey Geologic Names Committee, 2010). These formed a stable basis for collecting semantically matched GTS terms in different languages.

The second challenge is addressing syntax/lexicon. For semantically matched terms in different languages for the same concept, there may be several

---

[22] International Commission on Stratigraphy. http://www.stratigraphy.org [Accessed February 07, 2011].

synonyms describing the same concept in every language. It is not wrong to use synonyms in one language to record geodata as long as users can read them and understand their meanings in that language. However, for a multilingual geoscience thesaurus, many synonyms in different languages should be collected as much as possible so that they are all recognized when that thesaurus is used by a computer. For example, “Cainozoic” is a synonym of “Cenozoic” in English; “Paleogeno” is a synonym of “Paleógeno” in Spanish; “Quartaer” is a synonym of “Quartär” in German; and “ジュラ紀前期” is a synonym of “前期ジュラ紀” in Japanese, etc. Several of such synonyms in different languages were collected for the MLTGTS discussed here.

Related to the semantic and syntactic/lexical issues addressed in collecting multilingual GTS terms, there are two approaches commonly applied to match multilingual terms (Miles et al., 2001): (1) interlingual mapping or (2) multilingual labeling. The first approach can be used to address the lack of semantically matched multilingual terms. For instance, consider at least two independent monolingual thesauri covering the same or similar subject domain but with different hierarchical and associative relationships. Mappings between terms in each pair of thesauri can be performed via the first approach, but such mappings are time-consuming and, sometimes, even impossible. In contrast, the second approach deals with terms in different languages with the same conceptual structure (i.e., terms that have already been semantically matched). Thus, the second approach can be used to arrange multilingual terms in a thesaurus.

The multilingual labeling approach was applied in developing the MLTGTS because boundaries in the GTS proposed by ICS are accepted globally as a common conceptual schema in this subject domain. Standard GTS terms in seven languages (i.e., English, Dutch, German, Spanish, French, Chinese and Japanese) were collected by referring to the MTG and AMTG. However, some GTS terms are (a) not available in MTG and AMTG (e.g., terms at levels of “Series/Epoch” and “Stage/Age” in “Permian” and “Silurian”), (b) out of date (e.g., the Chinese term “晚第三纪” and Japanese term “新第三紀” of “Neogene” in AMTG). In addition, some GTS terms in the AMTG are mismatched (e.g., the Chinese geochronologic term “早泥盆世” (“Early Devonian Epoch”) is mismatched with the Japanese chronostratigraphic term “下部デボン系” (“Lower Devonian Series”) in the entry with English term “Lower Devonian” as the basic reference). In this regard and to make the collection of multilingual GTS terms complete and up-to-date, websites of geological institutions of different countries were searched and a multilingual GTS thesaurus[23] recently edited by the Geoscience Concept Definitions Task Group of the CGI-IUGS was referred.

Moreover, the two nomenclature systems for GTS terms were considered – one for chronostratigraphy (i.e., Eonothem, Erathem, System, Series and Stage) and the other for geochronology (i.e., Eon, Era, Period, Epoch and Age). In some western languages (e.g., English or Spanish) wherein the basic terms are the same, chronostratigraphic and geochronologic terms are often indistinct in actual applications, but in some other languages (e.g., Chinese or Japanese) GTS terms include units by which chronostratigraphic terms are distinguished from geochronologic terms. For example, in Chinese, the chronostratigraphic term "泥盆统" ("Devonian Series") corresponds to but is distinct from the geochronologic term "泥盆世" ("Devonian Epoch"). Another concern from the twofold nomenclature relates to GTS terms containing "Upper/Late" and "Lower/Early" at the level of "Series/Epoch". GTS terms containing "Upper" and "Lower" were originally proposed for chronostratigraphy, whereas terms containing "Late" and "Early" are for geochronology (Haile, 1987; U.S. Geological Survey Geologic Names Committee, 2010). In many actual works, "Upper" is equated to "Late" and "Lower" to "Early" and are used inter-changeably, causing confusions for other workers. To improve the semantic precision of the developed MLTGTS, in the study presented in this chapter terms containing "Upper" or "Lower" (e.g., "Upper Cretaceous", "Lower Triassic", etc.) are regarded as chronostratigraphic terms and, correspondingly, terms containing "Late" or "Early" (e.g., "Late Cretaceous", "Early Triassic", etc.) as geochronologic terms.

3.2.3 Extending SKOS-model to capture GTS structure

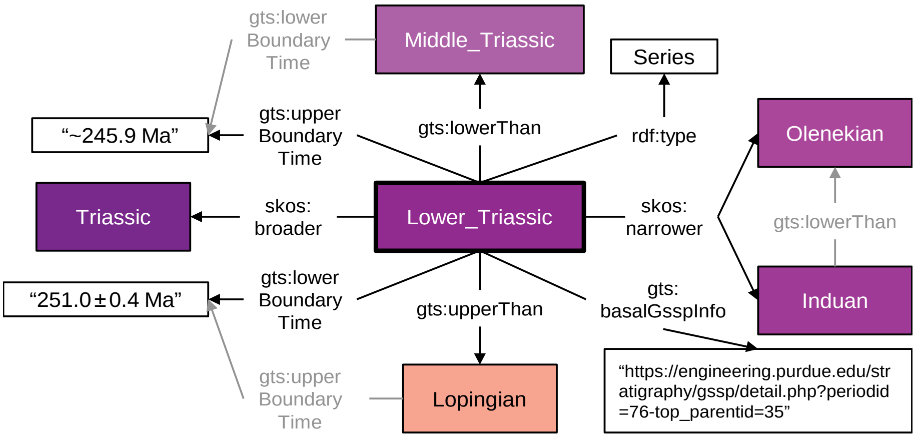
Properties "skos:prefLabel" and "skos:altLabel" of the SKOS model were used to capture multilingual GTS terms in the developed MLTGTS. To capture inter-relationships between GTS terms and add more semantic expressions, the SKOS-model was extended by adding several other object and datatype properties in the developed MLTGTS.

---

[23] https://www.seegrid.csiro.au/subversion/CGI_CDTGVocabulary/tags/SKOSVocabularies/ICS_TimeScale2008.rdf [Accessed February 07, 2011].

```
<skos:Concept rdf:ID="Lower_Triassic">
  <skos:definition>Lower/Early Triassic Series/Epoch; 251.0±0.4—~245.9 Ma</skos:definition>
  <skos:prefLabel xml:lang="cn">下三叠统</skos:prefLabel>
  <skos:prefLabel xml:lang="de">Untertrias</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Lower Triassic</skos:prefLabel>
  <skos:prefLabel xml:lang="es">Triásico Inferior</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">Trias Inférieur</skos:prefLabel>
  <skos:prefLabel xml:lang="jp">下部三畳系</skos:prefLabel>
  <skos:prefLabel xml:lang="nl">Onder Trias</skos:prefLabel>
  <skos:altLabel xml:lang="en">Early Triassic</skos:altLabel>
  <skos:altLabel xml:lang="jp">三畳紀前期</skos:altLabel>
  <skos:altLabel xml:lang="jp">前期三畳紀</skos:altLabel>
  <skos:altLabel xml:lang="jp">三畳系下部</skos:altLabel>
  <skos:altLabel xml:lang="cn">早三叠世</skos:altLabel>
  <skos:altLabel xml:lang="nl">Vroeg Trias</skos:altLabel>
  <skos:altLabel xml:lang="es">Triasico Inferior</skos:altLabel>
  <skos:broader rdf:resource="#Triassic"/>
  <skos:narrower rdf:resource="#Olenekian"/>
  <skos:narrower rdf:resource="#Induan"/>
  <rdf:type rdf:resource="#Series"/>
  <gts:lowerThan rdf:resource="#Middle_Triassic"/>
  <gts:upperThan rdf:resource="#Lopingian"/>
  <gts:upperBoundaryTime>~245.9 Ma</gts:upperBoundaryTime>
  <gts:lowerBoundaryTime>251.0±0.4 Ma</gts:lowerBoundaryTime>
  <gts:basalGsspInfo>
  https://engineering.purdue.edu/stratigraphy/gssp/detail.php?periodid=76-top_parentid=35
  </gts:basalGsspInfo>
  <skos:inScheme rdf:resource="urn:ICS:International_Stratigraphic_Chart:2010"/>
</skos:Concept>
```

(a) Source code.



(b) Graphic view of a part of (a).

Fig. 3.2. Definition of “Lower_Triassic” as a GTS concept with an extended SKOS model. Several pure SKOS properties in (a) (i.e., “skos:Concept”, “skos:definition”, “skos:prefLabel”, “skos:altLabel”, “skos:inScheme”) are omitted in (b) to show the difference between (b) and Fig. 3.1b. Text and symbols in gray color are not shown in (a), but are included in definitions of other GTS terms in the developed MLTGTS.

English chronostratigraphic terms were collected from the 2009 ICS International Stratigraphic Chart (2009 ICS chart) (Ogg, 2009), and were used as basic references of GTS concepts in the MLTGTS. For example, “Lower_Triassic” is encoded with “skos:Concept” in line 1 in Fig. 3.2a,. In line 2, the definition of “Lower_Triassic” is encoded with “skos:definition”. In lines 3 to 9, chronostratigraphic terms in seven languages are encoded as preferred labels of “Lower_Triassic” with “skos:prefLabel”. In lines 10, 12, 14 and 15, geochronologic terms in other languages are encoded as alternative labels of “Lower_Triassic” with “skos:altLabel”. In lines 11, 13 and 16, synonyms of both chronostratigraphic and geochronologic terms of the same SKOS concept are also encoded with “skos:altLabel”.

Unlike the definition of “Lower_Triassic” with the pure SKOS model (Fig. 3.1a), several external properties are used in the definition of that concept in the extended SKOS model (Fig. 3.2a) to represent the ordinal hierarchical structure of GTS. In line 20 in Fig. 3.2a, “Lower_Triassic” is defined as an instance of “Series” (a subclass of chronostratigraphic units) by using a RDF property “rdf:type”. In lines 21 and 22, two properties “gts:lowerThan” and “gts:upperThan” are used to represent that “Lower_Triassic” rocks are stratigraphically lower than “Middle_Triassic” rocks and stratigraphically upper than “Lopingian” rocks, respectively. In lines 23 and 24, properties “gts:upperBoundaryTime” and “gts:lowerBoundaryTime” are used to record, respectively, the upper and lower time boundaries of “Lower_Triassic”, which are derived from the 2009 ICS chart. In lines 25 to 27, “gts:basalGsspInfo” is used to record a web address pointing to the information of the basal Global Boundary Stratotype Section and Point (GSSP) [24] of “Lower_Triassic”. Fig. 3.2b shows the difference of Fig. 3.2a from Fig. 3.1a.

3.2.4 Summary of building the SKOS-based MLTGTS

The process consists of the following steps. First, a subject domain (i.e., GTS) of a thesaurus is chosen. Second, multilingual GTS terms are collected. To improve the interoperability of the resulting thesaurus, most terms were collected from the MTG and AMTG because they have been extensively reviewed and are widely accepted. Third, multilingual labeling (Miles et al., 2001) is used for organizing multilingual GTS terms because of the global common conceptual structure of GTS coordinated by the ICS; and English chronostratigraphic terms from the 2009 ICS chart are used as the basic references because this chart is accepted and used in the geoscience community globally. Fourth, an extended SKOS model is used to encode multilingual GTS terms and to represent the ordinal hierarchical structure of

---

[24] GSSP information is maintained by the Subcommission for Stratigraphic Information of the International Commission of Stratigraphy (https://engineering.purdue.edu/Stratigraphy/index.html) [Accessed February 07, 2011].

GTS. Finally, the MLTGTS is refined by adding more synonyms of multilingual GTS terms obtained from actual geodata. The workload for the first version of the MLTGTS is about 150 man-hours, with participation of geologists from different language background. However, the refinement of the MLTGTS is a continuous process, because there are certainly other synonyms of GTS terms existing in actual geodata in different languages. The size of the current MLTGTS is about 225 KB and it is still increasing slowly.

3.3 Recognizing and translating GTS terms retrieved from WMS

The primary functions of the developed MLTGTS are to recognize and translate GTS terms in GTS records retrieved from geological maps on WMS servers. These can be achieved in four steps (Fig. 3.3). Firstly, the GTS record of an area is retrieved from an online geological map provided by a WMS server. Secondly, GTS terms in that record and their languages are recognized. Thirdly, a recognized GTS term is chosen and information about it is searched in the MLTGTS, and then displayed on the user interface. Finally, other languages supported by the MLTGTS can be chosen on the user interface, such that the MLTGTS is re-searched and the displayed term and related information are translated into the chosen language. In this workflow, the SKOS-based MLTGTS is a RDF document and the user interface is a webpage encoded with HTML (HyperText Markup Language).

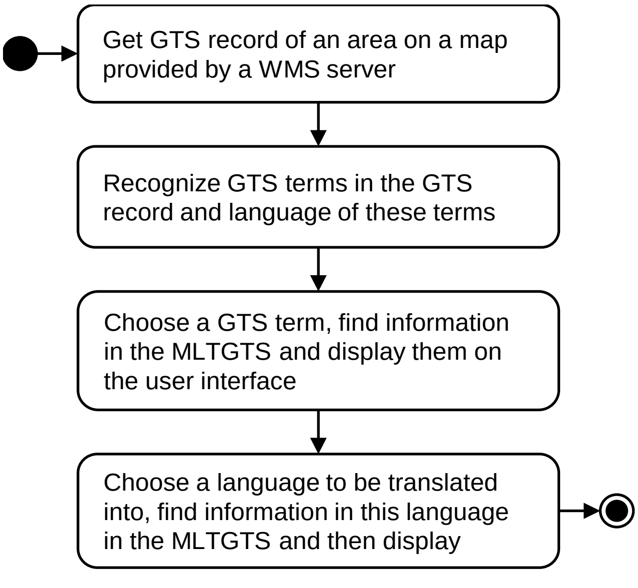


Fig. 3.3. Four-step workflow for recognizing and translating GTS terms in GTS records retrieved from geological maps on WMS servers.

JavaScript and Ajax (Asynchronous JavaScript and XML) techniques (Garrett, 2005) were used for accessing WMS and parsing the MLTGTS through the user interface, such that most of the designed functions in the workflow (Fig. 3.3) could be realized with one programming language. Functions for retrieving GTS records from WMS servers were developed in step 1. Because geological maps are produced by different countries and are in different languages, original GTS records are often presented in diverse styles. Thus, a function was developed to correct spelling errors (Kukich, 1992; Lam-Adesina and Jones, 2006) and to re-format texts of retrieved GTS records for compatibility with terms in the MLTGTS. Because, in step 2, a GTS term under operation has already been recognized in the MLTGTS, steps 3 and 4 simply involve finding the term again in the MLTGTS, and then retrieving information about that term in the chosen language. Compared to steps 1, 3 and 4, recognizing GTS terms in a GTS record (i.e., step 2) costs most of the time in the workflow.

To recognize certain terms from given texts, there are various methods of information retrieval (e.g., Gey et al., 2005; Lazarinis et al., 2009; Baeza-Yates and Ribeiro-Neto, 2011) that can be adopted and/or adapted. Here, the target is comparing a pre-processed GTS record with standard GTS terms in the MLTGTS and then recognizing all GTS terms in that record. To achieve this target, an algorithm (Fig. 3.4) was developed based on Boolean information retrieval (Radecki, 1983; Losee, 1997; Koubarakis et al., 2006). In this algorithm, a letter-by-letter equality (LBLE) condition was imposed for finding standard GTS terms in a GTS record and, accordingly, this involved an iterative process of letter-by-letter comparison. The LBLE condition was used because, in general, a GTS record is a phrase or a short sentence and, thus, letter-by-letter comparison between a GTS record and standard terms in the MLTGTS does not entail huge workloads for a computer. An ideal result of Boolean information retrieval is that the GTS record contains only one GTS term and it is recognized either as a “skos:prefLabel” term or a “skos:altLabel” term in the MLTGTS (Fig. 3.4).

However, multilingual GTS terms have their own spelling features, which defy a pure LBLE search. For example, “Trias” (in German) and “Triassic” (in English) are both “skos:prefLabel”s for the concept “Triassic” in the MLTGTS. If a GTS record is “Triassic to Jurassic” (in English), then, by using a pure LBLE search, “Trias” will also be recognized as a GTS term in that GTS record, which is not the case however. The terms “Lower/Early”, “Middle” and “Upper/Late” diversify GTS records and pose obstacles in a pure LBLE search. For example, a GTS record “Lower Triassic to Middle Triassic” may be written as “Lower to Middle Triassic”. A pure LBLE search can recognize “Middle

Triassic" from the abridged record, whereas the GTS term "Lower" (i.e., "Lower Triassic") is ignored. Worse still, a pure LBLE search will also recognize "Triassic" and "Trias" as GTS terms in that GTS record.

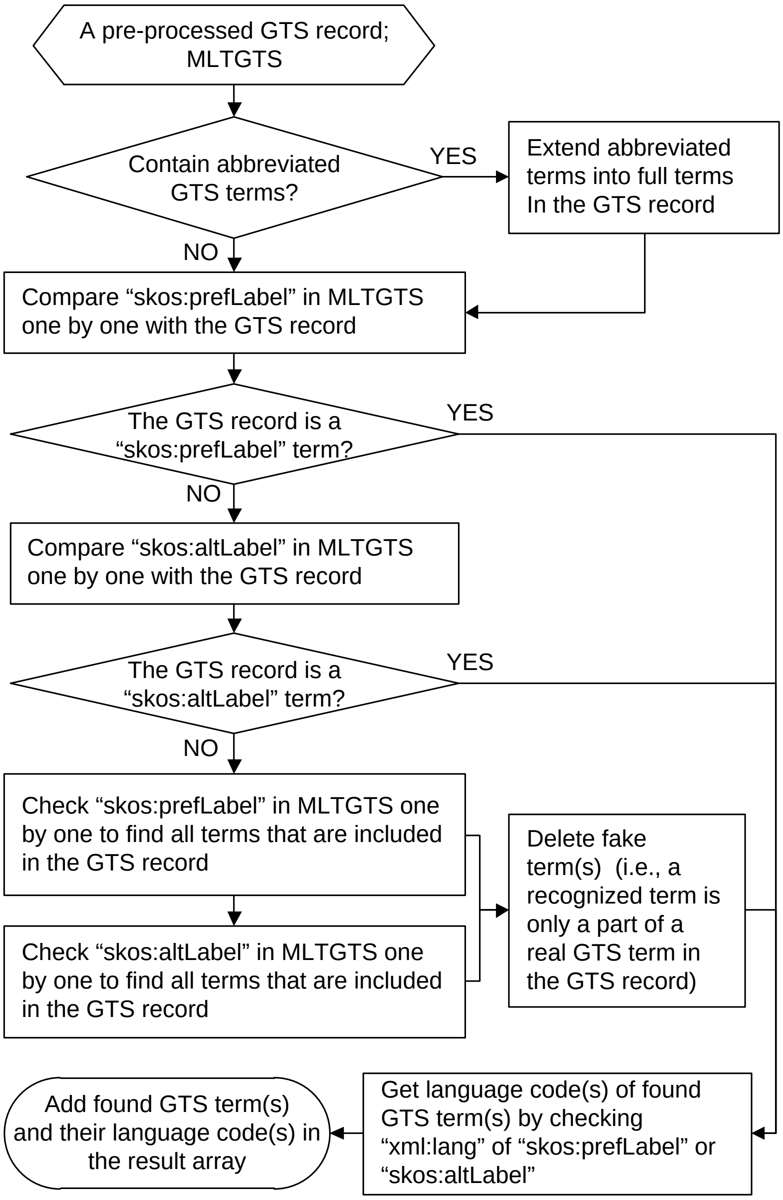


Fig. 3.4. Algorithm for recognizing GTS terms and their languages in a GTS record by using the developed MLTGTS.

To address those challenges and to complement the LBLE search, a group of other methods were developed forming what is called a characteristic-oriented term retrieval algorithm (Fig. 3.4), because they focus on the characteristics of GTS terms and GTS records. For example, one of the methods extends abridged terms in the GTS record into full terms before using the LBLE search. For instance, if a GTS record is “Lower to Upper Triassic”, it is extended to “Lower Triassic to Upper Triassic” prior to LBLE search. Another method deletes fake terms in the recognized term list. For instance, if a GTS record is “Lower Triassic to Lower Jurassic”, a pure LBLE search will recognize “Lower Triassic”, “Lower Jurassic”, “Triassic”, “Jurassic”, “Trias” and “Jura” as GTS terms in that record. The latter four terms will be recognized as fake terms and will be deleted by the algorithm shown in Fig. 3.4, leaving only “Lower Triassic” and “Lower Jurassic”, as desired, in the result. This characteristic-oriented term retrieval algorithm substantially improved the accuracy of GTS term recognition.

3.4 Pilot system, results and evaluation

A pilot system was set up to test the accuracy of the MLTGTS and the functionality of the JavaScript programs for recognizing and translating GTS terms in GTS records of online geological maps. A purpose of this pilot system, being consistent with the aforementioned workflow, is retrieving background knowledge of GTS terms from the MLTGTS and showing them in a way that is easy for both geologists and non-geologists to access. Datasets in the pilot system include the following geological maps (as ARCGIS .shp files) stored in a self-built WMS server:

- 1:200,000 Geological Map of Japan (GSJ-AIST, 2009), for testing recognition and translation of Japanese GTS terms;
- 1:5,500,000 Geological Map of South America (CGMW et al., 2003), for testing recognition and translation of Spanish GTS terms;
- 1:600,000 Superficial Rock Age Map of The Netherlands (Schokker, 2010), for testing recognition and translation of Dutch GTS terms;
- 1:625,000 Bedrock Age Map of United Kingdom (BGS, 2005), for testing recognition and translation of British English GTS terms; and
- 1:250,000 Geologic Map of New York (Dicken et al., 2008), for testing recognition and translation of American English GTS terms.

GTS information is retrieved by clicking polygons in a WMS map layer, and then GTS terms are recognized and translated by using the MLTGTS and JavaScript programs. Besides the self-built WMS server, from remote WMS servers maintained by geological surveys of several countries, the following geological map layers were also used. In the following list, layers from the first four WMS servers are used for testing recognition and translation of

English GTS terms, and layers in the last WMS server are used for testing recognition and translation of Dutch and German GTS terms.

- WMS: CCOP Combined Bedrock and Superficial Geology and Age[25];
  Layer: EASIA CCOP 1:2,000,000 Combined Bedrock and Superficial Geology and Age;
- WMS: GSJ Combined Bedrock and Superficial Geology and Age[26];
  Layer: JPN GSJ 1:1,000,000 Combined Bedrock and Superficial Geology and Age;
- WMS: BGS GSN Bedrock geology[27];
  Layer: NAM GSN 1:1,000,000 Bedrock Age;
- WMS: BGS Bedrock and Superficial geology[28];
  Layer: GBR BGS 1:625,000 Bedrock Age; and
- WMS: DinoMap Geological maps[29]
  Layers: Geological map of NL 600k (German legend); Geological map of NRW 100k (original map).

The following open-source or free software programs were used for developing the pilot system:

- Protégé 4.0.2[30] and SKOSEd-1.0-alpha (build04)[31], for editing the MLTGTS;
- Notepad++ 5.7[32], for revising the MLTGTS, and for editing JavaScript programs and the HTML file of the user interface;
- Firefox 3.6.8[33] and Firebug 1.5.4[34], for debugging JavaScript programs and browsing the HTML file of the user interface;
- GeoServer 2.0.1[35], for setting up a WMS server;
- uDig 1.1[36], for editing SLD (Styled Layer Descriptor) files of geological maps stored in GeoServer 2.0.1; and
- OpenLayers 2.9.1[37]: for retrieving spatial and attribute data from a WMS server.

---

[25] http://geodata1.geogrid.org/mapserv/CCOP_Combined_Bedrock_and _Superficial_Geology_and_Age/wms? [Accessed February 07, 2011].

[26] http://geodata1.geogrid.org/mapserv/GSJ_Combined_Bedrock_and _Superficial_Geology_and_Age/wms? [Accessed February 07, 2011].

[27] http://ogc.bgs.ac.uk/cgi-bin/BGS_GSN_Bedrock_Geology/wms? [Accessed February 07, 2011].

[28] http://ogc.bgs.ac.uk/cgi-bin/BGS_Bedrock_and_Superficial_Geology/wms? [Accessed February 07, 2011].

[29] http://www.dinoservices.nl/wms/dinomap/M07M0034? [Accessed February 07, 2011].

[30] http://protege.stanford.edu [Accessed February 07, 2011].

[31] http://code.google.com/p/skoseditor [Accessed February 07, 2011].

[32] http://notepad-plus-plus.org [Accessed February 07, 2011].

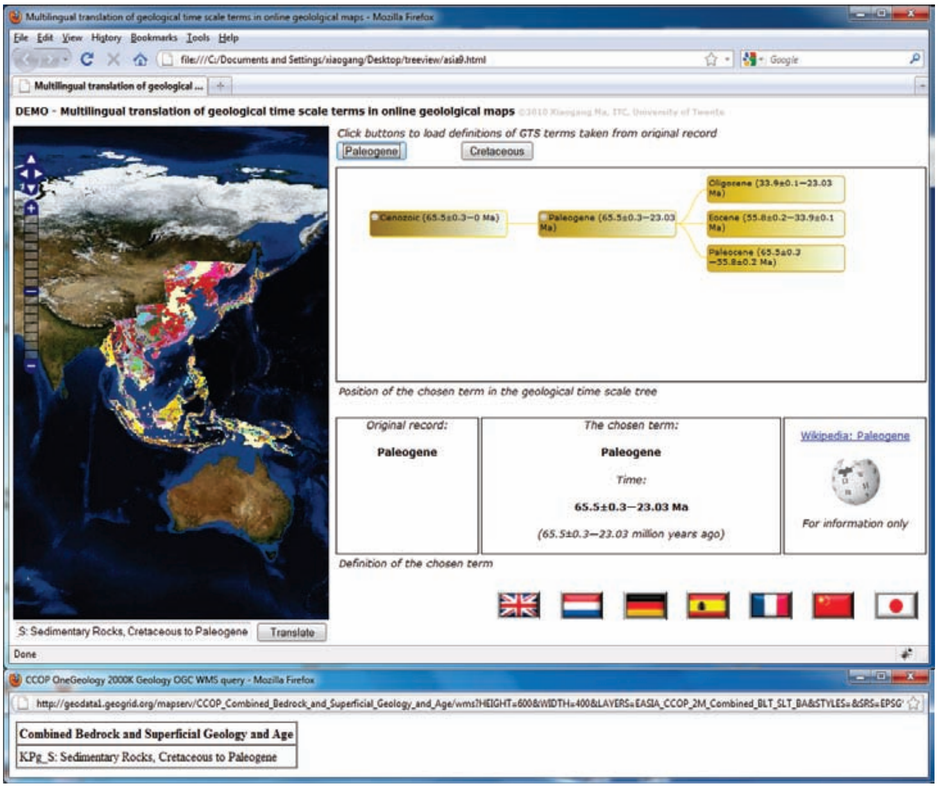
[33] http://www.mozilla-europe.org/en/firefox [Accessed February 07, 2011].

[34] http://www.getfirebug.com [Accessed February 07, 2011].

[35] http://www.geoserver.org [Accessed February 07, 2011].

[36] http://udig.refractions.net [Accessed February 07, 2011].

[37] http://www.openlayers.org [Accessed February 07, 2011].



file:///C:/Documents and Settings/xiaogang/Desktop/treeview/asia9.html

DEMO - Multilingual translation of geological time scale terms in online geololgical maps ©2010 Xiaogang Ma, ITC, University of Twente

Click buttons to load definitions of GTS terms taken from original record

Position of the chosen term in the geological time scale tree

Original record:

Paleogene

The chosen term:

Paleogene

Time:

65.5±0.3—23.03 Ma

(65.5±0.3—23.03 million years ago)

For information only

Definition of the chosen term

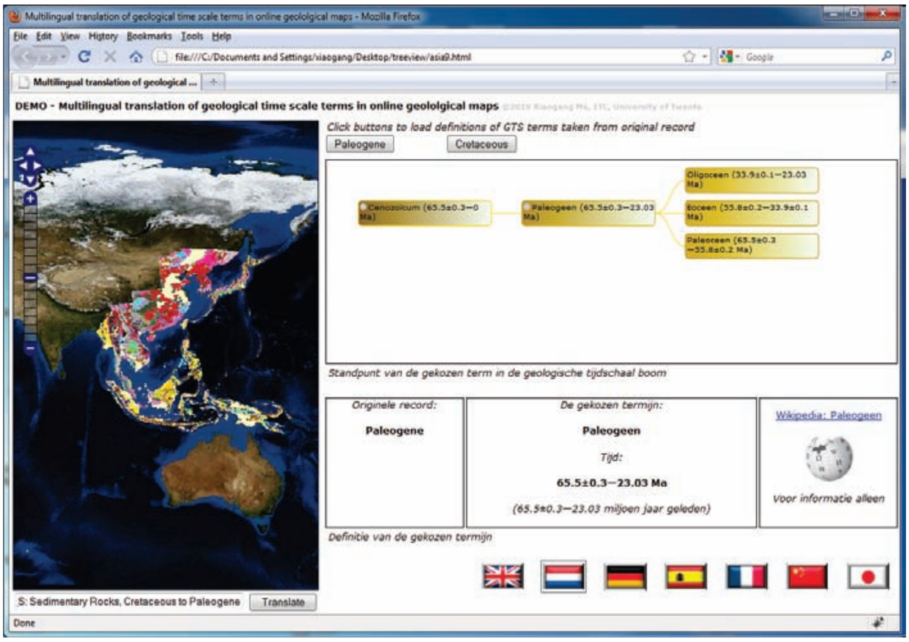
S: Sedimentary Rocks, Cretaceous to Paleogene

http://geodatal.geogrid.org/mapserv/CCOP_Combined_Bedrock_and_Superficial_Geology_and_Age/wms?HEIGHT=600&WIDTH=400&LAYERS=EASIA_CCOP_2M_Combined_BLT_SLT_BA&STYLES=&SRS=EPSG'

Combined Bedrock and Superficial Geology and Age
KPg_S: Sedimentary Rocks, Cretaceous to Paleogene

(a) Information of "Paleogene" shown in English.

Fig. 3.5. Running JavaScript programs and MLTGTS through a web browser to translate GTS terms in GTS records retrieved from a WMS server. (a) GTS terms retrieved from a WMS server are recognized and listed on the user interface. Users can choose a GTS term from the list and then its location in the GTS tree structure, its definition and a link to its corresponding Wikipedia page are shown. (b) (on the next page) Users can choose a preferred language by clicking a flag button and then the GTS tree structure, the definition of the GTS term and additional information on the user interface are translated into the chosen language. Geological map reproduced with the permission of the OneGeology secretariat & registered participants. All Rights Reserved.



DEMO - Multilingual translation of geological time scale terms in online geololgical maps

Click buttons to load definitions of GTS terms taken from original record

Standpunt van de gekozen term in de geologische tijdschaal boom

Originele record: Paleogene

De gekozen termijn: Paleogeen

Tijd: 65.5±0.3—23.03 Ma

(65.5±0.3—23.03 miljoen jaar geleden)

Voor informatie alleen

Definitie van de gekozen termijn

S: Sedimentary Rocks, Cretaceous to Paleogene

(b) Information of "Paleogene" shown in Dutch.

By operating the pilot system with geological maps on the self-built and remote WMS servers, the recognized GTS terms in English, Spanish, Dutch, German or Japanese and relevant information about them could be translated into any one of the seven languages supported by the MLTGTS. Fig. 3.5a shows the user interface of the pilot system with the layer "EASIA CCOP 1:2,000,000 Combined Bedrock and Superficial Geology and Age" retrieved from a remote WMS server. In this example, an area in the map was clicked first. However, because the cross-domain data access is limited by JavaScript currently[38], a pop-up window was used to show the retrieved GTS record (bottom part of Fig. 3.5a). Then, the record was copied manually to a text box at the bottom left part of the main user interface, and the button "Translate" was clicked. Two GTS terms, "Paleogene" and "Cretaceous", were recognized and listed at the top of the main user interface. Meanwhile, the language code (i.e., "en") of the two terms were recognized and recorded, but not shown on the user interface. Then, either recognized term can be clicked and information of this term is shown in its original language. Fig. 3.5a shows the result after the term "Paleogene" was clicked. After that, any one of the seven flag buttons at the bottom right part of the user interface can be clicked, and information shown on the user interface is translated into

---

[38] Technologies of cross-domain data access is now being worked on by the W3C community. http://www.w3.org/TR/cors [Accessed February 07, 2011].

the corresponding language. Fig. 3.5b shows the result after the Dutch flag button was clicked.

Table 3.2 GTS records in the Geological Map of Kumamoto that are successfully and unsuccessfully translated by the pilot system of the MLTGTS

Original record	**In English**	**Record type**	**Pilot system translation**
白亜紀	Cretaceous	1) Containing only one GTS term	Successful
シルル紀-デボン紀	Silurian to Devonian	2) Containing two GTS terms, without “Upper”, “Late”, “Middle”, “Lower” or “Early”	Successful
ペルム紀-前期白亜紀	Permian to Early Cretaceous	3) Containing two GTS terms, one or two with “Upper”, “Late”, “Middle”, “Lower” or “Early”	Successful
前-後期ジュラ紀	Early to Late Jurassic	4) Containing two GTS terms, one in abbreviated format	Successful
中期始新世	Middle Eocene	5) Containing only one GTS term, not found in 2009 ICS chart	Unsuccessful
中期始新世-前期漸新世	Middle Eocene to Early Oligocene	6) Containing two GTS terms, not found in 2009 ICS chart	Unsuccessful
時代未詳	Unknown age	7) Containing no GTS terms	Unsuccessful

Table 3.3 Results of recognizing and translating GTS terms in GTS records of some 1:200,000 Geological Maps of Japan

Map name	**GTS records**	**Differentiated GTS records**	**Successful translations**	**Failed translations**
Geological Map of Hyogo	3747	27	25	2
Geological Map of Ibaraki	2419	21	19	2
Geological Map of Miyazaki	4188	31	29	2
Geological Map of Niigata	4906	28	26	2
Geological Map of Osaka	1442	16	15	1

Applying the MLTGTS and JavaScript programs in the pilot system is not a “one station stop” work. Instead, evaluations and revisions on them are iterative. By operating the pilot system with GTS records retrieved from actual geological maps in different languages, insufficiencies of the MLTGTS and the JavaScript programs can be found. Then the results can be evaluated, and the MLTGTS and/or the JavaScript programs can be revised to improve their accuracy and functionality. For example, the Geological Map of Kumamoto was recently taken from the 1:200,000 Geological Map of Japan and the recognition and translation of Japanese GTS records were operated. This map contains 2822 original GTS records, which can be condensed into 32 differentiated GTS records (i.e., the 2822 original GTS records are just repetitions of these 32 differentiated records). With an earlier version of the

MLTGTS, the pilot system successfully translated GTS terms in 21 of those 32 differentiated GTS records but failed for the other 11 records. The compositions of the successfully translated 21 GTS records can be classified into four types, whereas the compositions of the unsuccessfully translated 11 GTS records can be classified into three types. Table 3.2 shows a list of examples of all these seven types. Nine of the 11 unsuccessfully translated GTS records contain valid GTS terms that are not included in the 2009 ICS chart. These unsuccessful pilot results allowed the revision of the MLTGTS by encoding those valid GTS terms that are not included in the 2009 ICS chart. With the revised MLTGTS, the pilot system successfully translated GTS terms in 30 (types 1-6) of the 32 differentiated GTS records but still failed for the other two records (type 7) that contain no GTS terms. The pilot system (with the revised MLTGTS) was further tested to recognize and translate GTS terms in GTS records of other 1:200,000 Geological Maps of Japan, and satisfactory results were obtained (Table 3.3).

Table 3.4 GTS records of Dutch-German border areas in the geological map of NL 600k (German legend) and the geological map of NRW 100k (original map) that are successfully and unsuccessfully translated by the pilot system of the MLTGTS

Map	**Original record**	**In English**	**Record type**	**Pilot system translation**
NL 600k	Boven Krijt	Upper Cretaceous	1.1) Containing only one GTS term	Successful
NL 600k	Tertiair	Tertiary	1.2) Containing only one GTS term, not found in 2009 ICS chart but found in MLTGTS	Successful
NL 600k	Weichseliën - Saaliën	Saalian to Weichselian	1.3) Containing “Weichseliën” and/or “Saaliën”	Unsuccessful
NRW 100k	Pliozän	Pliocene	2.1) Containing only one GTS term	Successful
NRW 100k	Miozän bis Oligozän	Miocene to Oligocene	2.2) Containing two GTS terms	Successful
NRW 100k	Unterpleistozän	Lower Pleistocene	2.3) Containing only one GTS term, not found in 2009 ICS chart but found in MLTGTS	Successful
NRW 100k	Weichselium	Weichselian	2.4) Containing “Weichselium” and/or “Saalium”	Unsuccessful

In the pilot system, it was also performed a case study of translating GTS terms in geological maps of areas across, say, borders between two countries. That is because geological mapping of border areas is an increasingly discussed topic in recent years (Satkunas and Graniczny, 1997; Asch, 2001; Podemski, 2005; OneGeology-Europe Consortium, 2010). Geological units are naturally independent of administrative borders, but geological maps in border areas usually exhibit certain inconsistencies, including GTS nomenclatures (Satkunas et al., 2004). There are vast challenges, including the linguistic barriers discussed above, in harmonizing geological maps in border areas (e.g., Delgado et al., 2001; Asch, 2001, 2005). The purpose of this case study was to check whether SKOS-based multilingual geoscience thesauri can address the challenge of linguistic barriers in harmonizing geological maps in border areas.

For this case study, two geological maps – geological map of NL 600k (German legend) and geological map of NRW 100k (original map) – were retrieved from the WMS server “DinoMap Geological maps” (see footnote 29). These two maps cover the Dutch-German border areas between the provinces Overijssel, Gelderland and Limburg of the Netherlands and the state North Rhine-Westphalia (NRW) of Germany. In the first map, 11 differentiated Dutch GTS records were found along the Dutch-German border areas; whereas in the second map, 14 differentiated German GTS records were found. Compositions of GTS records in the two maps can be classified into three and four types, respectively (Table 3.4).

With the earlier version of the MLTGTS, the pilot system successfully translated GTS terms in six (types 1.1 and 1.2) of the 11 differentiated GTS records in the first map and in 11 (types 2.1–2.3) of the 14 differentiated GTS records in the second map. In the first map, it was found that the five unsuccessfully translated GTS records (type 1.3) include two terms “Weichseliën” (Weichselian) and “Saaliën” (Saalian). These terms are not included in the 2009 ICS chart, but are used as Stage terms in regional subdivisions of the Pleistocene Series in North West Europe[39]. In the second map, the three unsuccessfully translated GTS records (type 2.4) include two terms “Weichselium” (Weichselian) and “Saalium” (Saalian), which are not included in the 2009 ICS chart. By referring to the chart (v. 2010) provided by the Subcommission on Quaternary Stratigraphy of ICS (see footnote 39), these terms and their multilingual versions were added into the MLTGTS. With the updated MLTGTS, the pilot systems successfully translated all GTS terms in the 25 differentiated GTS records from both maps. By operating the recognition and translation, original Dutch or German GTS terms can be translated into any one of the seven language supported by the MLTGTS. This would provide convenience to geologists working on geological maps of

---

[39] Stratigraphical charts for the Quaternary. http://www.quaternary.stratigraphy.org.uk/charts [Accessed February 07, 2011].

Dutch-German border areas. Results of this case study show the benefits of using SKOS-based multilingual geoscience thesauri in using geological maps of border areas, although it addresses only a part (i.e., linguistic barriers) of the challenges in harmonizing multisource geological maps.

3.5 Discussion

With proper extensions, the SKOS is functional for encoding multilingual geoscience thesauri into a format that is compatible with the Semantic Web, and SKOS-based multilingual geoscience thesauri are efficient for translating online geoscience records into any language that is supported by the thesauri. Only a multilingual thesaurus of GTS was developed in the study presented in this chapter, but it is transparent that other SKOS-based multilingual thesauri of different subject domains in geosciences can also be developed. By using these multilingual geoscience thesauri, geological maps in their native languages can be published online directly, while users in regional/global cooperative projects can translate the maps and browse the data in their preferred languages. In this way, linguistic barriers between online geological maps can be reduced and, thus, their interoperability can be improved. Results of the pilot system demonstrate the accuracy of the MLTGTS and the functionality of the JavaScript programs to recognize and translate GTS terms in multilingual geological maps. Meanwhile, background information of GTS terms retrieved from the MLTGTS is also displayed on the user interface in the chosen language. Because the multilingual terms and their definitions and relationships in the MLTGTS were collected from credible sources, users can get precise explanations of GTS terms from the MLTGTS. By reading the translated GTS terms and background information in preferred languages, users can access the information represented by the GTS records in an easier way.

It has been extensively discussed that the SKOS model has several advantages compared to other models for encoding thesauri in the Web context (van Assem et al., 2006; Miles and Pérez-Agüera, 2007; Pastor-Sanchez et al., 2009). SKOS model is based on RDF, making SKOS-based thesauri compatible with other standards and technologies of the Semantic Web. In recent years, SKOS model has been applied to build or rebuild thesauri in various fields, such as agriculture (Soergel et al., 2004), stratigraphy (Fils et al., 2009), authority files (Voss, 2009), and the aforementioned geoscience thesauri edited by the Geoscience Concept Definitions Task Group of the CGI-IUGS, etc. There are also significant studies on building thesauri from diverse resources (e.g., Broughton, 2006; Hepp, 2006; Fang et al., 2008; Tsuruoka et al., 2008), addressing semantic and syntactic issues from various aspects. In the study presented in this chapter, the semantics of the GTS has been defined by the 2009 ICS chart, which is accepted as a global standard. Instead of redefining semantics of the GTS, the existing standard was adopted (cf. McGuinness, 2003; Bibby, 2006) and then adapted with an extended SKOS model. Although the multilingual GTS terms were collected from different resources, they could be added into the MLTGTS easily, because the meanings (i.e., semantics) of these terms are defined by their time boundaries and, thus, their locations in the ordinal hierarchical structure GTS are clear.

For GTS thesauri, Cox and Richard (2005) discussed in detail components in the GTS and drew conceptual schemas for them by using Unified Modeling Language (UML). They also transformed the UML schemas into XML formats so that they can be used on the Web. The resulting schemas represent units, boundaries and GSSPs in the GTS and their relationships. Thus, those schemas not only represent GTS terms but also show how they were derived. Compared to the extended SKOS model used for the MLTGTS, the conceptual schemas of Cox and Richard (2005) are more thorough. The SKOS-based thesaurus of GTS (see footnote 23) edited by the Geoscience Concept Definitions Task Group of CGI-IUGS adapted the work of Cox and Richard (2005) by simplifying the components to fit the SKOS model. The current version of the CGI-IUGS GTS thesaurus also refers to the ICS chart and covers GTS terms in English, French, Italian and Slovakian. However, unlike the MLTGTS discussed here, the CGI-IUGS GTS thesaurus does not distinguish between chronostratigraphic and geochronologic terms. For some GTS concepts, the CGI-IUGS GTS thesaurus combines chronostratigraphic terms with geochronologic units in their definitions, which potentially causes confusion. For example, “Upper Cretaceous” in the CGI-IUGS GTS thesaurus is defined as “Upper Cretaceous Epoch”, whereas it should be “Upper Cretaceous Series” or “Late Cretaceous Epoch”. Such issues were discussed within the CGI-IUGS community in a recent workshop[40] and more international cooperation was proposed.

Another example of GTS thesaurus is the multilingual geological age thesaurus developed in the OneGeology-Europe project (1G-E)[41] recently (Asch, 2010). The 1G-E hosts a web portal[42] providing multilingual (i.e., 18 European languages) access to contents of semantically and technically interoperable1:1000,000 scale geological maps for the whole of Europe

---

[40] IUGS-CGI and OneGeology-Europe Geoscience Language Workshop (IGSL 2010). http://www.bgr.bund.de/cln_116/nn_1951520/EN/Themen/GG__geol__Info/IGSL2010 [Accessed February 07, 2011].

[41] http://www.onegeology-europe.org/home [Accessed February 07, 2011].

[42] http://onegeology-europe.brgm.fr/geoportal/viewer.jsp [Accessed February 07, 2011].

(Laxton et al., 2010). Such functions are bolstered by SKOS-based multilingual thesauri of lithology, age (geochronology), genesis, and structures and faults developed by the 1G-E Work Package 3 (1G-E WP3) (Asch, 2010). There are several differences between the SKOS-based thesauri of 1G-E and the MLTGTS discussed here. The MLTGTS in the study described in this chapter adopts the 2009 ICS chart and uses chronostratigraphic units. The geological age thesaurus of 1G-E uses geochronologic units, and subdivides the periods of Precambrian for the Europe and adds 27 new terms accordingly (Asch et al., 2010). Another difference is that the MLTGTS in this study includes two Asian languages, while the geological age thesaurus of 1G-E focuses on European languages. Besides these differences, the goal of using MLTGTS to translate GTS records of online geological maps in this study is similar to that of the 1G-E project, although the current literature of 1G-E shows little about whether or not its web portal applies a workflow of retrieval, recognition, translation and display that is similar to what was developed in this study (Fig. 3.3).

The benefits of embedding ontologies in Spatial Data Infrastructures (SDI) have also been discussed significantly in recent years (Ludäscher et al., 2003; Georgiadou, 2006; Lacasta et al., 2007; Sinha et al., 2007). By using ontologies in a SDI, heterogeneous geodata sources can be mapped to common models; meanings of inconsistent concepts can be harmonized and the semantic interoperability of geodata can be improved. Geoscience thesauri, as “simple ontologies”, are also functional for improving the interoperability of geodata (Ma et al., 2007; Ma et al., 2010b). Because SKOS-based thesauri are compatible with standards and technologies of the Semantic Web, using them can potentially lead to more features in a SDI. The results of the pilot system in the study presented in this chapter and the 1G-E web portal already show some of these features. Recently, the AuScope[43] project has built services using SKOS vocabularies for querying geodata (Woodcock et al., 2010). The AuScope vocabularies record synonyms of geoscience terms by using the label “skos:altLabel”. Even users input alternative names of geoscience terms for querying, the vocabulary services can find certain concepts and then retrieve desired geodata. The label “skos:altLabel” was also used in the study presented in this chapter for recording synonyms of GTS terms. However, compared to the vocabulary services/applications of AuScope, the application of MLTGTS in this study is not for querying geodata but for recognizing and translating GTS terms and showing background knowledge about them.

---

[43] http://www.auscope.org [Accessed February 07, 2011].

Several lessons are learnt from the study presented in this chapter. Firstly, SKOS provides a concise model for representing hierarchical structures, but it may be insufficient or inappropriate for structures of certain subject domains in geosciences and this may require an extension to the SKOS model in practice. Thus, because GTS is not a pure hierarchical structure but an ordinal hierarchical structure divided by time boundaries, the SKOS model was extended by adding several other properties, as described earlier, so that that the extended model can represent the ordinal hierarchical structure of GTS properly. Secondly, SKOS is good for encoding multilingual geoscience thesauri, but matching multilingual geoscience terms and building inter-relationships still need geoscience knowledge and cooperation of experts from different language background. Thus, the 2009 ICS chart was referred for GTS terms in English and the MTG and AMTG were referred for multilingual GTS terms because they are results of international cooperation that are accepted globally. However, because some GTS terms are mis-matched or missed in the MTG and AMTG, various other sources were also referred for collecting credible multilingual GTS terms. Thirdly, many synonyms in different languages should be collected as much as possible in geoscience thesauri. Although international standards or agreements on professional terms of a certain subject domain exist, synonyms are still used in current geoscience works. For example, in British English there are three GTS terms “Cainozoic”, “Palaeozoic” and “Archaean”, which correspond to “Cenozoic”, “Paleozoic” and “Archean”, respectively. Such synonyms were encoded in the MLTGTS so that if they are encountered in practice, they can be recognized by using the MLTGTS. Finally, “new” standards cannot be used to explain “old” data, denoting that if a concept’s meaning is changed in the thesaurus, it cannot be used to explain a record using the previous meaning of that concept. For example, in the 2009 ICS chart, the basal boundary of Quaternary is different from that in previous versions of ICS charts. The MLTGTS refers to the 2009 ICS chart for the most recently defined meaning of Quaternary. However, if a record “Quaternary” in a map refers to the 2008 ICS chart, then the definition of “Quaternary” in the MLTGTS cannot be used to explain the meaning of that record (cf. Mascarelli, 2009). This reminds the geoscience community that thesauri used by a geodata source could be attached along with the geodata, or at least, a record of used thesauri could be noted in the metadata of a geodata source.

Because SKOS-based multilingual geoscience thesauri are still an emerging topic in the field of geosciences, many further studies can be proposed. One possible work is collecting more synonyms for GTS terms not only in the seven languages considered in the study presented in this chapter but also in other languages to enrich the MLTGTS. Another work is incorporating results and lessons of this study with other efforts for developing multilingual geoscience thesauri, such as that of the Geoscience Concept Definitions Task Group of the CGI-IUGS. In a broader perspective, SKOS-based multilingual geoscience thesauri can be maintained by international task groups in the CGI-IUGS and published online. Meanwhile, they can be accessed and used by many different organizations and individuals globally for various applications (cf. Schäffer et al., 2010). New technologies for parsing SKOS-based thesauri can also be studied further. JavaScript programs are efficient for parsing the MLTGTS in this study because it is small; for parsing a large SKOS-based thesaurus or a group of large thesauri, those programs require further testing. Some other technologies for parsing thesaurus (e.g., SPARQL[44]) can be tested in the further studies. Transforming the SKOS-based MLTGTS into a RDF/OWL-based ontology of GTS is an open topic, because although SKOS and RDF/OWL are compatible in physical formats, a RDF/OWL-based ontology of GTS is capable of adding more semantic descriptions for concepts and relationships. The work of the geological time ontology[45] in the SWEET project (Raskin and Pan, 2005) can be referred to in this further study.

3.6 Conclusions

Fast evolving Web-based technologies provide not only platforms for building online geological data services but also opportunities for alleviating linguistic barriers to geological data use in regional/global environments. Among various proposed technologies, the SKOS model is advantageous as a start point for encoding and applying multilingual geoscience thesauri in the context of the Semantic Web, and it can be extended in conjunction with other approaches to express concepts and relationships of a subject domain properly. In the study presented in this chapter, a multilingual thesaurus of geological time scale was encoded with an extended SKOS model and, coupled with the thesaurus, methods of characteristic-oriented term retrieval were implemented in JavaScript programs for recognizing and translating geological time scale terms in online geological maps. The developed thesaurus and associated programs were used in a pilot system to recognize and translate geological time scale terms in actual geological maps. Results of the pilot system proved the accuracy of the developed multilingual thesaurus of geological time scale and the functionality of the JavaScript programs. This study shows that SKOS-based multilingual geoscience thesauri can be functional for alleviating linguistic barriers between online geological maps and, thus, improving their interoperability. However,

---

[44] http://www.w3.org/TR/rdf-sparql-query [Accessed February 07, 2011].

[45] http://sweet.jpl.nasa.gov/2.0/timeGeologic.owl [Accessed February 07, 2011].

background knowledge of a subject domain is essential when SKOS is used for building a multilingual geoscience thesaurus of that domain. In addition, it may be necessary to extend the SKOS model in order to obtain satisfactory semantic expressions in certain subject domains in geosciences. Methods of conceptual analysis are beneficial for the extension and/or enrichment of semantic expressions in an ontology. Chapter 4 will discuss how to integrate different conceptual models to develop computer programs and generate standard-compatible results for a subject in geology. Chapter 5 will describe a RDF/OWL-based ontology of geological time scale, in which the semantic expressions of geological time concepts are enriched compared to the SKOS-based thesaurus described in this chapter.

Chapter 4

Standard-compatible conceptual schemas for mine geological data

This chapter is based on: Ma, X., Carranza, E.J.M., van der Meer, F.D., Wu, C., Zhang, X., 2010. Algorithms for multi-parameter constrained compositing of borehole assay intervals from economic aspects. Computers & Geosciences 36 (7), 945–952.

4.1 Introduction

Conceptual analysis is important for subjects in the field of geology because the resulting conceptual schemas/models are often used as basic frameworks of massive geological data and, thus, are closely related to the interoperability of geological data. One of the data-intensive geological subjects is the compositing of borehole metal-grade intervals. Compositing of borehole intervals is an initial step in the cross-sectional method (Fig. 4.1) used for modeling orebodies and estimating mineral resources of a deposit (Sinclair and Blackwell, 2002; Xu and Dowd, 2003; Hustrulid and Kuchta, 2006). Compositing can be based on different criteria, such as geological conditions and economic aspects, etc. Compositing based on economic criteria (e.g., metal-grade) (Ranta et al., 1984; Green, 1991; Bonham-Carter, 1994), which is often used when well-defined geological boundaries are absent for defining orebodies, is a stepwise procedure that consists of (1) classifying raw intervals along a borehole by assayed grades; (2) combining them into composite intervals (i.e., composites) according to certain economic criteria and (3) classifying resulting composites. In the 3rd step, two properties of a composite – average grade and total length – are often used to determine, respectively, economic and mining feasibility of that composite (cf. Ranta et al., 1984; Diering, 1992). If values of those two properties of a composite are both above certain cut-off values (i.e., cut-off grade and minimum mining length, respectively), then that composite is considered to be economic and minable. If the average grade of a composite is above the cut-off value but its total length is not, then that composite may have to be diluted with adjoining intervals to further determine its economic and mining feasibility.

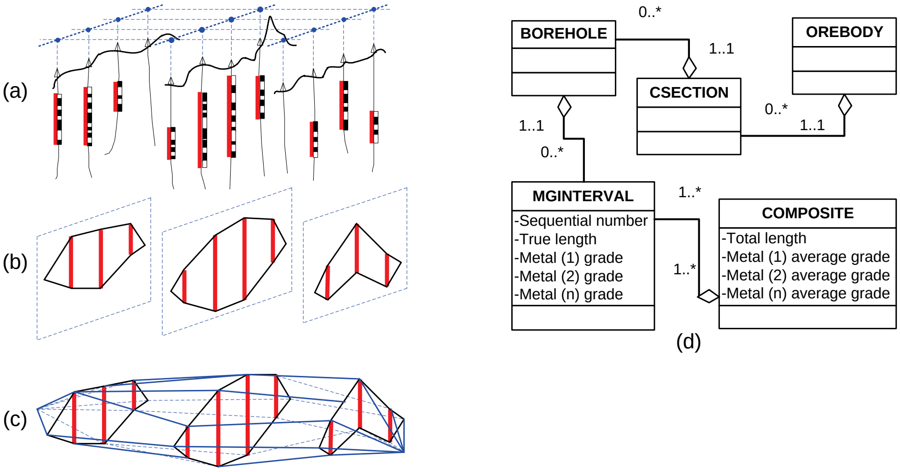


Fig. 4.1. Concept of orebody modeling based on borehole ore composites derived cross-sectional method: (a) minable ore composites determined in each borehole, in which short black and white blocks stand for borehole assay intervals, while a bold red line attached to a group of intervals stands for a minable ore composite; (b) orebody outline in each cross section based on minable ore composites; (c) 3D orebody model interpolated from cross sections; and (d) a model of objects, attributes and relationships in (a) to (c).This chapter focuses on works in step (a). In (d), “MGINTERVAL” and “COMPOSITE” may contain grade values of several metals, but this chapter deals with intervals and composites with a single metal-grade.

Dilution, in compositing of metal-grade intervals, means adding external intervals into a short economic composite in order to exceed the minimum mining length, while the reduced grade still exceeds the cut-off grade (Annels, 1991; Villaescusa, 1998; Sinclair and Blackwell, 2002). Dilution makes the compositing procedure complex and tedious for manual work, which is why computer programs are necessary to facilitate this work. To develop such computer programs, conceptual schemas/models are required because they show a map of concepts and their relationships in the compositing of borehole metal-grade intervals.

In this chapter, data-flow models (DFM) and object-oriented models (OOM) are used together to underpin a program for compositing borehole metal-grade intervals. Steps in the compositing procedure are followed in the DFM, and objects, sets/classes of objects and their interrelationships are identified in the OOM. DFM and OOM generated in this study are compatible with commonly used standards in the field of mineral resources estimation. This increases the interoperability of results produced by the developed program.

4.2 Data-flow and object-oriented models

4.2.1 Compositing procedure and objects involved

Borehole metal-grade intervals are basic elements (i.e., input data) in the compositing. Although an interval may have assayed grades of several metals (Fig. 1d), this study focuses on the compositing of intervals with a single metal-grade. Thus, any borehole metal-grade interval s has three attributes: sequential number id, metal-grade g_s and true length l_s (Table 4.1). Correspondingly, any composite c has two attributes: average grade g_c and total length l_c. An individual s can be classified into either the economic interval set (S^E) or the waste interval set (S^W) by using cut-off grade G_M as criterion. Accordingly, S^E and S^W are both subsets of the interval set S. A looping construct (Fig. 4.2) was set up as the DFM to classify metal-grade intervals in a borehole.

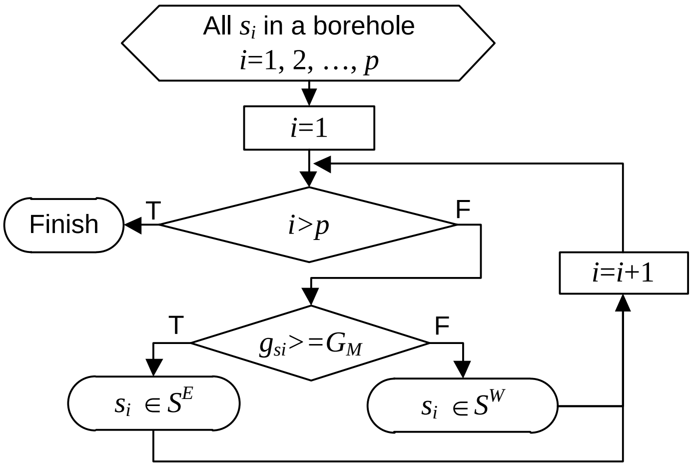


Fig. 4.2. Steps for classifying metal-grade intervals in a borehole. In the diagram $g_{si} = s_i.g_s$, and T means True and F means False. For meanings of other symbols see Table 4.1.

Table 4.1 Meanings of symbols used in the compositing procedure

Symbol	Meaning
S	Borehole metal-grade interval (a set or a class)
S^E	Economic interval (a set or a class)
S^W	Waste interval (a set or a class)
C	Composite (a set or a class)
L_M	Minimum mining length. The thinnest ore zone that can be extracted by the mining method employed
G_M	Cut-off grade. The lowest grade set to distinguish economic profitable ore from waste in a given deposit
id	Sequential number of an interval in a borehole
l_s	True length of an metal-grade interval. In a tabular deposit, if θ is the intersection angle between the orebody and the borehole axis, then l_s = original length of an interval × sin θ
g_s	Metal-grade of an interval
l_c	Length of a composite. The sum of l_s of intervals within a composite
g_c	Weighted average grade of a composite. The average of g_s weighted by l_s of every interval within a composite
C_{SUD}^E	Dilutable unminable short economic composite (a set or a class)
C_{SUU}^E	Undilutable unminable short economic composite (a set or a class)
C_D	Diluted composite (a set or a class)
C_D^N	Diluted non-economic composite (a set or a class)
C_D^E	Diluted minable economic composite (a set or a class)
$C_{D-Output}^E$	Economically optimized diluted minable economic composite (a set or a class)
C^W	Waste composite (a set or a class)
$C_{\mathrm{Re}sult}$	Result of compositing metal-grade intervals in a borehole (a set or a class)
C_M^E	Minable economic composite (a set or a class)
C_U^E	Unminable economic composite (a set or a class)
C^U	Non-economic composite (a set or a class)
$f_D(c_{su}^e, s_n)$	Dilution function
$f_{MAX}(SET_{Cd}^E)$	A function used to find the c_{d-Max}^e in a non-empty SET_{Cd}^E
SET_{Cd}	A set storing the $n+1$ elements of C_D in a dilution step

Symbol	Meaning
SET_{Cd}^{N}	A set storing elements of C_D^N in the dilution function
SET_{Cd}^{E}	A set storing elements of C_D^E in the dilution function
σ_D	$\sigma_D = \sum (g_s - G_M) \times l_s$, which represents total net economic profit value of all intervals in a diluted composite

Along the one dimensional top–down track of a borehole, each group of contiguous elements of S^E is composited as a pure economic composite c_p^e. A c_p^e can be classified into either the long economic composite set (C_L^E) or the short economic composite set (C_S^E) by using minimum mining length L_M as criterion. Accordingly, C_L^E and C_S^E are both subsets of the pure economic composite set C_P^E. A short economic composite c_s^e can be classified into either the minable short economic composite set (C_{SM}^E) or the unminable short economic composite set (C_{SU}^E) by using the value of $G_M \cdot L_M$ as criterion. Accordingly, C_{SM}^E and C_{SU}^E are both subsets of the short economic composite set C_S^E.

An unminable short economic composite c_{su}^e can be classified into either the dilutable unminable short economic composite set (C_{SUD}^E) or the undilutable unminable short economic composite set (C_{SUU}^E) by using a dilution function $f_D(c_{su}^e, s_n)$ as criterion. Accordingly, C_{SUD}^E and C_{SUU}^E are both subsets of the unminable short economic composite set C_{SU}^E. A c_{su}^e is classified into C_{SUD}^E means that it can be diluted by the dilution function $f_D(c_{su}^e, s_n)$. In the function, several operations are conducted to add a number of adjacent borehole intervals to c_{su}^e, and thus generate an economically optimized diluted minable economic composite $c_{d-Output}^e$ for the c_{su}^e.

Based on the above descriptions, a looping construct (Fig. 4.3) was set up as the DFM to classify and process economic metal-grade composites c_p^e in a borehole.

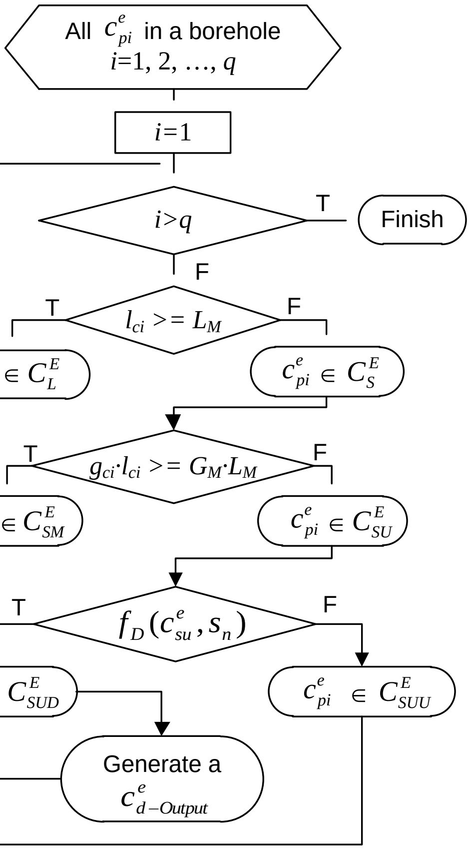


Fig. 4.3. Steps for classifying and processing economic composites in a borehole. In the diagram $l_{ci} = c_{pi}^e.l_c$ and $g_{ci} = c_{pi}^e.g_c$, and T means True and F means False. For meanings of other symbols see Table 4.1.

If each c_{su}^e in a borehole is processed by the dilution function $f_D(c_{su}^e, s_n)$, then all elements of C_L^E, C_{SM}^E, $C_{D-Output}^E$ and C_{SUU}^E in a borehole are identified. The remaining un-composited intervals are, therefore, all elements of S^W. Each group of contiguous elements of S^W is composited as a waste composite c^w. Elements of C_L^E, C_{SM}^E, $C_{D-Output}^E$, C_{SUU}^E and C^W comprise the output result of compositing metal-grade intervals in a borehole.

4.2.2 Dilution procedure and objects involved

In the dilution function $f_D(c_{su}^e, s_n)$, a finite number of external intervals s_n ($n = 1,2,3,..., p$) is added to a c_{su}^e, generating several diluted composites c_d, and then identifying whether these c_d are diluted economic composites c_d^e. If one or more c_d^e can be derived from a c_{su}^e, then the one with the maximum economic profit value can be chosen among them, and a $c_{d-Output}^e$ is derived from it.

Data input to $f_D(c_{su}^e, s_n)$ include a c_{su}^e and s_n, and there are two possible results when $f_D(c_{su}^e, s_n)$ is applied: a c_{su}^e is classified into either C_{SUU}^E or C_{SUD}^E. The intervals s_n include not only elements of S^W but also intervals in element composites of C_L^E, C_{SM}^E, $C_{D-Output}^E$ and C_{SU}^E, meaning that the intervals s_n may also include elements of S^E. Although elements of both S^W and S^E can be added to a c_{su}^e to generate c_d, each added element of S^W would eventually result in a non-economic c_d. Therefore, because all diluted economic composites c_d^e come from c_d, the number (n) of external intervals added to a c_{su}^e should be strictly limited in order to generate economically optimized c_d^e. Accordingly, $f_D(c_{su}^e, s_n)$ is designed as a stepwise function following the increasing number n ($n = 1,2,3,..., p$) of external intervals added. In this stepwise procedure a derived attribute σ was used, which is the total net economic profit value of all borehole intervals in a composite:

$$\sigma = \sum (g_s - G_M) \times l_s = (g_c - G_M) \times l_c .$$

Correspondingly, σ_D was defined as the profit value of a diluted composite (Table 4.1).

In the stepwise procedure of $f_D(c_{su}^e, s_n)$, a certain number (n) of external intervals s_n are added to the c_{su}^e at the n^{th} step. Consequently, there are

$n+1$ dilution choices to generate several c_d (Fig. 4.4), which comprise a set SET_{Cd}. The $n+1$ diluted choices of c_d at the n^{th} step are not suitable for the $(n+1)^{th}$ step, because the number of external intervals added at each step changes. Therefore, SET_{Cd} is cleared for every n^{th} step.

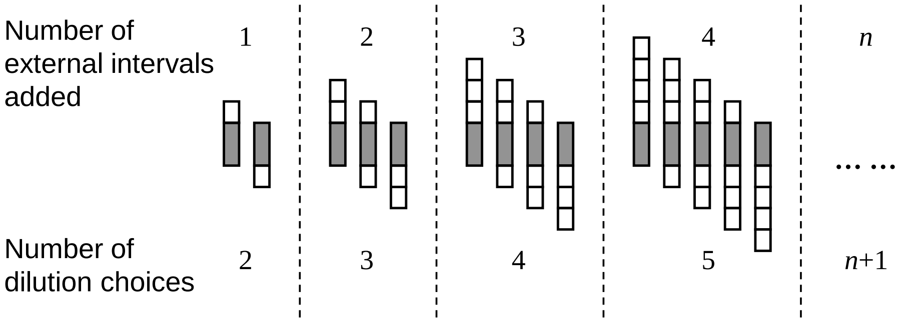


Fig. 4.4. Possible dilution choices in each dilution step following the increase of external intervals. The gray block in the middle is an unminable short economic composite (c_{su}^{e}) and white blocks are external intervals added in the dilution.

For a c_d in a certain SET_{Cd}, if $\sigma_D < 0$ (i.e., $c_d.g_c < G_M$), then this c_d is classified into the diluted non-economic composite C_D^N (e.g., the second choice (counted from left) in dilution step 2 in Fig. 4.5); and in the dilution function this c_d is recorded in a set SET_{Cd}^N. SET_{Cd}^N is not cleared for every n^{th} step, but is used in the following steps. If a c_d of following dilution steps contains a composite c_d^n stored in SET_{Cd}^N, whether its $\sigma_D < 0$ (e.g., the second choice in step 3 in Fig. 4.5) or $\sigma_D \geq 0$ (e.g., the third choice in step 3 in Fig. 4.5), this c_d should also be recorded in SET_{Cd}^N. For a c_{su}^e, if every c_d in the SET_{Cd} of a dilution step is recorded in SET_{Cd}^N (e.g., step 4 in Fig. 4.6), then $f_D(c_{su}^e, s_n)$ returns a $FALSE$ result, and this c_{su}^e is classified into the undilutable unminable short economic composite set C_{SUU}^E.

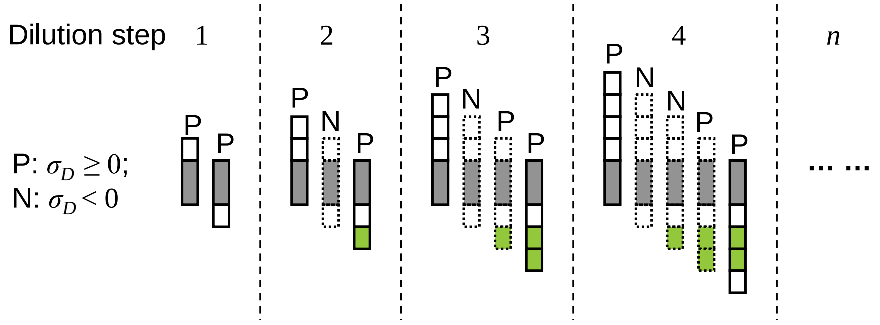


Fig. 4.5. Checking σ_D of each dilution choice c_d. The gray block in the middle is a c_{su}^{e}. White blocks are elements of S^W and green blocks are elements of S^E, representing external intervals added in the dilution procedure. A dotted outline means that a c_d is recorded in SET_{Cd}^{N}. P means a positive or zero σ_D and N means a negative σ_D.

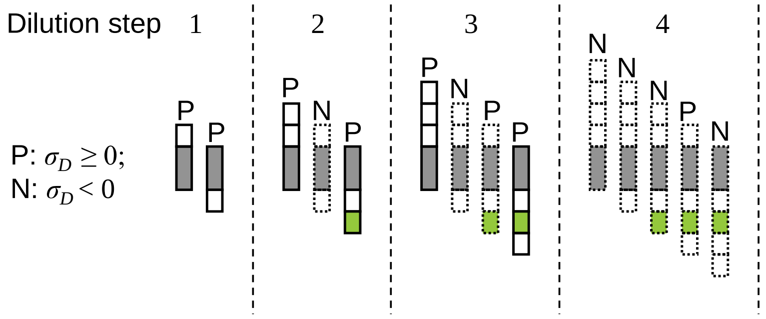


Fig. 4.6. Stopping point of a failed dilution case. The dilution procedure is stopped in step 4 because all resulting composites in this step are recorded in SET_{Cd}^{N}. P means a positive or zero σ_D and N means a negative σ_D.

For a c_d in a SET_{Cd}, if $c_d.g_c \geq G_M$ and $c_d.l_c \geq L_M$, and this c_d does not include any composites in the SET_{Cd}^{N}, then this c_d is classified into diluted minable economic composite C_D^E. In this SET_{Cd} there may be several c_d that fulfill these conditions and thus these c_d are elements of C_D^E that are all recorded in a set SET_{Cd}^{E}. If a SET_{Cd} can result in a non-empty SET_{Cd}^{E}, then a function $f_{MAX}(SET_{Cd}^{E})$ can be used to find a unique composite c_{d-Max}^{e}, which has the maximum value of σ_D, among all element composites in the non-empty SET_{Cd}^{E}.

It is possible that the c_{d-Max}^{e} is the economically optimized diluted minable economic composite $c_{d-Output}^{e}$ for a c_{su}^{e}. That may not be the case if the top and bottom boundaries of a c_{d-Max}^{e} are adjacent to existing composites that are elements of C_{L}^{E}, C_{SM}^{E}, $C_{D-Output}^{E}$ and C_{SU}^{E}, or if the c_{d-Max}^{e} includes intervals that have already been included in those existing composites. Accordingly, a set SET_{Cm} was used to represent such a group of existing composites of C_{L}^{E}, C_{SM}^{E}, $C_{D-Output}^{E}$ and C_{SU}^{E}; and c_m was used to represent elements in SET_{Cm}. If there is a SET_{Cm} related to a c_{d-Max}^{e}, this c_{d-Max}^{e} is first combined with all recorded c_m in the SET_{Cm}, so as to generate a new composite c_{d-New}^{e}, Then, as many as possible waste intervals $s_{Top\&Bottom}^{w}$ at the top and bottom of c_{d-New}^{e} are excluded, while keeping the length of the modified composite not shorter than L_M, so as to generate a $c_{d-Output}^{e}$ for a c_{su}^{e}.

For a certain c_{su}^{e}, if a $c_{d-Output}^{e}$ is derived, then $f_D(c_{su}^{e}, s_n)$ returns a $TRUE$ result, and this c_{su}^{e} is classified into the dilutable unminable short economic composite set C_{SUD}^{E}. SET_{Cd}, SET_{Cd}^{N}, SET_{Cd}^{E} and SET_{Cm} are all cleared at the beginning of each dilution case using $f_D(c_{su}^{e}, s_n)$. Thus, if a c_{su}^{e} is classified into C_{SUD}^{E} during dilution, then at least one element of C_{D}^{E} constituting a unique SET_{Cd}^{E} can be obtained. In the SET_{Cd}^{E} there is a unique c_{d-Max}^{e}, which may result in a unique c_{d-New}^{e}, although each c_{d-Max}^{e} or c_{d-New}^{e} results in a unique $c_{d-Output}^{e}$. Consequently, the result of compositing consists of three super-sets of composites (i.e., minable economic composite set C_{M}^{E}, unminable economic composite set C_U^E , and non-economic composite set C^U). C_M^E includes C_L^E , C_{SM}^E , and $C_{D-Output}^E$; C_U^E includes C_{SUU}^E ; and C^U includes C^W . Based on the above descriptions, a flow chart (Fig. 4.7) was set up as the DFM to dilute a c_{su}^e .

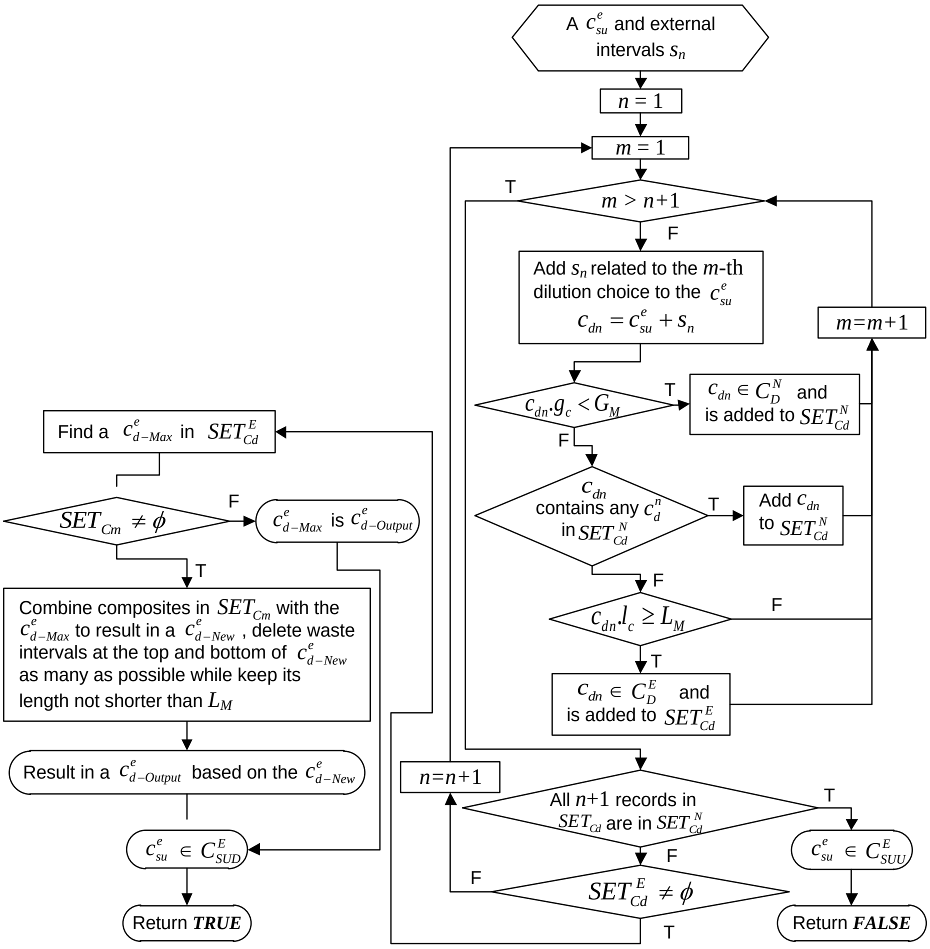


Fig. 4.7. Steps in dilution of an unminable short economic composite c_{su}^e . T means True and F means False.

4.2.3 Object-oriented models

By expressing the DFM in Unified Modeling Language (UML), OOM of classes and sets of borehole metal-grade intervals and composites, and inter-relationships among those classes and sets were defined (Fig. 4.8). Compared to the DFM, the OOM are concise as they omit steps in the compositing procedure but address the classes, states, and characteristics of different types of intervals and composites and the relationships between them.

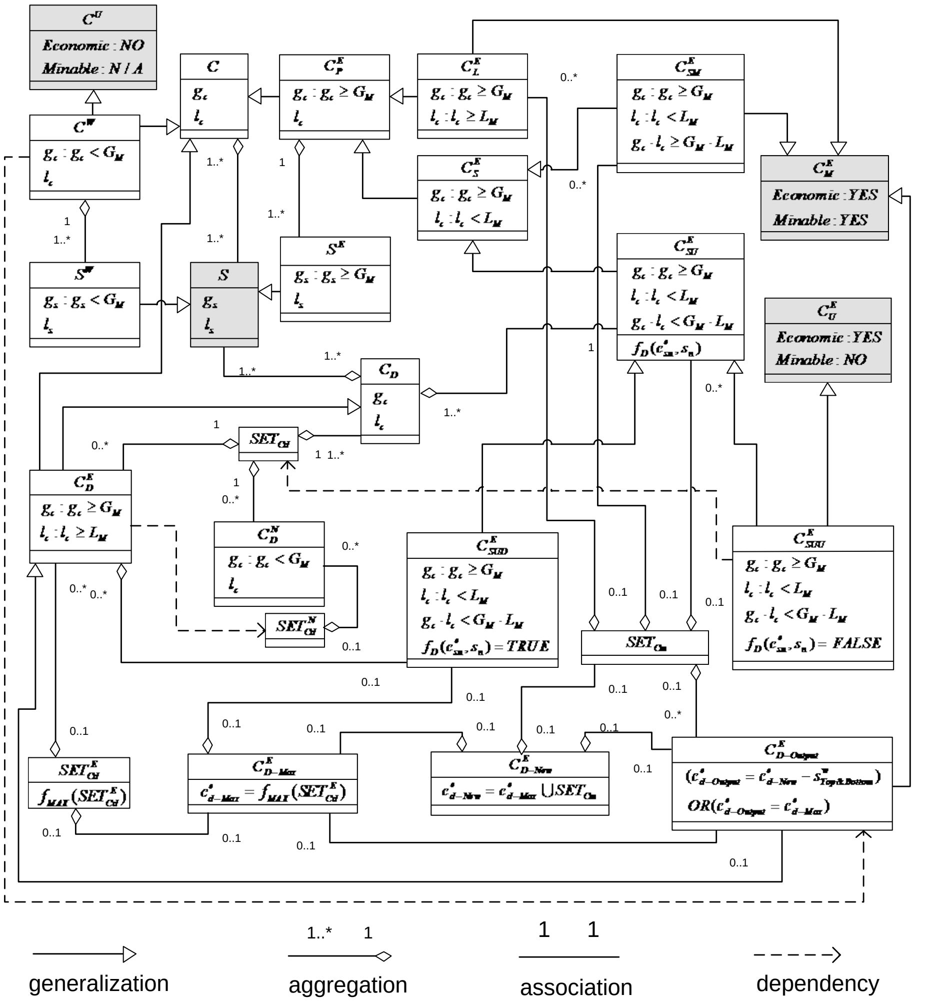


Fig. 4.8. Objects and their relationships in compositing of borehole metal-grade intervals. Boxes in gray color show classes of input data (i.e., S) and output data (i.e., C_M^E, C_U^E and C^U) in the compositing.

Relationships of "class-subclass" (e.g., S^E is a subclass of S) are denoted as "generalization" in Fig. 4.8. Inferential relationships between C^E_{SUU} and SET_{Cd} and between C^E_D and SET^N_{Cd} are denoted as "dependency". The notation "dependency" is also used to represent relationship between C^W and $C^E_{D-Output}$, which however is different from that between C^E_{SUU} and SET_{Cd} or between C^E_D and SET^N_{Cd}. That is because the "dependency" between C^W and $C^E_{D-Output}$ applies to overall instances but not to individual instances of C^W and $C^E_{D-Output}$ in a borehole (i.e., instances of C^W in a borehole can only be identified after all instances of $C^E_{D-Output}$ in that borehole have been determined), whereas the "dependency" between C^E_{SUU} and SET_{Cd} or between C^E_D and SET^N_{Cd} applies to individual instances of C^E_{SUU} and SET_{Cd} or C^E_D and SET^N_{Cd}.

Relationships between SET_{Cm} and C^E_L, C^E_{SM}, $C^E_{D-Output}$ and C^E_{SU} are denoted as "aggregation", because C^E_L, C^E_{SM}, $C^E_{D-Output}$ and C^E_{SU} are not subclasses of SET_{Cm}, but their instances are recorded in the SET_{Cm} set. Furthermore, relationships between SET^E_{Cd} and C^E_{D-Max}, between C^E_{D-Max} and C^E_{D-New}, and between C^E_{D-New} and $C^E_{D-Output}$ are also denoted as "aggregation". The relationship between C^E_{D-Max} and $C^E_{D-Output}$ is denoted as "association", because although a $c^e_{d-Output}$ derives from a c^e_{d-Max}, it is possible that a c^e_{d-Max} is a $c^e_{d-Output}$, or that only a part of a c^e_{d-Max} is included in the derived $c^e_{d-Output}$, which is due to the exclusion of waste intervals $s^w_{Top\&Bottom}$ at the top and bottom of a corresponding c^e_{d-New}.

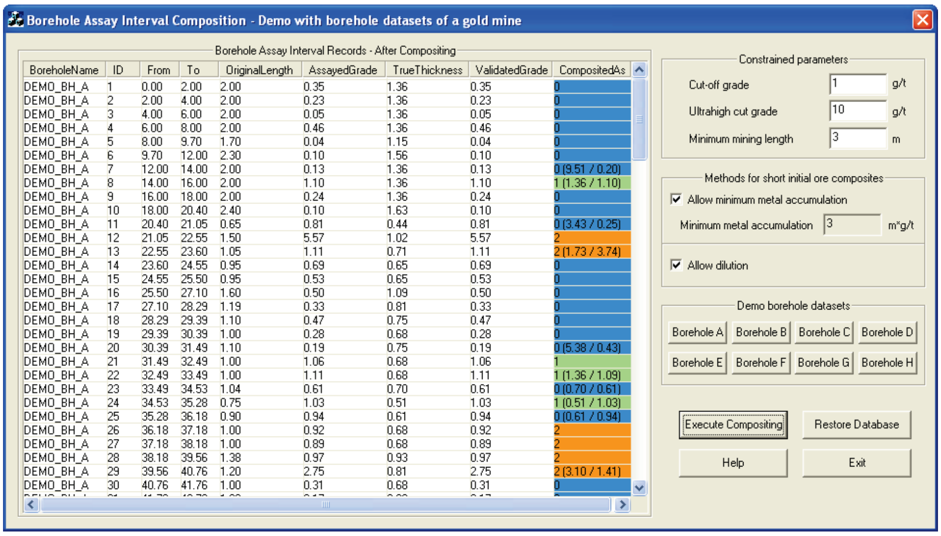
Notations of multiplicity (i.e., zero or more instances "0..*", one or more instances "1..*", zero or one instance "0..1", and exactly one instance "1") are also used in the UML diagrams (Fig. 4.8) . For example, notations of "1" to “1..*”, “1..*” to “1”, and “1” to “0..*” are used, respectively, between C_{SU}^{E} and C_D, between C_D and SET_{Cd}, and between SET_{Cd} and C_D^E, showing that an instance of C_{SU}^{E} may result in zero or more instances of C_D^E. Notations of “0..*” to “0..1” are used between C_D^E and C_{SUD}^{E}, which indicates that if an instance of C_{SU}^{E} cannot result in any instance of C_D^E, then it is not a C_{SUD}^{E}.

SET_{Cm} is a subset of the union of instances of C_L^E, C_{SM}^{E}, $C_{D-Output}^{E}$ and C_{SU}^{E}. In Fig. 4.8, “aggregation” is used to show the relationships from C_L^E, C_{SM}^{E}, $C_{D-Output}^{E}$ and C_{SU}^{E} to SET_{Cm}, then “0..*” is used at the side of C_L^E, C_{SM}^{E}, $C_{D-Output}^{E}$ and C_{SU}^{E}, and “0..1” is used at the side of SET_{Cm}. Here the “0..1” means that, depends on a certain c_{d-Max}^{e}, there may or may not exist an instance of SET_{Cm}, and the “0..*” indicates that in an instance of SET_{Cm} there may be zero or more instances of C_L^E, C_{SM}^{E}, $C_{D-Output}^{E}$ and C_{SU}^{E}.

Notations used between S^W and C^W are “1..*” to “1” because an instance of C^W includes at least one instance of S^W, and conversely, one or more instances of S^W can be included in exactly one instance of C^W. This explanation also applies to the notations between S^E and C_P^E. Nevertheless, the notations used between S and C are “1..*” to “1..*” because each instance of S may be included in one or more instances of C. This is caused by dilution in the compositing procedure. For example, by adding external intervals (i.e., s_n) to a c_{su}^{e}, an instance of S may be included in several instances of C_D. Thus the notations of “1..*” to “1..*” are also used for the relationship between S and C_D.

4.3 Pilot system and results

The aforementioned DFM and OOM were applied in the development of a pilot system by using C++ (Fig. 4.9). Compositing can be performed via a user interface consisting of four parts: constrained parameters setting, demo borehole choosing, control buttons, and compositing result. Two methods are set optional: (1) allowing minimum metal accumulation (i.e., allowing minable short economic composites c_{sm}^{e} in the result) and (2) allowing dilution.


Borehole Assay Interval Composition - Demo with borehole datasets of a gold mine

Borehole Assay Interval Records - After Compositing

BoreholeName	ID	From	To	OriginalLength	AssayedGrade	TrueThickness	ValidatedGrade	CompositedAs
DEMO_BH_A	1	0.00	2.00	2.00	0.35	1.36	0.35	0
DEMO_BH_A	2	2.00	4.00	2.00	0.23	1.36	0.23	0
DEMO_BH_A	3	4.00	6.00	2.00	0.05	1.36	0.05	0
DEMO_BH_A	4	6.00	8.00	2.00	0.46	1.36	0.46	0
DEMO_BH_A	5	8.00	9.70	1.70	0.04	1.15	0.04	0
DEMO_BH_A	6	9.70	12.00	2.30	0.10	1.56	0.10	0
DEMO_BH_A	7	12.00	14.00	2.00	0.13	1.36	0.13	0 (9.51 / 0.20)
DEMO_BH_A	8	14.00	16.00	2.00	1.10	1.36	1.10	1 (1.36 / 1.10)
DEMO_BH_A	9	16.00	18.00	2.00	0.24	1.36	0.24	0
DEMO_BH_A	10	18.00	20.40	2.40	0.10	1.63	0.10	0
DEMO_BH_A	11	20.40	21.05	0.65	0.81	0.44	0.81	0 (3.43 / 0.25)
DEMO_BH_A	12	21.05	22.55	1.50	5.57	1.02	5.57	2
DEMO_BH_A	13	22.55	23.60	1.05	1.11	0.71	1.11	2 (1.73 / 3.74)
DEMO_BH_A	14	23.60	24.55	0.95	0.69	0.65	0.69	0
DEMO_BH_A	15	24.55	25.50	0.95	0.53	0.65	0.53	0
DEMO_BH_A	16	25.50	27.10	1.60	0.50	1.09	0.50	0
DEMO_BH_A	17	27.10	28.29	1.19	0.33	0.81	0.33	0
DEMO_BH_A	18	28.29	29.39	1.10	0.47	0.75	0.47	0
DEMO_BH_A	19	29.39	30.39	1.00	0.28	0.68	0.28	0
DEMO_BH_A	20	30.39	31.49	1.10	0.19	0.75	0.19	0 (5.38 / 0.43)
DEMO_BH_A	21	31.49	32.49	1.00	1.06	0.68	1.06	1
DEMO_BH_A	22	32.49	33.49	1.00	1.11	0.68	1.11	1 (1.36 / 1.09)
DEMO_BH_A	23	33.49	34.53	1.04	0.61	0.70	0.61	0 (0.70 / 0.61)
DEMO_BH_A	24	34.53	35.28	0.75	1.03	0.51	1.03	1 (0.51 / 1.03)
DEMO_BH_A	25	35.28	36.18	0.90	0.94	0.61	0.94	0 (0.61 / 0.94)
DEMO_BH_A	26	36.18	37.18	1.00	0.92	0.68	0.92	2
DEMO_BH_A	27	37.18	38.18	1.00	0.89	0.68	0.89	2
DEMO_BH_A	28	38.18	39.56	1.38	0.97	0.93	0.97	2
DEMO_BH_A	29	39.56	40.76	1.20	2.75	0.81	2.75	2 (3.10 / 1.41)
DEMO_BH_A	30	40.76	41.76	1.00	0.31	0.68	0.31	0

Constrained parameters

Cut-off grade 1 g/t

Ultrahigh cut grade 10 g/t

Minimum mining length 3 m

Methods for short initial ore composites

☑ Allow minimum metal accumulation

Minimum metal accumulation 3 m*g/t

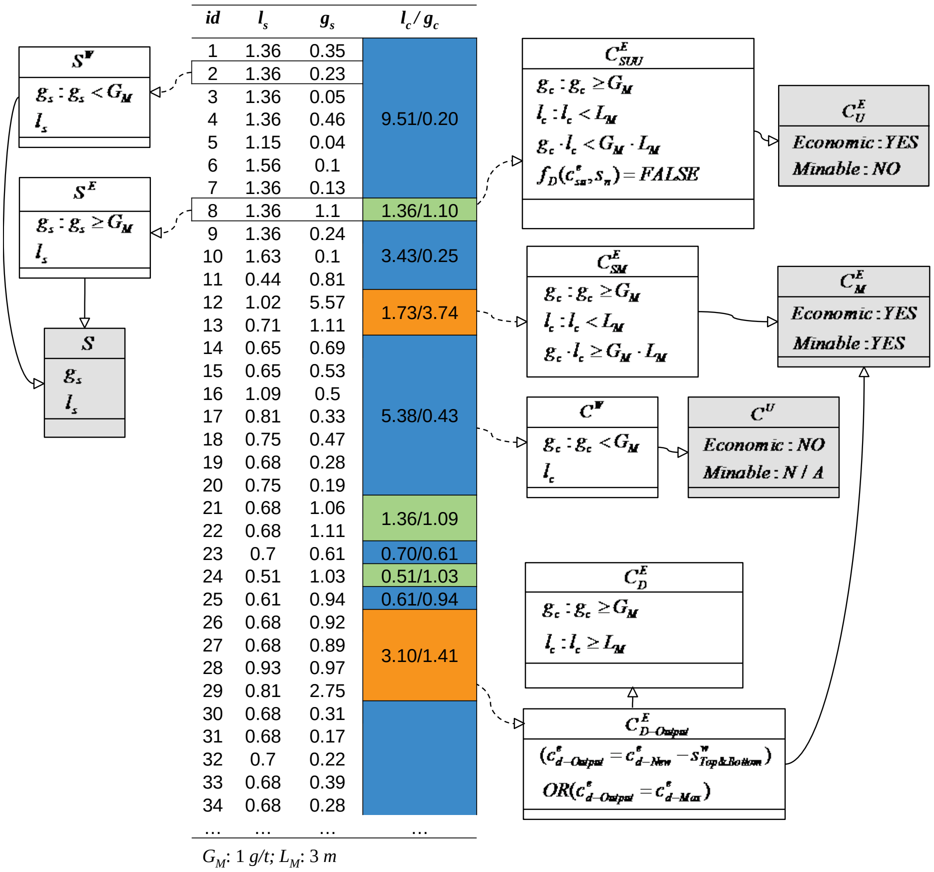
☑ Allow dilution

Demo borehole datasets

Fig. 4.9. User interface of a pilot system. Compositing results are shown in column "CompositedAs": Label "2" and red filling color stand for minable economic composite C_M^E; Label "1" and green filling color stand for unminable economic composite C_U^E; and label "0" and blue filling color stand for non-economic composite C^U. Total length and average grade of each composite are given at the end interval of a composite.

The developed program was tested using eight borehole dataset from a gold mine and desired results were obtained as designed in the DFM and OOM. For the case study data (Fig. 4.10), the defined composites constitute three superclasses C_M^E, C_U^E and C^U. However, in actual practice of reserves/resources estimation, data attributes of intervals comprising every instance of a composite are required for assessing the accuracy and reliability of classifications of reserves/resources (UNECE, 1997, 2001; Dominy, 2002). Data attributes and characteristics of objects in classes and sets recorded in the OOM can be organized in computer programs and can then be used to trace the data attributes of intervals that comprise every instance of a composite. For example, there are two instances of C_M^E in Fig. 4.10, one of them belongs to the subclass C_{SM}^E and the other belongs to the subclass

$C^{E}_{D-Output}$. Although both C^{E}_{SM} and $C^{E}_{D-Output}$ belong to C^{E}_{M}, they may have different weights in the classification of estimates of metal reserves. By using OOM in computer programs for compositing borehole metal-grade intervals, the differences between C^{E}_{SM} and $C^{E}_{D-Output}$ can be recorded.



id	l_s	g_s	l_c/g_c
1	1.36	0.35	9.51/0.20
2	1.36	0.23	
3	1.36	0.05	
4	1.36	0.46	
5	1.15	0.04	
6	1.56	0.1	
7	1.36	0.13	
8	1.36	1.1	1.36/1.10
9	1.36	0.24	3.43/0.25
10	1.63	0.1	
11	0.44	0.81	
12	1.02	5.57	1.73/3.74
13	0.71	1.11	
14	0.65	0.69	5.38/0.43
15	0.65	0.53	
16	1.09	0.5	
17	0.81	0.33	
18	0.75	0.47	
19	0.68	0.28	
20	0.75	0.19	
21	0.68	1.06	1.36/1.09
22	0.68	1.11	
23	0.7	0.61	0.70/0.61
24	0.51	1.03	0.51/1.03
25	0.61	0.94	0.61/0.94
26	0.68	0.92	3.10/1.41
27	0.68	0.89	
28	0.93	0.97	
29	0.81	2.75	
30	0.68	0.31	
31	0.68	0.17	
32	0.7	0.22	
33	0.68	0.39	
34	0.68	0.28	
…	…	…	…

G_M: 1 g/t; L_M: 3 m

Fig. 4.10. Objects of metal-grade intervals and composites in actual borehole data. The column with head l_c / g_c shows the output of compositing, in which the records filled with red, green and blue colors are composites respectively belonging to superclasses C^{E}_{M}, C^{E}_{U} and C^{U} (see Fig. 4.9).

The OOM can also be used to explain objects involved in a certain dilution case. For example, in Fig. 4.11 there is an instance of C^{E}_{SU}. In the first step of the dilution function $f_D(c^{e}_{su}, s_n)$, only one external interval (i.e., $n = 1$) is added and there are two dilution choices. Because the two dilution choices are both C_D^N, this dilution case returns a $FALSE$ result. In contrast, in Fig. 4.12, another dilution case returns a $TRUE$ result. There are three instances of C_D^E among four dilution choices in the third step (i.e., $n = 3$) of the dilution function. These three instances of C_D^E constitute an instance of SET_{Cd}^E, in which an instance of C_{D-Max}^E is chosen, which in turn becomes an instance of $C_{D-Output}^E$.

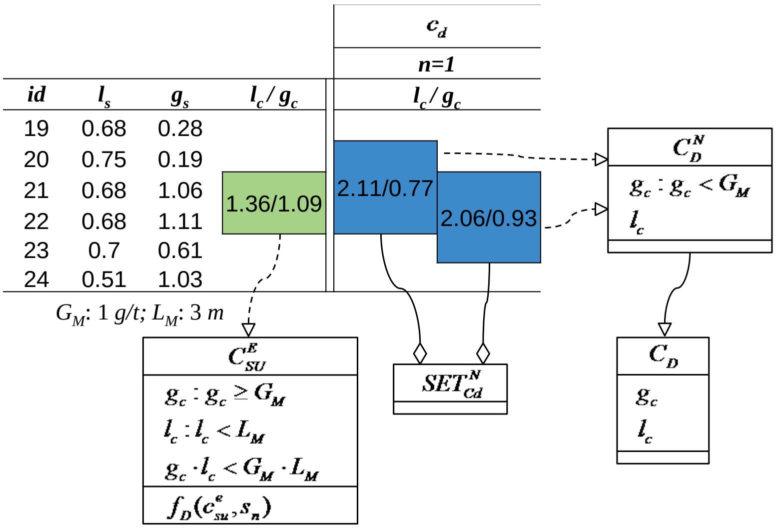


Fig. 4.11. Objects of metal-grade intervals and composites in a dilution case that returns a $FALSE$ result. Input data of the dilution case is taken from Fig. 4.10.

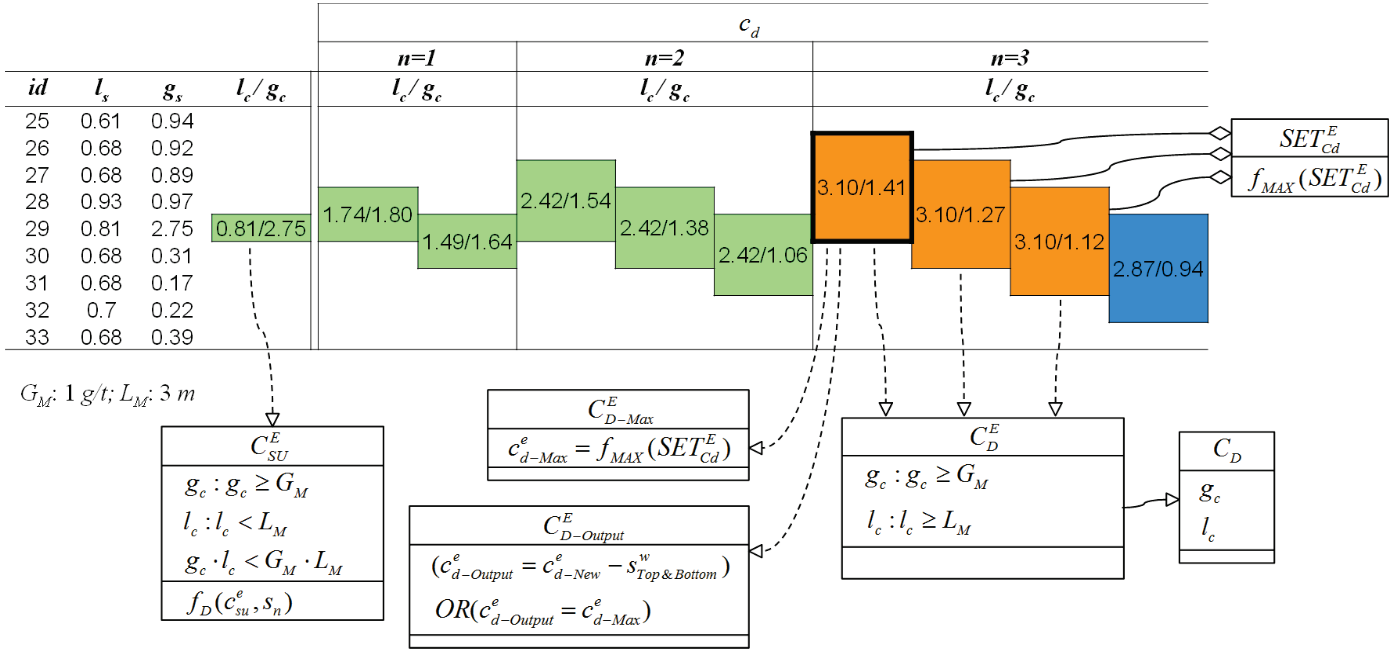


Fig. 4.12. Objects of metal-grade intervals and composites in a dilution case that returns a *TRUE* result. The record with red filling color and bolded outline in step 3 is the output result of this dilution case. Input data of the dilution case is taken from Fig. 4.10.

4.4 Discussion

In the DFM, steps in the procedure of compositing are defined, which are a solid support for transforming the procedure into a computer program. In the OOM, classes and sets of intervals and composites are defined, and inter-relationships among them are identified. DFM and OOM complement each other in the aforementioned work of developing a program for compositing borehole metal-grade intervals.

The DFM show a stepwise procedure, in which each step consists of three parts: input data, data processing and output data. The output data of a foregoing step is often a part of the input data of a following step. A foregoing step may generate diverse output data, and these may lead to several branches of following steps. The DFM specify certain stopping points for each branch in the stepwise procedure when desired results are generated. Such DFM concentrate on the sequence of steps, the method for data processing and the desired result of each step (Ma et al., 2010a). In contrast, the OOM concentrate on the identification of objects and their relationships. Intervals and composites in the compositing are classified into classes, subclasses, as well as sets and subsets. Relationships among classes, sets and instances of classes and sets, such as generalization, aggregation, association and dependency, are identified with UML in Fig. 4.8. Such characteristics of OOM provide a clear view of objects and their relationships in the compositing, and thus are complementary to DFM. A closer relationship between end-users and software developers is the trend for developing mine planning software (Kapageridis, 2009). OOM and DFM show different aspects of a software program and can facilitate discussion between end-users and software developers when a program is under development.

Nowadays, OOM are widely used in software design and development (Lu and Jin, 1997; Booch et al., 2007; Din and Idris, 2009; Sun, 2010). OOM are easy to understand and can be reused or linked to other models (Mylopoulos et al., 1999). For example, in the OOM of this study the objects of minable economic composite C_M^E , unminable economic composite C_U^E and non-economic composite C^U were proposed, which are compatible with the commonly used schemas for mineral resources estimation (Fig. 4.13 and Table 4.2). Nevertheless, each method of computerized analysis and modeling has its own characteristics and limitations (Green and Rosemann, 2000). Therefore, DFM cannot be omitted completely. Although OOM present a concise expression of objects and their relationships in compositing of borehole metal-grade intervals, they are weaker than DFM in explaining steps in the compositing procedure and the derivation of objects in the procedure.

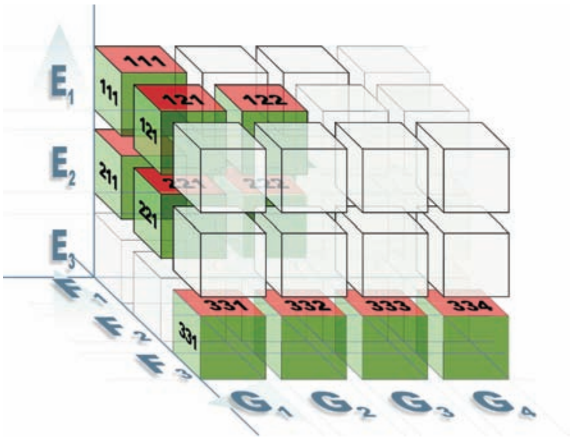


Fig. 4.13. United Nations Framework Classification for Energy and Mineral Resources (UNFC) as applied to coal, uranium and other solid minerals (from UNECE, 1997). For meanings of economic (E1, E2, E3), feasibility (F1, F2, F3) and geological (G1, G2, G3, G4) categories see Table 4.2.

Table 4.2 Categories for coal, uranium and other solid minerals (from UNECE, 1997)

Categories	**Meanings**
E1	Economic
E2	Potentially Economic
E3	Intrinsically Economic
F1	Mining Report and/or Feasibility Study
F2	Pre-feasibility Study
F3	Geological Study
G1	Detailed Exploration
G2	General Exploration
G3	Prospecting
G4	Reconnaissance Study

There is no single method that can address all aspects of modeling objects and their relationships in the compositing of borehole metal-grade intervals. Villa et al. (2009) discussed that knowledge incorporated in models is rarely self-contained enough for them to be understood and used – by humans or machines – without the modeler's mediation. They also proposed that ontology-driven approaches will become more important for formal organization of domain knowledge and for remedying insufficiencies in current models. Studies of an ontology spectrum (McGuinness, 2003; Obrst, 2003; Uschold and Gruninger, 2004) show that higher order logics (e.g., description logic (DL)) allow more expressivity for semantics of objects and their relationships. Precise restrictions can be defined by using DL-based languages (e.g., Resource Description Framework (RDF) or Web Ontology Language (OWL)) and online applications can be developed based on RDF/OWL. Therefore, a possible direction for further studies is to investigate applications of RDF/OWL in building conceptual models for compositing borehole metal-grade intervals.

In this study, the DFM and OOM follow the principle of Boolean logic (i.e., every meaningful proposition is either true or false, and nothing in between). For example, an instance s of S belongs to either S^E or S^W. In contrast, fuzzy logic (FL) permits intermediate values between completely true and completely false for a certain proposition (Zadeh, 1965, 2008). Tutmez (2007) proposed a FL-based methodology for estimation of ore grades and uncertainties of estimates. Although the application of FL in Tutmez (2007) is different from the general idea of Boolean logic in this study (i.e., to identify economic metal-grade borehole intervals and composites and their inter-relationships), the advantage of FL for calculating uncertainties in ore grade estimates suggests further work to investigate how FL can be used to supplement or complement the current study for generating models of borehole metal-grade composites.

4.5 Conclusions

Integrating different methods of conceptual analysis can generate detailed conceptual schemas/models for a subject in geology, such as the compositing of borehole metal-grade intervals. As boreholes and borehole intervals of a mine site are numerous, it is desirable to have a computer program that can generate standard-compatible results in borehole metal-grade interval compositing. In the study presented in this chapter, data-flow and object-oriented models were developed to represent the procedure of compositing borehole metal-grade intervals. A pilot system was then developed to implement the designed models and the program was tested using borehole datasets from a gold mine. This study shows that object-oriented models and data-flow models complement each other in developing the program for compositing borehole metal-grade intervals. By using the developed models and programs, results of composited borehole metal-grade intervals are compatible with commonly used standards in the field of mineral resources estimation. Since approaches driven by higher order logics are potentially useful for remedying insufficiencies of current data-flow and object-oriented models, investigating the application of higher order logic languages may prove useful in the conceptual analysis for compositing borehole metal-grade intervals. In addition, investigating the application of fuzzy logic may also prove useful to complement the current models.

It is also noteworthy that as a part of the cross-sectional method, the compositing method discussed here assumes that values (e.g., assayed metal-grade and true length) of borehole intervals to be composited are free of errors. It tends to be applied in the initial modeling of orebodies in a mineral deposit, without comprehensive and precise consideration of distribution and covariance of samples as what geostatistics does. Therefore, in order to avoid misleading results, a reliable estimation of mineral resources may apply several methods and make compare between the separate results. The works of conceptual analyses and conceptual schemas/models in this chapter are used in the local contexts of mine sites. Chapter 5 will shift the context to the cyber-infrastructure and will describe the featured functions developed with a RDF/OWL-based ontology of geological time scale for promoting the interoperability of online geological data.

Chapter 5

Ontology-aided management of information from online geological data

This chapter is based on: Ma, X., Carranza, E.J.M., Wu, C., van der Meer, F.D., Ontology-aided annotation, visualization and generalization of geological time scale information from online geological map services. Computers & Geosciences (2011), doi:10.1016/j.cageo.2011.07.018

5.1 Introduction

Cyber-infrastructure enables faster and easier creation, storage and transfer of data, yet services facilitating efficient information retrieval and knowledge discovery are still underdeveloped (Hey and Trefethen, 2005; Stafford, 2010). In the field of geology, it has been extensively discussed that a geoscience cyber-environment includes not only digitized geological data but also expertise and tools that support the transformation of data to knowledge (Brodaric and Gahegan, 2006; Howard et al., 2009; Loudon, 2009; McGuinness et al., 2009). Such services of expertise and tools are useful for studies of geology within a cyber-environment and, more importantly, they provide supports for addressing geology-related societal challenges, such as resources exploration, urban development and hazards mitigation, etc., in the context of the cyber-infrastructure (Broome, 2005; OneGeology-Europe Consortium, 2010; Sinha et al., 2010).

Ontologies, as shared conceptualizations of domain knowledge (Gruber, 1995; Guarino, 1997b), can help to improve the interoperability of geological data and facilitate the transformation of geological data into geological knowledge in the cyber-infrastructure (Loudon, 2000; Brodaric and Gahegan, 2006; Galton, 2009; Reitsma et al., 2009). There are several forms of geological ontology with varying semantic richness (i.e., precision of meanings of concepts and relationships between concepts). Following a general direction from informal to formal semantics, geological ontologies include controlled vocabularies (e.g., Bibby, 2006; Richard and Soller, 2008; Ma et al., 2010b), conceptual schemas (e.g., Brodaric, 2004; NADM Steering

Committee, 2004; Richard, 2006) and RDF[46]/OWL[47]-based ontologies (e.g., Ludäscher et al., 2003; Raskin and Pan, 2005; Tripathi and Babaie, 2008), etc.

In several recent projects, ontologies have been applied to provide featured functions in geological data infrastructures, thereby promoting services of geological data and tools that support information retrieval and knowledge discovery. In the GEON project (www.geongrid.org), ontologies were used to mediate conceptual schemas of heterogeneous geological maps and enable semantic integration among them (Ludäscher et al., 2003; Baru et al., 2009). The AuScope project (www.auscope.org.au) built vocabulary-based services for querying geological maps, which overcame differences in geoscience terms due to language, spelling, synonyms and local variations and, thus, helped users to find desired information (Woodcock et al., 2010). The OneGeology (1G) project (www.onegeology.org) promoted the GeoSciML (Sen and Duffy, 2005) as a common conceptual schema, which improved the interoperability of online geological maps distributed globally (Jackson, 2007). GeoSciML was also applied in the OneGeology-Europe (1G-E) project (www.onegeology-europe.eu) and, compared to the 1G, the 1G-E extended vocabulary-based services and enabled multilingual annotation and translation of geological map contents among 18 European languages (Asch et al., 2010; Laxton et al., 2010).

Through the aforementioned projects, substantial developments have been made in conceptualizing geological knowledge into ontologies and using defined ontologies to mediate and/or integrate heterogeneous geological data. However, services using ontologies to support the interpretation of geological data are still underdeveloped. Provision of those services is necessary, nevertheless, because they are vital for comprehending the usability (i.e., as an essential part of interoperability (Bishr, 1998; Harvey et al., 1999)) of geological data served in a data infrastructure. Services using ontologies enable users, especially those who are not familiar with geology, not only to find desired data but also to understand and use the data appropriately (cf. Broome, 2005; Bond et al., 2007; Gahegan et al., 2009).

In the present study, an ontology of geological time scale is applied to support annotation, visualization, filtration and generalization of geological time scale (GTS) information from online geological map services. The

---

[46] Resource Description Framework. http://www.w3.org/RDF [Accessed February 04, 2011].

[47] Web Ontology Language. http://www.w3.org/2004/OWL [Accessed February 04, 2011].

present study aimed to: (1) show methods of using proper datatype and object properties to represent the structure of a domain (i.e., GTS) in geosciences; (2) develop functions of ontology-based annotation and visualization to help users to understand GTS contents of online geological maps; (3) develop ontology-based interactive functions to help users retrieve GTS information and discover GTS knowledge in online geological maps; and, as a whole, (4) show a novel way of using ontologies to improve geological data interoperability and facilitate geological knowledge discovery in the context of the Semantic Web.

5.2 Building and visualizing a GTS ontology

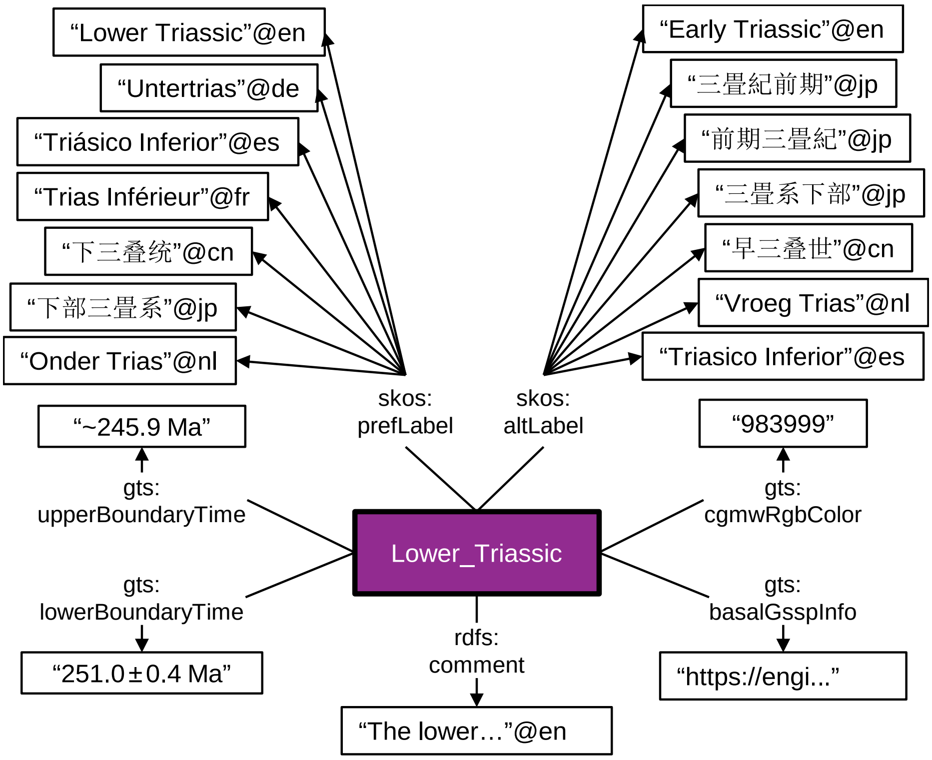
5.2.1 Incorporating annotations in a GTS ontology

The GTS ontology was developed with a Resource Description Framework (RDF) model (Fig. 5.1a). Properties used in the ontology include two parts: datatype properties and object properties. The former are used to define differentiating qualities of concepts (Fig. 5.1b) and the latter are used to define relationships between concepts (Fig. 5.1c). All GTS concepts are defined as instances of GTS classes in the current ontology, including “Supereonothem”, “Eonothem”, “Erathem”, “System”, “Subsystem”, “Series” and “Stage”. “Supereonothem” and “Subsystem” are special, because the former is only used to define the GTS concept “Precambrian” whereas the latter is only used to define GTS concepts “Pennsylvanian” and “Mississippian”. The other five classes follow the common chronostratigraphic units used globally. The example concept “Lower_Triassic” in Fig. 5.1 is an instance of the class “Series”.

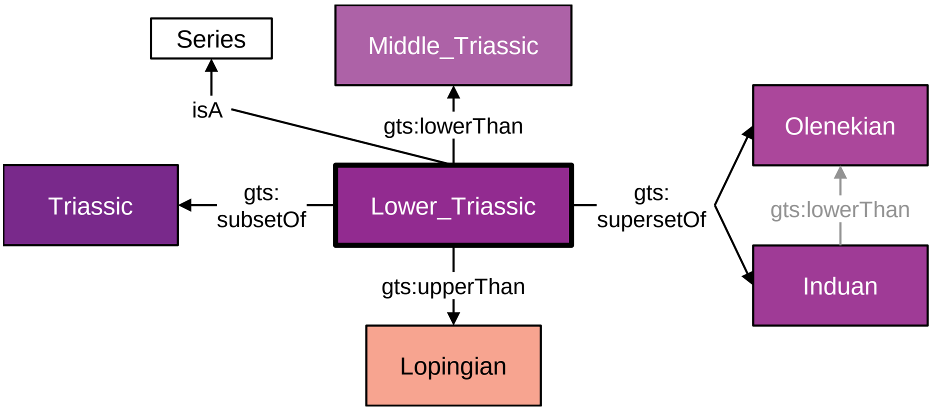
```
<gts:Series rdf:ID="Lower_Triassic">
    <skos:prefLabel xml:lang="en">Lower Triassic</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Untertrias</skos:prefLabel>
    <skos:prefLabel xml:lang="es">Triásico Inferior</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">Trias Inférieur</skos:prefLabel>
    <skos:prefLabel xml:lang="cn">下三叠统</skos:prefLabel>
    <skos:prefLabel xml:lang="jp">下部三畳系</skos:prefLabel>
    <skos:prefLabel xml:lang="nl">Onder Trias</skos:prefLabel>
    <skos:altLabel xml:lang="en">Early Triassic</skos:altLabel>
    <skos:altLabel xml:lang="jp">三畳紀前期</skos:altLabel>
    <skos:altLabel xml:lang="jp">前期三畳紀</skos:altLabel>
    <skos:altLabel xml:lang="jp">三畳系下部</skos:altLabel>
    <skos:altLabel xml:lang="cn">早三叠世</skos:altLabel>
    <skos:altLabel xml:lang="nl">Vroeg Trias</skos:altLabel>
    <skos:altLabel xml:lang="es">Triasico Inferior</skos:altLabel>
    <rdfs:comment xml:lang="en">The lower series of the Triassic System of the Standard
Global Chronostratigraphic Scale, above the Permian System of the Paleozoic Erathem and below
the Middle Triassic Series. Also the time during which these rocks were formed, the Middle
Triassic Epoch.</rdfs:comment>
    <gts:cgmwRgbColor>983999</gts:cgmwRgbColor>
    <gts:subsetOf rdf:resource="#Triassic"/>
    <gts:supersetOf rdf:resource="#Olenekian"/>
    <gts:supersetOf rdf:resource="#Induan"/>
    <gts:lowerThan rdf:resource="#Middle_Triassic"/>
    <gts:upperThan rdf:resource="#Lopingian"/>
    <gts:upperBoundaryTime>~245.9 Ma</gts:upperBoundaryTime>
    <gts:lowerBoundaryTime>251.0±0.4 Ma</gts:lowerBoundaryTime>
    <gts:basalGsspInfo>
https://engineering.purdue.edu/stratigraphy/gssp/detail.php?periodid=76-top_parentid=35
[Subcommission for Stratigraphic Information of ICS, 2010, GSSP Table]</gts:basalGsspInfo>
</gts:Series>
```

(a) Source code.

Fig. 5.1. Definition of “Lower_Triassic” as an instance of “Series” in a GTS ontology. (b) and (c) (on the next page) present graphic views, respectively, of datatype properties and object properties of “Lower_Triassic” defined in (a). The “lowerThan” relationship (in gray color) between “Induan” and “Olenekian” are not defined in (a), but are included in the definition of the GTS concept “Induan”.



(b) Graphic view of datatype properties.



(c) Graphic view of object properties.

According to the International Stratigraphic Chart[48], there is a hierarchal structure among GTS concepts. Meanwhile, because the time boundaries of

---

[48] https://engineering.purdue.edu/Stratigraphy/charts/chart.html [Accessed February 04, 2011].

GTS concepts form a continuous temporal sequence, there is also an ordinal structure among these GTS concepts (Cox and Richard, 2005; Michalak, 2005). The hierarchical structure among GTS concepts was encoded with two object properties "gts:supersetOf" and "gts:subsetOf" (Fig. 5.1c), and the ordinal structure was encoded with other two object properties "gts:lowerThan" and "gts:upperThan". Two datatype properties "gts:upperBoundaryTime" and "gts:lowerBoundaryTime" were used to record the time boundaries of each GTS concept (Fig. 5.1b). With these definitions, the developed ontology represents the ordinal hierarchical structure of GTS.

Two SKOS[49] datatype properties "skos:prefLabel" and "skos:altLable" in Fig. 5.1b were used to encode preferred and alternative labels of the concept "Lower_Triassic". These multilingual labels were adopted from a previous work of a SKOS-based thesaurus of GTS (Ma et al., 2011; also see Chapter 3). Because the SKOS model is compatible with the RDF, the two properties "skos:prefLabel" and "skos:altLable" can be imported directly into the GTS ontology discussed here. Another datatype property "gts:cgmwRgbColor" was used to encode the related RGB (red-green-blue) code (in hexadecimal format) of a GTS concept (Fig. 5.1b). These RGB codes of GTS concepts are specified by the CGMW[50] and are used in the International Stratigraphic Chart (see footnote 48). The standard RGB codes of CGMW were followed in the study presented in this chapter to improve the compatibility of the developed GTS ontology.

Recent studies (Lumb et al., 2009; Reitsma, 2010) have shown that including commonly accepted explanations of concepts as annotations in an ontology enhances the compatibility of that ontology and those annotations are useful for explaining and interpreting geoscience data. Two datatype properties "rdfs:comment" and "gts:basalGsspInfo" were used to encode annotations in the developed GTS ontology (Figs. 5.1a, b). The property "rdfs:comment" recorded explanations of GTS concepts. Definitions of most GTS concepts were retrieved from the Glossary of Geology, 5th edition (Neuendorf et al., 2005), which is a reliable resource for explanations of geological concepts. For some concepts not defined in that glossary, explanations were edited for them (e.g., the explanation of "Lower_Triassic" recorded by "rdfs:comment" in Fig. 5.1) following the style of the glossary. The other property "gts:basalGsspInfo" recorded webpages of basal GSSP (Global Boundary Stratotype Section and Point) information of GTS concepts. The website of

---

[49] Simple Knowledge Organization System. http://www.w3.org/2004/02/skos [Accessed February 04, 2011].

[50] Commission for the Geological Map of the World. http://www.cgmw.net [Accessed February 04, 2011].

the Subcommission for Stratigraphic Information of the International Commission on Stratigraphy [51] was referred for reliable ratified GSSP information.

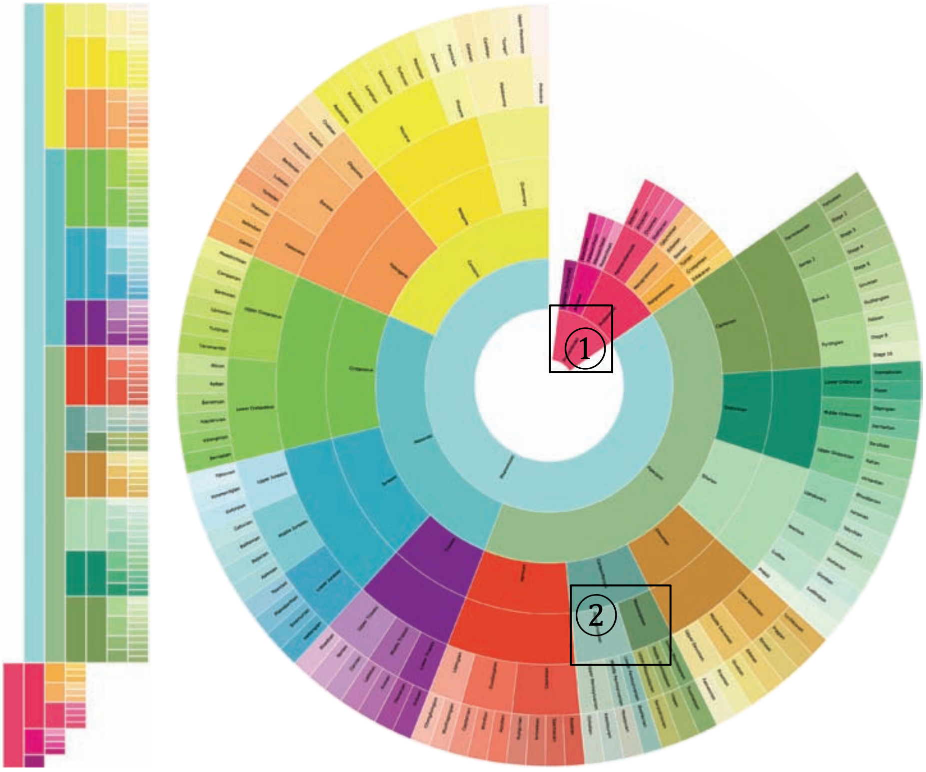
By representing the ordinal hierarchical structure, collecting preferred and alternative multilingual labels, and organizing reliable annotations of GTS concepts in the GTS ontology, a stable basis was set up for the following works of building a GTS ontology-based animation and developing interactive functions between the ontology, animation and online geological map services.

5.2.2 An animation based on developed GTS ontology

Animation is an interactive way of interpreting geological data and conveying geological knowledge (Kulawiak et al., 2010; Reitsma, 2010). Because the GTS is a global reference system of time used in geological maps, building an animation for the GTS and then using it to complement online geological maps may lead to more interactive functions and, thus, improve the process of geological data retrieval and geological knowledge discovery. Intuitively, visualizing concepts and relationships in the developed RDF-based GTS ontology can set up a framework for the GTS animation discussed here.

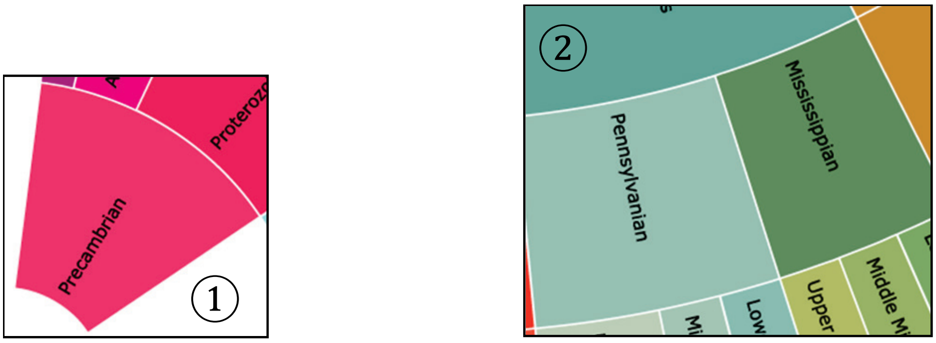
---

[51] https://engineering.purdue.edu/Stratigraphy [Accessed February 05, 2011].



(a) *Rooted tree.*

(b) *Radial tree.*



(c) *Node of "Supereonothem" concept.*

(d) *Nodes of "Subsystem" concepts.*

Fig. 5.2. Layout of the developed GTS animation with details of two parts. "Precambrian" is the only node at the level of "Supereonothem", as shown in (b) and (c), and "Mississippian" and "Pennsylvanian" are the only two nodes with names at the level of "Subsystem", as shown in (b) and (d).

There are vast methods and techniques for visualizing ontologies (Katifori et al., 2007; Krivov et al., 2007). The ordinal hierarchical structure of the developed GTS ontology requires both an intuitive layout and an efficient space usage on a user interface. Among the commonly used layouts of hierarchical visualizations (e.g., rooted tree, radial tree, balloon tree, and tree-map, etc.) (Holten, 2006), the rooted tree and the radial tree were chosen for developing the GTS animation in order to achieve those requirements. Web-based visualizations and animations can be realized with various technologies, such as JavaScript, SVG (Scalable Vector Graphics) and Flash, etc., for many of which open-source libraries are available on the Web (D'Ambros et al., 2010). SVG (e.g., Ipfelkofer et al., 2006) and JavaScript (e.g., Ma et al., 2011) have already been used to visualize ontologies and interact with online maps. Although Flash has been used to publish online maps (e.g., Kraak, 2004; Youn et al., 2008) and to visualize ontologies (e.g., Geroimenko and Geroimenko, 2006), the application of Flash-based ontologies and interactions with online map services is underdeveloped. Therefore, Flash was chosen as the format of the GTS animation and the ActionScript language and the Flare[52] library were used to develop it. The developed GTS animation has two parts: a rooted tree (Fig. 5.2a) and a radial tree (Fig. 5.2b). The radial tree is the main user interface and the rooted tree is set as a complementary view. Nodes in the rooted tree are parallel to those in the radial tree, and animations in both trees are synchronized.

Nodes in the developed animation represent GTS concepts and their relationships defined in the GTS ontology. Arrangements of nodes in the animation represent relationships between GTS concepts. From left to right in the rooted tree and from core to edge in the radial tree, the hierarchical layouts of nodes move from the higher to the lower levels of GTS concepts in the ontology. Meanwhile, the bottom-up and clockwise arrangements of nodes in, respectively, the rooted and radial trees follow the ordinal (i.e., earlier to later) sequence of geological time. In the radial tree, English names of GTS concepts were labeled in each node, whereas in the rooted tree these names were omitted due to the limitation of space here. The English names of nodes in the animation were retrieved from English labels encoded with "skos:prefLabel" in the GTS ontology. The filling colors of nodes were retrieved from RGB codes recorded with "gts:cgmwRgbColor". "Precambrian" is the only instance of the class "Supereonothem" in the GTS ontology, so in both rooted and radial trees in the animation there is only one node at this level (Figs. 5.2b, c). "Mississippian" and "Pennsylvanian" are the only two instances of the class "Subsystem", so there are only two named nodes in the radial tree and both have their unique filling colors (Figs. 5.2b, d). The equivalent nodes of "Mississippian" and "Pennsylvanian" in the rooted tree have no names here, but have their unique filling colors. In both trees, the

---

[52] http://flare.prefuse.org [Accessed February 05, 2011].

filling colors of un-named nodes at the "Subsystem" level are the same as their parent nodes to show that those un-named nodes represent no "Subsystem" concepts.

The developed GTS animation is not a static graph. Instead, several functions were incorporated into it, which can change the layouts of the animation dynamically according to the input queries. For example, a basic function is collapsing or expanding the two trees into different levels in the GTS hierarchy (Fig. 5.3), as triggered by a query of chronostratigraphic unit such as Eonothem, Erathem or System, etc.

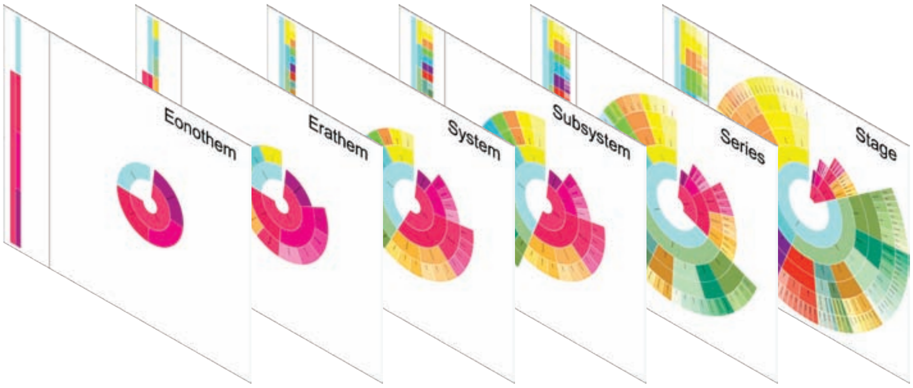
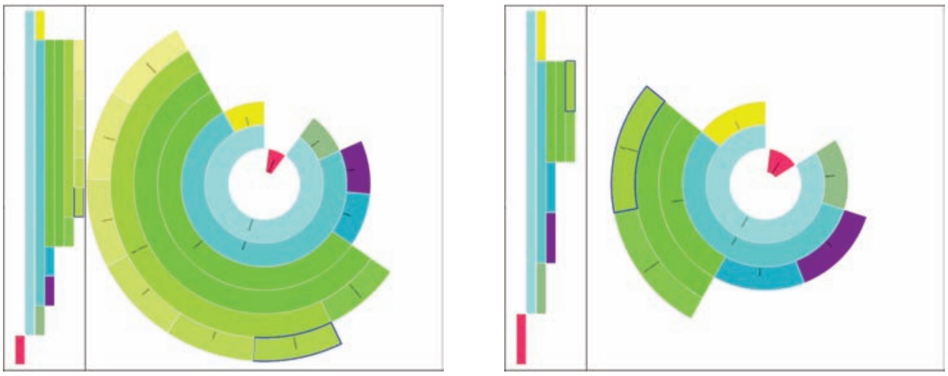


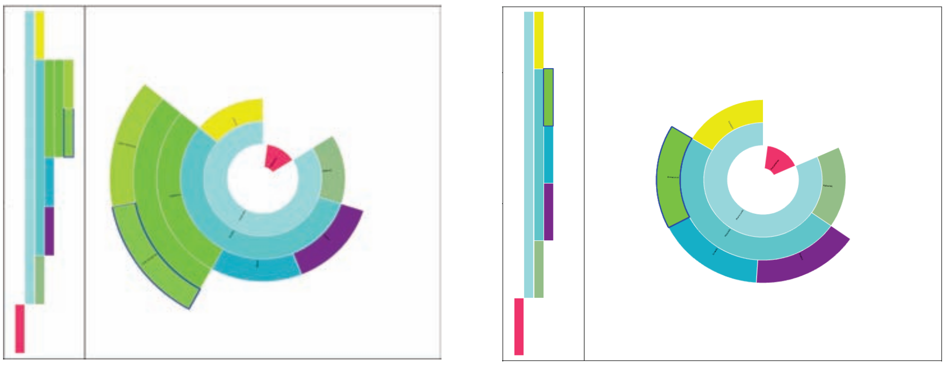
Fig. 5.3. Screenshots of the developed GTS animation showing that it collapses or expands to different levels of GTS concepts. Levels of GTS concepts in this figure follow a sequence of higher to lower chronostratigraphic units Eonothem, Erathem, System, Subsystem, Series and Stage.

Another function is collapsing into and highlighting a node in both trees synchronously (Fig. 5.4), as triggered by a query of GTS concept name. The located node is highlighted with a blue outline. The rules of the collapse (or animation) function are (1) showing the located node, its brother nodes, ancestor nodes and brother nodes of ancestor nodes, while (2) hiding all other nodes. A blank brother node of "Precambrian" was set at the level of "Supereonothem" (Figs. 5.1, 2c), as the father node of "Phanerozoic", to make perfect the implementation of the designed rules of the collapse function. Other capabilities of this GTS animation include zooming in and zooming out of view, showing a highlighted node in expanded trees, hiding or showing certain nodes in the trees, highlighting several nodes after semantic inferences based on input data, etc. Some of these developed functions are used to implement interactions between the GTS ontology, the GTS animation and online geological map services in the study presented in this chapter.



(a) Cenomanian.

(b) Upper Cretaceous.



(c) Lower Cretaceous.

(d) Cretaceous.

Fig. 5.4. Collapsing into and highlighting a node in the developed GTS animation. Layout of each of the four diagrams is triggered by an input GTS concept name: (a) Cenomanian, (b) Upper Cretaceous, (c) Lower Cretaceous, (d) Cretaceous. Nodes highlighted in both trees are equivalent in each diagram. For rules of collapse see text.

5.3 Interactions between GTS ontology, GTS animation and online geological map services

An essential feature of ontologies is their rich semantics and the ability of using semantic inferences (i.e., logical reasoning operations using definitions of concepts and relationships between concepts) to reach conclusions and produce new information (Katifori et al., 2007). Incorporating functions of semantic inferences into visualized ontologies has been increasingly studied in recent years, leading to novel features in vast applications. The OZONE (Suh and Bederson, 2002) visualizes query conditions and provides interactive, guided searching and browsing of ontological information. The OntoTrack (Liebig and Noppens, 2005) provides a graphical layout for handling ontologies, in which each editing step is synchronized with an external “reasoner” and then the reasoning feedback is shown instantly with animations and colorful marks. The CRAFT (Gruen et al., 2008) represents collective knowledge of cooperating analysts and handles reasoning tasks via interconnected graphical models built upon a shared evolving ontology. The Wivi (Lehmann et al., 2010) visualizes the structure of visited online articles and emphasizes relevant topics, acting as a guide for exploring larger information networks. Although significant progress has been made in incorporating semantic inferences into visualized ontologies, relevant studies are limited in the field of geological ontologies, and methods of using semantic inferences of visualized ontologies to interact with online geological map services are wanting.

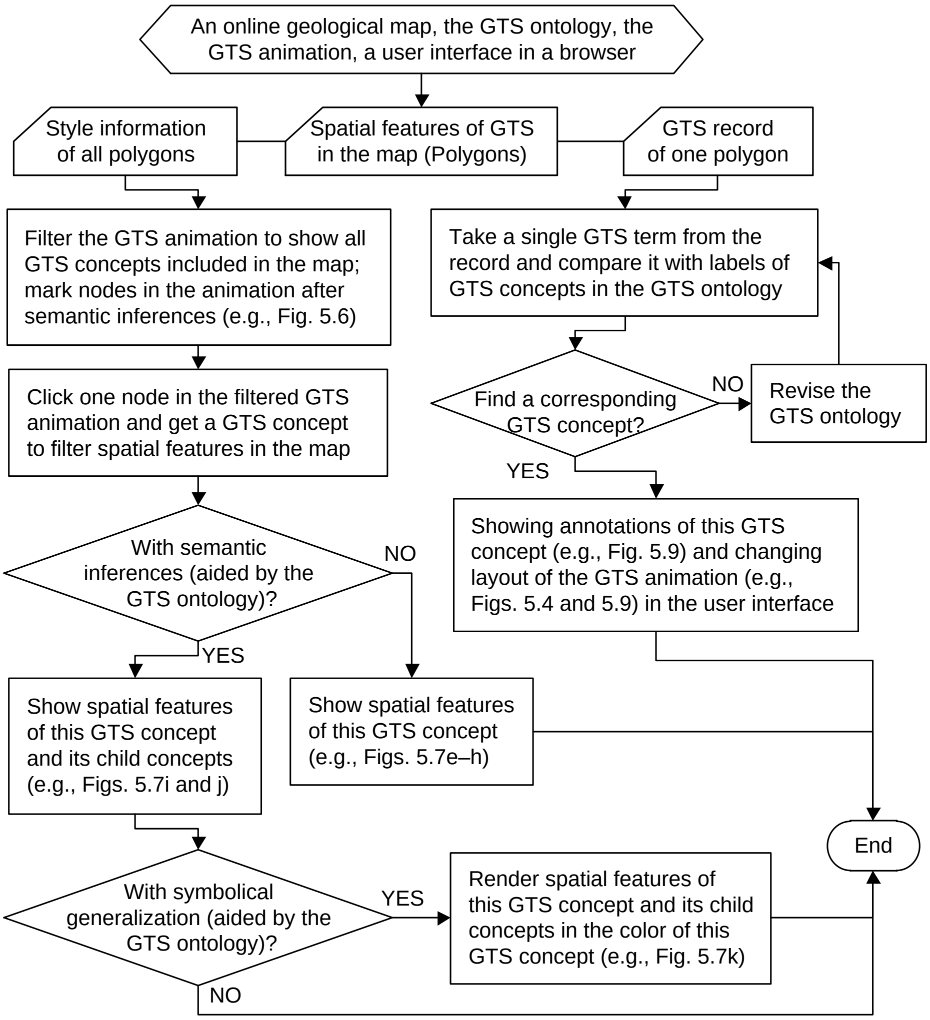


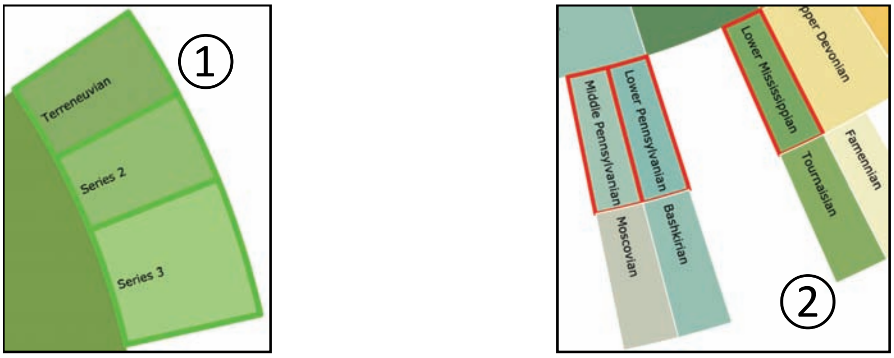
Fig. 5.5. Workflow for interactions between the developed GTS ontology, GTS animation and online geological maps.

A workflow was designed in the study presented in this chapter to conduct interactions between the GTS ontology, the GTS animation and Web Map Services (WMS) of geological maps (Fig. 5.5). One part of the interactions (right of Fig. 5.5) is explaining the GTS record retrieved from a polygon in a geological map. GTS terms are first recognized from the original GTS record. For every GTS term, the GTS ontology is searched to find a corresponding GTS concept. Then, annotations (e.g., time span in numbers, definition in text and links to GSSP and Wikipedia webpages) of this GTS concept are retrieved from the GTS ontology and shown in the user interface, and the layout of the GTS animation is changed instantly to highlight this GTS concept. In one of the previous works (Ma et al., 2011), methods were introduced to recognize GTS terms in a GTS record and explain meanings of these terms with aid of a SKOS-based GTS thesaurus. Similar methods were applied for recognizing GTS terms and arrange annotations for these terms in the workflow discussed above, but necessary updates were made, because the aid used in this study was changed from a thesaurus to an ontology and, many object and datatype properties used in the ontology are different from those used in the thesaurus.

The other part of the interactions (left of Fig. 5.5) is showing all GTS concepts included in a geological map with a filtered GTS animation and, in turn, using this filtered GTS animation as a panel to conduct symbolical generalizations of GTS contents in the map. In order to obtain all GTS concepts included in an online WMS geological map, a function was developed to (a) retrieve scripts of the GTS style information (i.e., map legend) of all polygons in a map; (b) parse the style information and recognize all GTS concepts; (c) find corresponding GTS concepts by searching the GTS ontology; and (d) send a list of found GTS concepts to the GTS animation. If an original GTS term is a synonym, it is identified by the GTS ontology by semantic inferences and then a note is attached in the list sent to the GTS animation. After receiving such a list of GTS concepts, a function in the GTS animation (a) hides nodes whose corresponding GTS concepts are not included in the received list (Figs. 5.6a, b); (b) marks nodes whose original GTS terms are synonyms as noted in the received list with green outlines (Fig. 5.6c); and (c) shows and marks nodes, whose corresponding GTS concepts are not included in the received list but whose child nodes are not hidden, with red outlines (Fig. 5.6d). Step (c) of the described function in the GTS animation also uses semantic inferences based on relationships between nodes in the animation. The filtered GTS animation (Figs. 5.6a, b) represents a legend of GTS features in a WMS geological map.

(a) Filtered rooted tree.

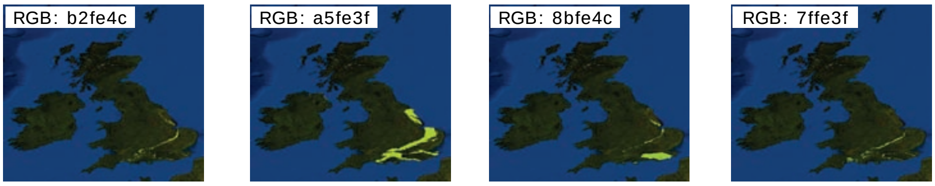
(b) Filtered radial tree.



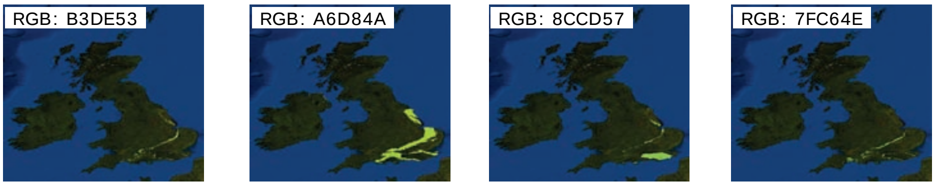
(c) Nodes with green outlines.

(d) Node with red outlines.

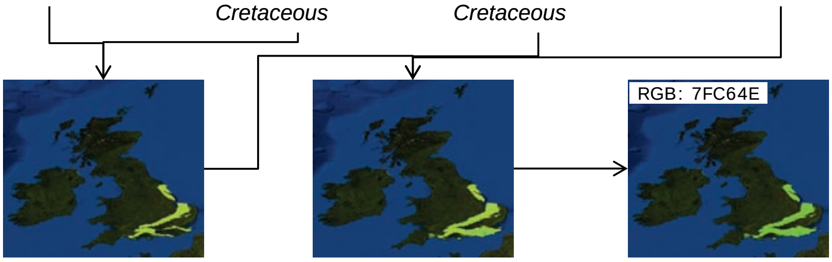
Fig. 5.6. Filtered GTS animation with marked results of semantic inferences after analyzing GTS data retrieved from a geological map. (c) and (d) are enlarged parts of (b), showing in detail the results of semantic references. In (c) a node with a green outline means the GTS concept shown in the node is included in the map contents, but the original records in the map use a synonym as the name of this GTS concept. In (d) a node with a red outline means the GTS concept shown in the node does not exist in the map, but is shown here because one or more of its child concepts are included in the map contents. For methods of filtering and semantic inferences see text.



(a) Cenomanian *(b) Upper Cretaceous* *(c) Lower Cretaceous* *(d) Cretaceous*



(e) Cenomanian *(f) Upper Cretaceous* *(g) Lower Cretaceous* *(h) Cretaceous*



(i) Upper Cretaceous after semantic inference *(j) Cretaceous after semantic inference* *(k) Cretaceous after generalization*

Fig. 5.7. Filtering out and generalizing GTS features of an online geological map aided by developed GTS ontology and GTS animation. (a)–(h) show results of direct filtering out (i.e., each filtered map shows features of only one GTS concept and polygons in the map are rendered in only one color). (a)–(d) use RGB codes from the style information retrieved from the online geological map, and (e)–(h) use RGB codes from the developed ontology. (i) shows a combination of polygons in (e) and (f) after semantic inference, because “Cenomanian” is a child concept of “Upper Cretaceous”. (j) shows a combination of polygons in (e)–(h), because “Lower Cretaceous” and “Upper Cretaceous” are both child concepts of “Cretaceous”. (k) shows a symbolical generalization of (j) (i.e., polygons in (j) are rendered in four colors, and polygons in (k) are rendered in only one color). RGB codes shown in this figure are in hexadecimal format. Original geological map (1:625,000 scale onshore bedrock age map of United Kingdom) reproduced with the permission of British Geological Survey © NERC. All Rights Reserved.

Besides the aforementioned functions, another function (left middle part of Fig. 5.5) was developed with the filtered GTS animation to filter out and generalize GTS features (i.e., polygons) in the original geological map. This function operates with the following steps: (a) a node in the filtered radial tree of the GTS animation is clicked and the user is provided two options by a question “With semantic inferences?”; (b) if the user chooses “NO”, the GTS animation sends only the name of this node (i.e., the label of the corresponding GTS concept) to a function outside the GTS animation; if the user chooses “YES”, the GTS animation sends a list of names of this node and all its visible child nodes (by using semantic inferences) to the function outside; (c) after receiving a GTS name or a name list, the function searches the “gts:cgmwRgbColor” properties in the GTS ontology and finds RGB codes for each GTS concept in the list and, then, the function creates a Styled Layer Descriptor (SLD) file following OGC® standards (OGC, 2007a; OGC/ISO, 2010) and sends it to the WMS geological map for filtering out and rendering GTS features (Figs. 5.7e–j); and (d) if there are more than one GTS concept received from the GTS animation, a symbolical generalization can be done by replacing the RGB codes of all GTS concepts with that of the top GTS concept in the SLD file and then sending it to the WMS geological map (Fig. 5.7k). In steps (c) and (d) of the aforementioned filtration and generalization function, an alternative operation is to parse the original style information obtained from the WMS geological map and to get RGB codes for each GTS concept, which can then be used to filter out and render GTS features (Figs. 5.7a–d) and do symbolical generalizations.

In the source code of a SLD file generated automatically in this study for filtering out and rendering GTS features of “Cretaceous” in a WMS geological map (Fig. 5.8), the element “<sld:Name>” (line 3) records the WMS map to which the SLD file is sent. The element “<sld:Rule>” (lines 6–19) records the conditions for filtering out GTS features recorded as “CRETACEOUS” (lines 7–12) and for rendering the GTS features filtered out (lines 13–18). The result generated by this SLD file is shown in Fig. 5.7h. The developed function can add more elements of “<sld:Rule>” in the SLD file for filtering out and rendering features of more than one GTS concept in the same WMS geological map. For example, in the SLD file for Fig. 5.7j (i.e., Cretaceous after semantic inference), there are four elements of “<sld:Rule>”, which set filtering and rendering conditions for GTS concepts “Cenomanian”, “Upper Cretaceous”, “Lower Cretaceous” and “Cretaceous”, respectively, each with a unique GTS concept name and a unique filling color. In the SLD file for Fig. 5.7k, there are also four elements of “<sld:Rule>”, with four different GTS concept names but only one filling color (i.e., RGB code of “Cretaceous”) to finish the symbolical generalization.

```
1    <sld:StyledLayerDescriptor version="1.1.0">
2      <sld:NamedLayer>
3        <sld:Name>GBR_BGS_625k_BA</sld:Name>
4        <sld:UserStyle>
5          <sld:FeatureTypeStyle>
6            <sld:Rule>
7              <ogc:Filter>
8                <ogc:PropertyIsEqualTo>
9                  <ogc:PropertyName>AGE_ONEGL</ogc:PropertyName>
10                  <ogc:Literal>CRETACEOUS</ogc:Literal>
11               </ogc:PropertyIsEqualTo>
12             </ogc:Filter>
13             <sld:PolygonSymbolizer>
14               <sld:Fill>
15                  <sld:CssParameter name="fill">#7FC64E</sld:CssParameter>
16                  <sld:CssParameter name="fill-opacity">1</sld:CssParameter>
17               </sld:Fill>
18             </sld:PolygonSymbolizer>
19           </sld:Rule>
20         </sld:FeatureTypeStyle>
21       </sld:UserStyle>
22     </sld:NamedLayer>
23   </sld:StyledLayerDescriptor>
```

Fig. 5.8. Source code of a SLD file sent to an online geological map for filtering out and rendering GTS features of “Cretaceous”. The result is shown as Fig. 5.7h. The RGB code in line 15 is retrieved from the developed GTS ontology.

Several programming languages and open-source libraries were adopted in the study presented in this chapter. RDF and SKOS were used to encode the GTS ontology, and ActionScript and the Flare library were used to develop the GTS animation in Flash format. JavaScript and the OpenLayers[53] library were used to access online geological maps, SLD was used to filter out and generalize GTS features of these maps, and HTML (HyperText Markup Language) was used to develop the user interface (as a part of the pilot system). Techniques of communication between Flash, JavaScript and HTML (Elst et al., 2006) were applied to transfer data between the GTS ontology, the GTS animation and online geological maps in the designed workflow. About 320 man-hours were spent for developing these works and setting up a pilot system, and extra time was also spent to do tests and evaluations.

5.4 Pilot system, results and evaluation

As mentioned above, a primary objective of the study presented in this chapter is to facilitate GTS information retrieval and knowledge discovery by annotating, visualizing, filtering and generalizing GTS information of online

---

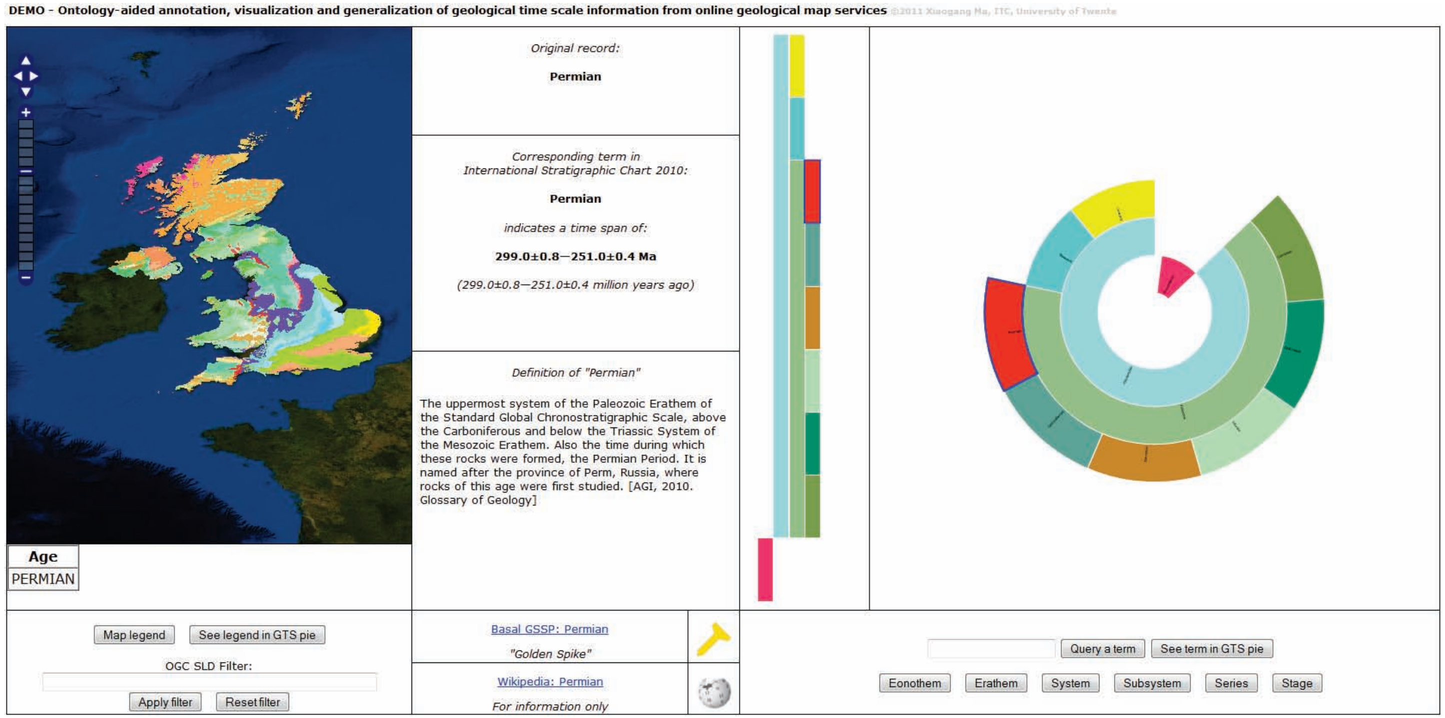
[53] http://openlayers.org [Accessed February 11, 2011].

geological maps. Although several examples were already used to demonstrate the functions of the developed GTS ontology, GTS animation and their interactions with WMS geological maps, the usability and usefulness of these works still need further tests and evaluation. A pilot system was set up to do so.

In the pilot system, a WMS server[54] provided by the British Geological Survey (BGS) was linked to, from where the 1:625,000 scale onshore bedrock age map of United Kingdom was retrieved and then shown in a map window (left part of Fig. 5.9) in a user interface. The system gets the GTS record of a polygon after a click on this polygon in the map window. The retrieved GTS record (e.g., "PERMIAN") is shown below the map window, while the functions incorporated in the system parse the GTS ontology and retrieve annotations (middle part of Fig. 5.9) for each GTS concept (e.g., "Permian") recognized from the GTS record. These functions also generate links to the GSSP information and the Wikipedia page of a GTS concept. The former link provides formal and ratified information and the latter can provide, more multilingual information. Meanwhile, the rooted tree and radial tree in the GTS animation (right part of Fig. 5.9) collapse into the corresponding node (e.g., "Permian") of the GTS concept and highlight it with a blue outline. Besides the example shown in Fig. 5.9, the examples in Fig. 5.4 were also generated by the pilot system.

---

[54] http://ogc.bgs.ac.uk/cgi-bin/BGS_Bedrock_and_Superficial_Geology/wms [Accessed August 10, 2010].



DEMO - Ontology-aided annotation, visualization and generalization of geological time scale information from online geological map services ©2011 Xiaogang Ma, ITC, University of Twente

Age: PERMIAN

Original record:

Permian

Corresponding term in International Stratigraphic Chart 2010:

Permian

indicates a time span of:

299.0±0.8—251.0±0.4 Ma

(299.0±0.8—251.0±0.4 million years ago)

Definition of "Permian"

The uppermost system of the Paleozoic Erathem of the Standard Global Chronostratigraphic Scale, above the Carboniferous and below the Triassic System of the Mesozoic Erathem. Also the time during which these rocks were formed, the Permian Period. It is named after the province of Perm, Russia, where rocks of this age were first studied. [AGI, 2010. Glossary of Geology]

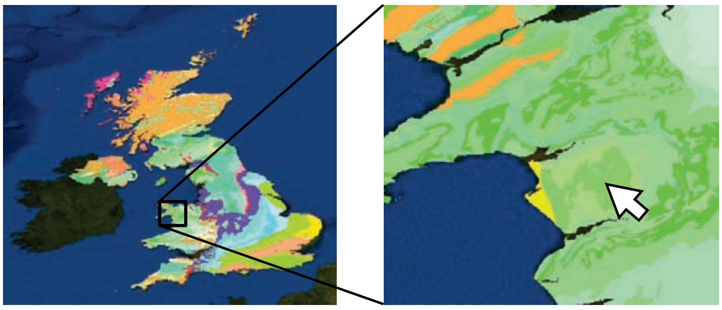
Basal GSSP: Permian

"Golden Spike"

Wikipedia: Permian

For information only

Fig. 5.9. User interface of developed pilot system. The user interface includes three interactive parts: an online geological map in the left, annotations of GTS concepts in the middle, and a GTS animation in the right. Original geological map reproduced with the permission of British Geological Survey © NERC. All Rights Reserved.

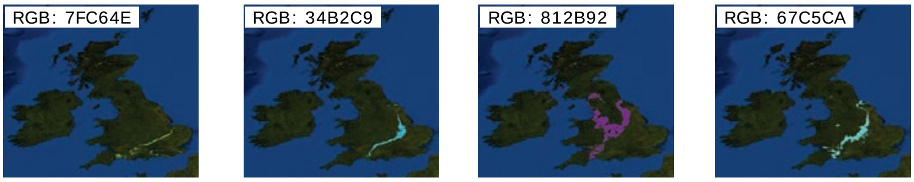


(a) An example of GTS features recorded as "LOWER CAMBRIAN"

(b) GTS animation collapsed to a union of "Terreneuvian" and "Series 2" (see bottom right part of the rooted tree and top right part of the radial tree)

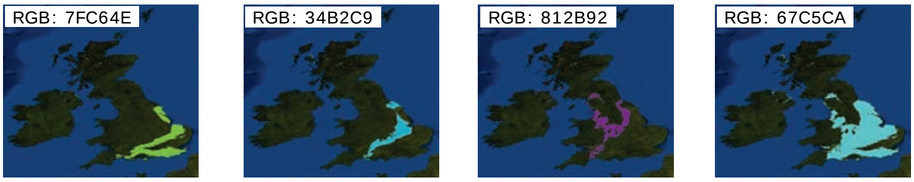
Fig. 5.10. Nodes highlighted with green outlines due to a synonym used in an original GTS record. (a) shows an example of GTS features of "LOWER CAMBRIAN" in the 1:625,000 scale onshore bedrock age map of United Kingdom. "Lower Cambrian" is not a standard term in the International Stratigraphic Chart, but with developed GTS ontology, "Lower Cambrian" is recognized as a union of "Terreneuvian" and "Series 2". Then in (b) the GTS animation collapse into nodes of these two GTS concepts and highlight them with green outlines, indicating the GTS term used in the original record is a synonym. Original geological map reproduced with the permission of British Geological Survey © NERC. All Rights Reserved.

A GTS record may use a synonym (i.e., terms not included in the International Stratigraphic Chart (see footnote 48)) as the name of a GTS concept. If this synonym is recorded as a "skos:altLabel" in the GTS ontology, then by parsing the GTS ontology the system can find its corresponding “skos:prefLabel” as a standard name. The GTS animation will then collapse into the corresponding node and will highlight them with green outlines (Fig. 5.10).



(a) Cretaceous (b) Jurassic (c) Triassic (d) Mesozoic

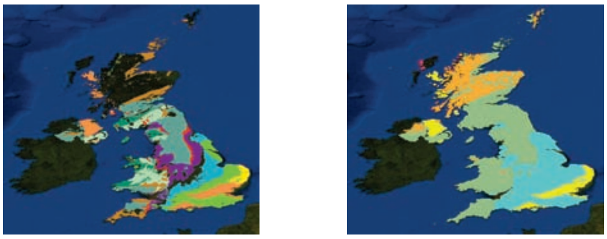
(e) Cretaceous after semantic inference (f) Jurassic after semantic inference (g) Triassic after semantic inference (h) Mesozoic after semantic inference



(i) Cretaceous after generalization (j) Jurassic after generalization (k) Triassic after generalization (l) Mesozoic after generalization

Fig. 5.11. Filtering results and symbolical generalizations of GTS features in the 1:625,000 scale onshore bedrock age map of United Kingdom with RGB codes from developed GTS ontology. (h) is equivalent to a combination of (d), (e), (f) and (g). (l) is a symbolical generalization of (h). Original geological map reproduced with the permission of British Geological Survey © NERC. All Rights Reserved.

A case study of filtering out and generalizing GTS features in the 1:625,000 scale onshore bedrock age map of United Kingdom was conducted. Figs. 5.11a–d and 11e–h are results of filtered out GTS features by clicking nodes in the filtered GTS animation (Fig. 5.6b), which is generated by clicking the button “See legend in GTS pie” in the bottom left part of the user interface (Fig. 5.9). Figs 5.11i–l are results of symbolical generalizations using RGB codes from the GTS ontology.



(a) Generalization at System level

(b) Generalization at Erathem level

(c) Generalization at Eonothem level

(d) Generalization of Precambrian and Phanerozoic

Fig. 5.12. Symbolical generalizations of different levels of GTS features in the 1:625,000 scale onshore bedrock age map of United Kingdom with RGB codes from developed GTS ontology. The major difference between (c) and (d) is the area of Na h-Eileanan Siar (or Western Isles) at the top left part of the two maps. Original geological map reproduced with the permission of British Geological Survey © NERC. All Rights Reserved.

In the same map, spatial features at different GTS levels (i.e., Supereonothem, Eonothem, Erathem, System, Series and Stage from higher to lower) were also generalized. The level of detail of GTS records in the map influences the results of symbolical generalizations. For example, the map was generalized at the System level and it was found that some areas in the generalized map were blank (Fig. 5.12a). That is because the original GTS records of those blank areas are GTS concepts whose levels are higher than System (i.e., Erathem, Eonothem and Supereonothem) and, therefore, they cannot be generalized to a lower level (i.e., System). Similar cases became apparent when the same map was generalized at the Erathem level (Fig. 5.12b) and the Eonothem level (Fig. 5.12c). Only in the generalization of "Precambrian" (a Supereonothem concept) and "Phanerozoic" (an Eonothem concept) were all polygons in the original map filtered out and re-rendered (Fig. 5.12d).

From the perspective of developers, the GTS ontology and GTS animation were integrated with a WMS server of BGS in the pilot system with light adaptations to the developed interactive functions. From the perspective of users, only a short GTS record could be obtained by a click operation in the original WMS geological map, but in the pilot system, supports were provided to help users to understand GTS information and discover GTS knowledge in the map. To evaluate the usefulness of key functions developed in the pilot system, a user-survey was made wherein 19 PhD students participated. The particular objective of this survey was to determine if users of the system, especially those who are unfamiliar with geology, are able to comprehend usability of geological data (i.e., as an essential part of their interoperability) (Bishr, 1998; Harvey et al., 1999; Broome, 2005; Bond et al., 2007; Gahegan et al., 2009). The 19 participants in the user-survey are located at five departments (i.e., earth observation science, earth systems analysis, geo-information processing, natural resources, and water resources) in the Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente. They fall into two groups: (1) nine are familiar with GTS and (2) 10 are unfamiliar with GTS. It was expected that these two groups have different opinions about the usefulness of the system in terms of its five functions (Appendix 5-I), with the “familiar” group’s opinions as references for usefulness.

On average, the “familiar” group scored all the individual functions, except the Annotation function, lower than the “unfamiliar” group (Tables 5.1, 5.2). The average scores by the “familiar” were mostly between “Useful” and “Very Useful”, except that their average scores for the Collapse (or Animation) and Legend functions tend only toward the “Useful” category (Table 5.2, Appendix 5-I). In contrast, the average scores by the “unfamiliar” group tend more toward the “Very useful” category. An explanation for this is that geologists, compared to other earth scientists, historically tend to be reluctant in using computer technology (cf. Hubaux, 1973; Rock, 1991; Huff, 1998; Clegg et al., 2006). Results of two-sample t-tests show that the scores given by the two groups are not significantly different for all functions except Collapse (Appendix 5-II, Table 5.2). In particular, both groups similarly found that Annotation, Visualization and Filtering are “Useful” to “Very Useful”, and both groups similarly found that Legend is more “Useful” than “Very useful”. However, the “familiar” group found that Collapse (or Animation) is just “Useful” where the “unfamiliar” group found this function to be “Very useful”. The results show, therefore, that the developed GTS ontology and the associated functions to facilitate GTS information retrieval and knowledge discovery are useful not only for those who are familiar but also to those who unfamiliar with geology. Nevertheless, the results of the survey indicate that the Collapse and Legend functions need further re-thinking to improve their usefulness for geologists.

Table 5.1 Scores given by participants on usefulness of functions in the GTS (geological time scale) pilot system. For meanings of column heads and scores see Appendix 5-I.

Participant	**Annotation**	**Visualization**	**Collapse**	**Legend**	**Filtering**
Familiar with GTS					
1	3	3	3	3	3
2	3	3	3	2	2
3	2	2	2	2	3
4	3	2	3	3	3
5	3	3	3	2	3
6	2	3	2	3	3
7	3	1	1	0	2
8	3	3	2	2	2
9	2	3	1	3	3
Unfamiliar with GTS					
10	3	3	3	3	3
11	3	3	3	3	3
12	1	2	3	3	2
13	3	3	3	2	3
14	3	3	3	3	3
15	3	2	3	1	3
16	3	3	3	2	3
17	3	3	3	3	3
18	2	3	3	3	3
19	2	2	3	3	3

Table 5.2 Results of two-sample t-tests on scores given by the two groups of participants (Table 5.1). For meanings of variables and details of the t-tests see Appendix 5-II.

Function	$\bar{x}_f$	s_f	$\bar{x}_u$	s_u	$\|t\|$	df	t^*	P-**value**
Annotation	2.667	0.500	2.600	0.699	0.241	16.236	2.120	0.187
Visualization	2.556	0.726	2.700	0.483	0.504	13.704	2.160	0.378
Collapse	2.222	0.833	3.000	0.001	2.800	8.000	2.306	0.977
Legend	2.222	0.972	2.600	0.699	0.963	14.410	2.145	0.648
Filtering	2.667	0.500	2.900	0.316	1.200	13.268	2.160	0.749

5.5 Discussion

The works of developing the GTS ontology, GTS animation and their interactive functions with online geological maps show the capabilities of ontologies and visualization techniques for complementing online geological map services. Results of the pilot system and the user-survey prove the usability and usefulness of the developed works for promoting geological data interoperability and geological knowledge discovery in the Semantic Web, though some functions can be further updated.

Compared to the previous work of a SKOS-based GTS thesaurus (Ma et al., 2011), the GTS ontology in the study presented in this chapter provides more precise definitions on meanings of GTS concepts and relationships between GTS concepts. RGB codes, GSSP information and definitions in glossaries were collected to enrich annotations in the GTS ontology. For the relationships between GTS concepts, the GTS ontology developed in the SWEET project (Raskin and Pan, 2005) was referred, such that "Supereonothem", "Eonothem", "Erathem", "System", "Subsystem", "Series" and "Stage" were defined as classes in the GTS ontology. In the previous work of the GTS thesaurus, all GTS concepts are defined as instances of "skos:Concept". For example, both "Precambrian" and "Phanerozoic" are instances of "skos:Concept". In the GTS ontology developed in this study, GTS concepts are defined as instances of the pre-defined classes, for example, "Precambrian" is an instance of "gts:Supereonothem" while "Phanerozoic" is an instance of "gts:Eonothem". The meanings of object properties "gts:subsetOf" and "gts:supersetOf" used in the GTS ontology are also clearer than those of the "skos:broader" and "skos:narrower" in the GTS thesaurus.

Annotations in ontologies and vocabularies have been increasingly studied in recent years. In the field of genetic ontologies, it has been extensively discussed that using commonly accepted annotations in an ontology can enhance the interoperability of datasets underpinned by this ontology (Camon et al., 2003; Dimmer et al., 2008; Hong et al., 2008). Recently, it was also discussed (Rhee et al., 2008) that incorrect annotations in gene ontology may lead to incorrect results and conclusions of studies that use this ontology and, thus, it is crucial to collect annotations from reliable sources and provide metadata of them. The approaches for arranging annotations in the GTS ontology described in this chapter are similar to those in studies of genetic ontologies. The International Stratigraphic Chart was adopted as the foundation of the GTS ontology, and definitions of GTS concepts were collected from the Glossary of Geology, RGB codes were collected from CGMW and GSSP information were collected from Subcommission for Stratigraphic Information of the International Commission on Stratigraphy, so that the developed ontology can provide reliable explanations of GTS concepts to users. Similar opinions on annotations were also expressed in several recent studies in the field of geoscience ontologies (Visser et al., 2002; Klien, 2007; Lumb et al., 2009). In the field of GTS ontologies/vocabularies, the 1G-E project presented featured services of using annotations to facilitate data interoperability and information retrieval. Such services are underpinned by a geological vocabulary in which several international and regional standards are adopted and/or adapted (Asch et al., 2010; Laxton et al., 2010). The work of combining ontology-based annotations with WMS geological maps described in this study is similar to that of the 1G-E project. The difference is that the GTS ontology in this study provides more information on GTS concepts than what the 1G-E geological time vocabulary does.

Semantic inference or logical reasoning is one of the key features of ontologies. Incorporating functions of reasoning into visualized ontologies is important to support interactive learning and knowledge discovery (Min et al., 2009). In Section 5.3 some related studies are already discussed, such OZONE (Suh and Bederson, 2002), OntoTrack (Liebig and Noppens, 2005), CRAFT (Gruen et al., 2008) and Wivi (Lehmann et al., 2010). In recent years, such ontology-based visualization and reasoning techniques have been put into practice. Some typical examples can be found in the field of medical research. Zillner et al. (2008) incorporated external semantics into patient data visualization and realized semantic facet browsing and semantic tree-map visualization using class-based reasoning. Gonçalves et al. (2009) developed an application of ontology for representation, reasoning and visualization of heart electrophysiology on the Web. Dupplaw et al. (2009) developed an ontology-driven framework with multimedia processing, annotation and reasoning to support multidisciplinary meetings that take place during breast cancer screening for diagnosing the patient. In a recent geospatial study, Willems et al. (2010) developed a system for analyzing the behavior of moving objects, which can abstract and simulate trajectory sensor data in an ontology, fuse multiple heterogeneous data sources into a knowledge base, and then conduct reasoning visual analysis of the combined data sources. For coupling reasoning and visualization with ontologies, the approach applied in the works of the GTS ontology and the GTS animation in this study is similar to the mentioned studies. However, the background of this study is geology and the GTS ontology and GTS animation developed are used to complement online geological map services.

Using ontology-driven approaches in map generalization has been extensively discussed recently. Understanding the meanings and inter-relationships of concepts represented by map features is essential for users to explore information and knowledge contained in maps (Kraak, 2008; Neun et al., 2008). Ontologies help users to understand map features and can be used to generalize maps (Kulik et al., 2005). Vast algorithms have already been studied in approaches to ontology-driven map generalization, such as genetic algorithms (Ware et al., 2003), supervised Bayesian inference (Lüscher et al., 2009), and heuristic methods for generalization of large datasets (Haunert and Wolff, 2010). Compared to these sophisticated algorithms, the method of ontology-based map generalization in the study presented in this chapter is simpler because the method used directly the hierarchical structure among GTS concepts defined in the ontology. Another difference is that, in the developed generalization functions, the outlines of polygons in the original WMS map are not changed, but the filling colors of sub-class concepts are changed into colors of super-class concepts to realize the symbolical generalization. The web portal[55] of the 1G-E project (Laxton et al., 2010) provides a vocabulary-supported geological map generalization service by which users can create a custom map using self-assigned RGB codes of GTS. The work of ontology-based map generalization in this study is similar to that of the 1G-E project, but the difference is that in this study a GTS animation is provided as an operation panel to simplify the operations of generalization and the RGB codes of GTS used in this study are controlled by the GTS ontology.

Two lessons are learned from the study presented in this chapter. The first is that detailed metadata from data sources can support efficient use and re-use of data in the context of the Semantic Web. Metadata is a much discussed topic both in general computer science (e.g., Gray et al., 2005; Schofield et al., 2009) and in geo-information science (e.g., Green and Bossomaier, 2002; Ma et al., 2007; Tilmes et al., 2010). Here, it is noteworthy the convenience that metadata can bring to ontology-based online geological data applications. If a geological data source provides detailed metadata about subjects (e.g., GTS), languages (e.g., English), and standards used (e.g., the International Stratigraphic Chart), etc., by maps published on its data server, users can apply corresponding ontologies (e.g., a GTS ontology) in applications after they retrieve data from the server. For WMS maps, a request “GetCapabilities” sent to a WMS server can return some of this metadata, but it depends on what are registered by the data providers. Another way is providing Catalog Service for the Web (CSW) (OGC, 2007b) on a data server. There are already studies of CSW in geosciences (Chen et al., 2010; Gebhardt et al., 2010), and their functions in online geological map services can be further studied. The second lesson learned is that standardization of data in geological map services influences the results of ontology-based applications. In this study, satisfactory results were obtained in ontology-based GTS concept recognition, visualization and symbolical generalization of GTS contents in a geological map on the BGS

---

[55] http://onegeology-europe.brgm.fr/geoportal [Accessed February 27, 2011].

WMS server. This is mainly due to (1) the compatibility of the GTS ontology (i.e., many synonyms of GTS terms are collected in it) and (2) the high standardization of GTS contents in this geological map (i.e., most GTS records in this map are standard GTS terms from the International Stratigraphic Chart). If the geological data on a server does not address standardization strictly and result in heterogeneous datasets, either the ontology should be updated in order to recognize concepts in these datasets or, if the results are still not satisfactory, strategies and methods of applying ontology-based tools with online geological map services may be redesigned.

From the study presented in this chapter, directions for further studies can also be recommended. The first is the multilingual annotation of GTS concepts. In the GTS ontology, multilingual labels of GTS concepts were already collected, which enable the ontology to recognize GTS terms in their multilingual formats, but the annotations of concepts in current GTS ontology are in English only. Enhancing multilingual annotations in the GTS ontology can potentially broaden the scope of applications of the ontology. The second direction is collecting more conceptual mapping cases in the GTS ontology. By accessing the 1:625,000 scale onshore bedrock age map of United Kingdom two mapping cases were collected: “Lower Cambrian = Terreneuvian + Series 2” and “Middle Cambrian = Series 3”. If more other geological maps are accessed, more such mapping cases can be collected and they will be useful for understanding and mediating heterogeneous datasets. The third direction is studying methods for filtering out a map feature with several GTS concepts in its attribute record. In the works described in this chapter, by clicking a node in the filtered GTS animation to filter a map (Figs. 5.6b, 7), only those map features with one GTS concept in its attribute record can be retrieved by the current program. Each GTS feature in the 1:625,000 scale onshore bedrock age map of United Kingdom has only one GTS concept, so satisfactory results of GTS feature filtering and generalization were obtained in the pilot system. In order to make the functions of filtering and generalization apply also to map features with several GTS concepts, new methods should be developed. Finally, in this study ontology-based tools were only applied to WMS geological maps, using ontologies and visualization techniques with WFS (Web Feature Service) and KML (Keyhole Markup Language) geological maps can also be considered in further studies (cf. De Paor and Whitmeyer, 2011).

5.6 Conclusions

Geological data infrastructures have been widely used in publication and sharing of geological data, whereas tools and services for information retrieval and knowledge discovery are underdeveloped compared to the massive geological data available online. In the study presented in this chapter, a RDF-based ontology of geological time scale, an animation based on this ontology, and interactive functions among the ontology, the animation and online geological map services were developed. A pilot system was built with the developed ontology, animation and interactive functions, and positive results were obtained in a user-survey on usefulness of the developed works. This study shows that annotations in an ontology and ontology-based visualizations are useful in helping people to understand concepts defined in the ontology. In addition, incorporating ontology-based annotations, visualizations and interactive functions with online geological map services is helpful for both geologists and non-geologists to understand information in a map and to conduct further operations of knowledge discovery. The case studies in Chapters 2–5 covers geological data interoperability issues at local, regional and global levels, but they do not take into account the evolution of geological data and ontologies. In a long-term perspective, local geological ontologies and geological data will evolve in their respective contexts and, thus, new challenges will arise. Chapter 6 will discuss these issues and propose solutions.

Appendix 5-I

1) Meanings of abbreviated function names

Annotation	Showing annotations of GTS concepts using the GTS ontology
Visualization	Showing the conceptual structure of GTS with a rooted tree and a radial tree (both expanded)
Collapse (or Animation)	Collapsing into and highlighting a chosen GTS concept in the rooted tree and the radial tree
Legend	Showing legend of GTS contents in a map with the rooted tree and the radial tree
Filtering	Filtering out of certain GTS features in a map with the rooted tree and the radial tree

2) Meanings of scores on usefulness of a function

3	Very useful
2	Useful
1	Somewhat useful
0	Not useful at all

Appendix 5-II

1) Reasons for using the two-sample t-test

The "familiar" and "unfamiliar" groups are independent, and their sizes are small and different ($n_f = 9$, $n_u = 10$). Variances of scores by either group are unknown and are assumed unequal.

2) Hypotheses of the two-sample t-test

$H_0 : \mu_f \neq \mu_u$	Null hypothesis: The "familiar" and "unfamiliar" groups have different opinions about the usefulness of a function.
$H_a : \mu_f = \mu_u$	Alternative hypothesis: The "familiar" and "unfamiliar" groups have similar opinions about the usefulness of a function

3) Meanings of variables used in the two-sample t-test

n_f	Size of "familiar" group
n_u	Size of "unfamiliar" group
$\bar{x}_f$	Mean of "familiar" group's scores on the usefulness of a function
$\bar{x}_u$	Mean of "unfamiliar" group's scores on the usefulness of a function
s_f	Standard deviation of "familiar" group's scores
s_u	Standard deviation of "unfamiliar" group's scores
t	Two-sample t-value
df	Estimated degrees of freedom using the Welch–Satterthwaite equation
P-value	Probability (two-sided) that the null hypothesis is true
$t*$	Critical t-value at the significance level of 0.05

4) Equations for calculating t and df

$$t = \frac{\bar{x}_f - \bar{x}_u}{\sqrt{\frac{s_f^2}{n_f} + \frac{s_u^2}{n_u}}} \; ; \; df = \frac{\left(\frac{s_f^2}{n_f} + \frac{s_u^2}{n_u}\right)^2}{\frac{1}{n_f - 1}\left(\frac{s_f^2}{n_f}\right)^2 + \frac{1}{n_u - 1}\left(\frac{s_u^2}{n_u}\right)^2}$$

Chapter 6

Pragmatic interoperability approach for distributed geological data

This chapter is based on: Ma, X., Carranza, E.J.M., Wang, X., Wu, C., van der Meer, F.D., Pragmatic interoperability approach for distributed geodata. Submitted.

6.1 Introduction

Interoperability of geological data (geodata) is essential for sharing geodata, retrieving geoinformation and discovering geo-knowledge within a cyberinfrastructure (Harvey et al., 1999; Nambiar et al., 2006; Zhao et al., 2009). A general requirement for the interoperability of geodata is that data provided by a geodata source can be accessed, decoded, understood and appropriately used by external users (Brodaric and Gahegan, 2006; Loudon and Laxton, 2007; Gahegan et al., 2009).

Considering the machine-readable aspects of geodata, various researchers (e.g., Bishr, 1998; Ouksel and Sheth, 1999; Sheth, 1999) have discussed that there are generally four levels of geodata interoperability, namely system, syntax, schematics, and semantics. In practice, these four levels of geodata interoperability are related, respectively, to the platform, encoding, structure, and meaning of geodata (Ludäscher et al., 2003; Brodaric and Gahegan, 2006; Laxton et al., 2010). Systemic, syntactic, and schematic interoperability issues, such as cross-database access and correct decoding of geodata, have benefited extensively from developments in general information technologies. Examples of these technologies are as protocols for web-based data transfer, programs for inter-conversion of file formats, interfaces for mapping between database schemas, etc. However, semantic interoperability issues, such as accurate understanding and appropriate using of geodata, are less-developed because they often involve one or more subject domains in geosciences and, therefore, require specification, conversation, and collaboration on domain-specific knowledge.

In the past decades, ontology-driven approaches have been studied significantly to address issues of semantic interoperability of geodata (Ludäscher et al., 2003; Agarwal, 2005; Lutz et al., 2009). Ontologies, which are shared conceptualizations of domain knowledge (Gruber, 1993; Gruber, 1995; Guarino, 1997b), are of different types (e.g., top-level ontologies, domain ontologies, application ontologies, etc.) (Guarino, 1997a) and different forms (e.g., glossaries, thesauri, conceptual schemas, logical theories, etc.) (McGuinness, 2003; Uschold and Gruninger, 2004; Ma et al., 2010). Many geodata producers maintain their local application ontologies in order to promote standardization and consistency in local databases. However, semantic heterogeneities often arise between local application ontologies, even if they are derived from equivalent subject domains in geosciences. In contrast, common ontologies have been studied in various geoscience subject domains (e.g., NADM Steering Committee, 2004; Raskin and Pan, 2005; McGuinness et al., 2006; Tripathi and Babaie, 2008; Ma et al., 2011) as an approach for reconciling heterogeneous local ontologies and databases. Ontology mapping (e.g., mapping between local ontologies, mapping local ontologies to a common ontology, etc.) has been studied and found to be effective for achieving semantic interoperability between local ontologies and, thus, between distributed geodata sources (e.g., Ludäscher et al., 2003; Verheyden et al., 2005; Lutz et al., 2009).

Nevertheless, there are still challenges in the research field of interoperability of geodata. Local ontologies and databases are not static and absolute, but are evolving in the real world (cf. Oreskes et al., 1994; Ding and Foo, 2002; Noy and Klein, 2004; Haase et al., 2005). For example, new conflicts will arise between two distributed geodata sources even if semantic interoperability has been achieved between them before. This is because local ontologies and databases are context-dependent as they are related to and affected by pragmatic elements of local contexts (Guarino, 1997a; van Heijst et al., 1997; Bouquet et al., 2004). Context is the situation or settings in which something happens. The contexts of distributed geodata sources (i.e., geological data contexts/geodata contexts) consist of local ontologies and databases as well as other pragmatic elements (e.g., people, intentions, methods, procedure of working, etc.). Therefore, in order to achieve the interoperability between distributed geodata contexts for consensus on understanding, use and potential result of shared geodata (i.e., pragmatic interoperability), representations of geodata contexts are required (cf. Tolk et al., 2006; Manso et al., 2009).

In this chapter, a model of geodata context is demonstrated, and a procedure of semantic negotiations for achieving pragmatic interoperability of distributed geodata is proposed. To reach these goals, information agents, objective facts, and subjective dimensions are proposed as elements of a conceptual model of geodata contexts. This model is then used to design a semantic negotiation procedure for achieving pragmatic interoperability of distributed geodata. The discussed conceptual model and semantic negotiation procedure were applied and tested in the National Mineral Resources Assessment (NMRA) project of China to achieve pragmatic interoperability among various geodata sources and researchers involved in the project.

6.2 Motivation

In 2006, the China Geological Survey (CGS) initiated the 5-year NMRA project to inventorize the spatial distributions and potential resources of 13 types of mineral deposits in China. The NMRA project consists of 47 sub-projects undertaken by CGS agencies at provincial and national levels (i.e., 30 provincial sub-projects, 11 national sub-projects, and six sub-projects of coordinating groups between the two levels). The essential geodata in the NMRA project are digital maps, covering the subjects of geology, metallogenesis, geophysics, geochemistry, remote sensing, as well as records of detailed exploration (e.g., borehole logs), etc. Most provincial sub-projects (i.e., except a few provinces with under-developed geological works) will complete all these maps and then provide them to national sub-projects for integration and syntheses.

Besides the short-term goal of assessing the potential resources of various types of mineral deposits in China, one of the long-term goals of the NMRA project is to investigate a mechanism for responding to challenges concerning interoperability of geodata among CGS agencies, who conducted not only the sub-projects of the NMRA project but also many other CGS projects in the past. Many of these challenges involve at least one the three information agents (i.e., human, machine, nature) of a geodata context (Brodaric, 2007). Provincial and national sub-projects are regarded as individual geodata contexts (Fig. 6.1). The nature information agent stands for an earth domain (e.g., geology, etc.). The human information agent stands for the staff of a project and their tacit knowledge. Staff members observe and study a certain earth domain and express their findings and knowledge through the machine information agent, which consists of a local ontology and a local database. The local ontology is used to promote standardization and consistency of geodata in the local database. Geodata sharing is operated between machine information agents of the two geodata contexts. For example, Researchers in a national sub-project use the shared geodata from a provincial sub-project, without or with less observation of the real earth, to generate their understandings of the same part of the earth and conduct further studies.

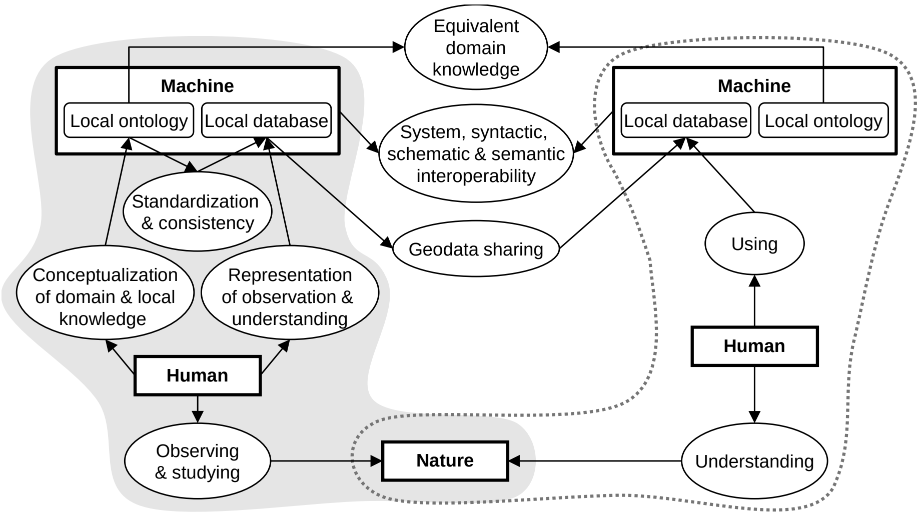


Fig. 6.1. Geodata sharing as an activity between two geodata contexts. The gray area at the left side is the geodata context of a provincial sub-project. The area surrounded by a dotted line at the right side is the geodata context of a national sub-project. The geodata context of the national sub-project is simplified, showing only how the shared geodata are used.

Because the two geodata contexts described in Fig. 6.1 share equivalent domain knowledge in geosciences, issues of systemic, syntactic, schematic, and semantic interoperability are addressed among the two machine information agents in order to promote appropriate and effective use of the shared geodata. In the geodata context of a provincial sub-project, the local ontology and database in the machine information agent represent objective facts about the nature information agent and subjective understandings of the human information agent. Although the nature information agent is stable, the tacit knowledge of the human information agent is not static and, thus, may lead to changes in the ontology and database of a provincial sub-project as new studies are continuously conducted. For example, several new methods of regional mineral assessments (e.g., Wang, 1990; Zhai, 2003; Ye, 2004; Zhao, 2006; Chen, 2007; Cheng, 2007) have been applied in China in recent years. The meanings of some geological concepts also evolved. An example is the basal boundary of Quaternary, which changed several times within the International Commission on Stratigraphy in recent years, and in turn led to changes in the International Stratigraphic Chart and geological maps (cf. Mascarelli, 2009). Moreover, similar changes may also happen in human and machine agents of the geodata context of a national sub-project (cf. Kuhn, 2009). Such changes in both geodata contexts may lead to the heterogeneities of understandings, uses and results of shared data. In order to address this challenge, studies of pragmatic interoperability of geodata are conducted in the NMRA project.

Table 6.1 Issues of geodata interoperability addressed in NMRA project

Interoperability issues	**Detailed approaches**
Pragmatic	Periodic semantic negotiations between evolving local ontologies and databases; Understanding and interoperability between local contexts; Consensus on understanding, use and potential result of shared geodata.
Semantic	Agreements on using a common ontology; Consistency of meanings of concepts among local ontologies; Improved understandings of shared geodata.
Schematic	Agreements on the general technical manual of involved map themes; Balance between standardization and flexibility: (1) Standardization of attributes table for each map layer; and (2) Flexibility of adapting structures of layers of map themes in local databases.
Syntactic	National and international standards-based geoscience terms and codes, map legends, coordinate systems and metadata schemas, etc.
Systemic	Internet-based geodata network among provincial and national sub-projects; Commonly used GIS software programs.

Although operated together with approaches for systemic, syntactic, schematic and semantic interoperability of geodata in the NMRA project, the approaches for pragmatic interoperability are different from the others (Table 6.1). Current approaches (e.g., those described in Section 1) for the interoperability of systems, syntaxes, schemas and semantics concentrate on the interoperability between machine information agents and, thus, their functions are limited for the pragmatic interoperability between the two geodata contexts. Moreover, in the case of more than two geodata contexts (i.e., 30 provincial sub-projects and 10 national sub-projects) of the NMRA project, semantic negotiations between them can become a tedious work. In view of these challenges, addressing geodata interoperability only between machine information agents is not an appropriate choice for either the NMRA project only or for the CGS agencies as a whole. To promote long-term geodata interoperability, pragmatic elements of contexts of distributed geodata must be taken into account and methods to achieve pragmatic interoperability between them must be investigated.

6.3 Achieving pragmatic interoperability of geodata

6.3.1 Representing geodata contexts

To achieve pragmatic interoperability of geodata, a proper representation of individual geodata contexts is essential. Researchers in linguistics (e.g., Levinson, 1983; Stalnaker, 1998; Mey, 2003) and philosophy (e.g., Morris, 1938; Kaplan, 1989; Crasnow, 2000) discussed that pragmatic elements in a context are a synthesis of objective facts (i.e., people, time, and location) and subjective dimensions (i.e., intention, procedure, and consequence), but they did not discuss in detail the methods to represent those object facts and subject dimensions. Ram and Park (2004) discussed, from the point of view of computer science, an extendable metadata schema for describing pragmatic elements of contexts and reducing semantic conflicts between contexts. Recently, Brodaric (2005, 2007) proposed three core aspects for representing pragmatic elements of contexts in geosciences: dimensions (i.e., origin, use, and effect), agents (i.e., human, machine, and nature) and entities (e.g., concepts, individuals, and states).

Brodaric's (2007) ideas about using human, machine, and nature as information agents in a geodata context are adapted in the study presented in this chapter, in which the machine agent includes a local ontology and a local database. A method is proposed for representing objective facts and subjective dimensions within a geodata context. Objective facts, including people, time, and location, represent elements related to credibility and responsibility of the local ontology and local database. Subjective dimensions, including intention, procedure, and consequence, represent elements related to systemic, syntactic, schematic, and semantic contents of the local ontology and local database.

Second-order logic (SOL) statements (Boolos, 1975; Hinman, 2005) are used for representing those pragmatic elements of a geodata context. The information agent I^A is a subset of a geodata context S, and I^A includes human information agent I_H^A, machine information agent I_M^A, and nature information agent I_N^A. I_M^A is a subset of I_H^A because the ontology and database in a geodata context are built by staff of that context. I_H^A is a subset of I_N^A because the tacit knowledge of every staff represents only a part of nature. Thus,

$\left\langle I_H^A, I_M^A, I_N^A \right\rangle \subseteq I^A \subseteq S$, and

$I_M^A \subset I_H^A \subset I_N^A$.

Symbols used for representing geodata contexts and semantic negotiations in this chapter are listed in Table 6.2.

Table 6.2 Symbols for representing geodata contexts and semantic negotiations

Symbol	Meaning
S	Geodata context
I^A	Information agent of S
I_H^A	Human information agent of S
I_M^A	Machine information agent of S
I_N^A	Nature information agent of S
O	Local ontology context of S
D	Local database context of S
O^O	Objective fact of O
O^S	Subjective dimension of O
D^O	Objective fact of D
D^S	Subjective dimension of D
O_P^O	People element in O^O
O_T^O	Time element in O^O
O_L^O	Location element in O^O
O_I^S	Intention element in O^S
O_P^S	Procedure element in O^S
O_C^S	Consequence element in O^S
D_P^O	People element in D^O
D_T^O	Time element in D^O
D_L^O	Location element in D^O
D_I^S	Intention element in D^S
D_P^S	Procedure element in D^S
D_C^S	Consequence element in D^S
C_D	Domain knowledge (A subset of I_H^A)

Symbol	**Meaning**
C_L	Local knowledge (A subset of I_H^A)
$f(C_D, C_L)$	Function for capturing and synthesizing C_D and C_L to generate O_C^S
C_C	Concept component of O_C^S
C_R	Relationship component of O_C^S
C_N	Name component of C_C
C_M	Meaning component of C_C
O_i, O_j	Two instances of O
$O_i.v1$, $O_i.v2$, $O_i.v3$, ...	Different versions of O_i
$O_j.v1$, $O_j.v2$, $O_j.v3$, ...	Different versions of O_j
N	Semantic negotiation between instances of O
$N.v1$, $N.v2$, ...	Different versions of N
O_n	Common ontology context of N
I_{Hn}^A	Human information agent of N
I_{Mn}^A	Machine information agent of N
I_{Nn}^A	Nature information agent of N
O_{Pn}^O	People element in O_n
O_{Tn}^O	Time element in O_n
O_{Ln}^O	Location element in O_n
O_{In}^S	Intention element in O_n
O_{Pn}^S	Procedure element in O_n
O_{Cn}^S	Consequence element in O_n
C_{Cn}	Concept component of O_{Cn}^S
C_{Rn}	Relationship component of O_{Cn}^S

Meanings of symbols for different versions of elements and components of O_i *,* O_j *and* N *can be derived from above symbols and are not listed here*

It is regarded in this study that a geodata context S includes (but is not limited to) two sub-contexts, namely a sub-context O focusing on a local ontology and a sub-context D focusing on a local database. The local ontology context O includes objective fact O^O and subjective dimension O^S . Likewise, the local database context D includes objective fact D^O and subjective dimension D^S . Thus,

$\langle O,D\rangle \subset S$,

$O = \langle O^O, O^S \rangle$, and

$D = \langle D^O, D^S \rangle$.

For a local ontology context O , the objective fact O^O includes people O_P^O , time O_T^O , and location O_L^O , whereas the subjective dimension O^S includes intention O_I^S , procedure O_P^S , and consequence O_C^S . Likewise, for a local database context D , the objective fact D^O includes people D_P^O , time D_T^O , and location D_L^O , whereas the subjective dimension D^S includes intention D_I^S , procedure D_P^S , and consequence D_C^S . Thus,

$O^O = \langle O_P^O, O_T^O, O_L^O \rangle$, $O^S = \langle O_I^S, O_P^S, O_C^S \rangle$, and

$D^O = \langle D_P^O, D_T^O, D_L^O \rangle$, $D^S = \langle D_I^S, D_P^S, D_C^S \rangle$.

People (i.e., O_P^O and D_P^O) managing the ontology and database are regarded as subsets of human information agent I_H^A in a geodata context. Locations (i.e., O_L^O and D_L^O) related to the ontology and database are regarded as subsets of nature information agent I_N^A . Consequences (i.e., O_C^S and D_C^S) of the ontology and database are regarded as subsets of machine information agent I_M^A . Thus,

$\langle O_P^O, D_P^O \rangle \subset I_H^A$,

$\langle O_L^O, D_L^O \rangle \subset I_N^A$, and

$\langle O_C^S, D_C^S \rangle \subset I_M^A$.

In subjective dimensions O^S of a local ontology context, the intention O_T^S is capturing both domain knowledge C_D and local knowledge C_L. The C_D and C_L are subsets of the tacit knowledge of human information agent I_H^A. Elements in the procedure O_P^S include people O_P^O, time O_T^O, location O_L^O, and a function $f(C_D, C_L)$ for capturing and synthesizing C_D and C_L in order to generate the consequence O_C^S. The O_C^S includes concept component C_C and relationship component C_R. Each concept consists of two components: name C_N and meaning C_M. Thus,

$O_I^S = C_D \bigcup C_L$, $\langle C_D, C_L \rangle \subset I_H^A$,

$O_P^S = \langle O_P^O, O_T^O, O_L^O, f(C_D, C_L) \rangle$,

$O_C^S = \langle C_C, C_R \rangle$, $C_C = \langle C_N, C_M \rangle$.

For a geodata context S, the local database consequence D_C^S is underpinned by the local ontology consequence O_C^S. If a method exists for achieving pragmatic interoperability between local ontology contexts of different geodata sources, then that method can also promote the pragmatic interoperability between local database contexts of these geodata sources.

6.3.2 Preconditions for semantic negotiations

Guha and McCool (2003) discussed that semantic negotiation is a process of conversations and agreements by which two programs (or contexts) can share larger vocabularies. Similarly, Garruzzo and Rosaci (2008, 2009) also discussed that semantic negotiation is a process by which multiple agents try to reach acceptable definitions mutually. In the study presented in this chapter, methods are investigated to use the pragmatic elements of geodata contexts in a procedure of semantic negotiations for achieving pragmatic interoperability between those geodata contexts. It is assumed that there are two geodata contexts S_i and S_j, each standing for a geodata source. Correspondingly, there are two local ontology contexts O_i and O_j. Thus, $S_i \supset \langle O_i, D_i \rangle$; $O_i = \langle O_i^O, O_i^S \rangle = \langle O_{Pi}^O, O_{Ti}^O, O_{Li}^O, O_{Ii}^S, O_{Pi}^S, O_{Ci}^S \rangle$; and $S_j \supset \langle O_j, D_j \rangle$; $O_j = \langle O_j^O, O_j^S \rangle = \langle O_{Pj}^O, O_{Tj}^O, O_{Lj}^O, O_{Ij}^S, O_{Pj}^S, O_{Cj}^S \rangle$.

The semantic negotiations between O_i and O_j is taken as a key part of the semantic negotiations between S_i and S_j, because O_i and O_j involve all elements of objective facts and subjective dimensions of S_i and S_j, and also because D_i and D_j are underpinned by O_i and O_j, respectively. Before holding a semantic negotiation between O_i and O_j, it is intuitively desired the credibility and responsibility of the two geodata sources that S_i and S_j stand for. For example, in the NMRA project, provincial sub-projects are handled by provincial agencies of CGS. Only these agencies are qualified to be provincial geodata sources for national sub-projects. The credibility and responsibility of a local geodata source are revealed by its professional capability, position in a social system, duty in a sub-project, etc. Such information is recorded in the objective facts (i.e., O_{Pi}^O, O_{Ti}^O, O_{Li}^O, O_{Pj}^O, O_{Tj}^O, O_{Lj}^O) of O_i and O_j. If the objective facts of S_i and S_j are both qualified, then the subjective dimensions of them can be analyzed. Generally, subjective dimensions of two local ontology contexts are different:

$$\forall O_i \forall O_j \left(\langle O_{Ii}^S, O_{Pi}^S, O_{Ci}^S \rangle \subset O_i ; \langle O_{Ij}^S, O_{Pj}^S, O_{Cj}^S \rangle \subset O_j \middle| \begin{array}{l} O_{Ii}^S \neq O_{Ij}^S ; \\ O_{Pi}^S \neq O_{Pj}^S ; O_{Ci}^S \neq O_{Cj}^S \end{array} \right).$$

Comparing similarities between O_i and O_j is an essential requirement for semantic negotiations between S_i and S_j. As mentioned above, the intention in a local ontology context is capturing both domain knowledge and local knowledge. For instance, the intention O_{Ii}^S of O_i is capturing the domain knowledge C_{Di} and local knowledge C_{Li}, which are both generated from the tacit knowledge of human information agent I_{Hi}^{A}. The procedure O_{Pi}^{S} of O_i synthesizes both C_{Di} and C_{Li}, and the resulting ontology O_{Ci}^{S} will be a collection of the concept component C_{Ci} and relationship component C_{Ri}. Thus,

$O_{Ii}^{S} = C_{Di} \bigcup C_{Li}$; $\left\langle C_{Di}, C_{Li} \right\rangle \subset I_{Hi}^{A}$;

$O_{Pi}^{S} = \left\langle O_{Pi}^{O}, O_{Ti}^{O}, O_{Li}^{O}, f\left(C_{Di}, C_{Li}\right) \right\rangle$; and

$O_{Ci}^{S} = \left\langle C_{Ci}, C_{Ri} \right\rangle$.

Similarly, there are intention O_{Ij}^{S}, procedure O_{Pj}^{S} and consequence O_{Cj}^{S} in the other local ontology context O_j. Both O_i and O_j are regarded as not static, but continuously evolving. There are different versions of O_i (i.e., $O_i.v1$, $O_i.v2$, $O_i.v3$, ...) and O_j (i.e., $O_j.v1$, $O_j.v2$, $O_j.v3$, ...) in the procedure of evolution.

The difference between local knowledge components C_{Li} and C_{Lj} leads to the difference between local ontology consequences O_{Ci}^{S} and O_{Cj}^{S}:

$$\forall O_i \forall O_j \left(\begin{array}{l} \left\{ C_{Li} \subset O_i, C_{Lj} \subset O_j \middle| C_{Li} \neq C_{Lj} \right\} \Rightarrow \\ \left\{ O_{Ci}^{S} \subset O_i, O_{Cj}^{S} \subset O_j \middle| O_{Ci}^{S} \neq O_{Cj}^{S} \right\} \end{array} \right).$$

However, there exist O_i and O_j having equivalent domain knowledge components C_{Di} and C_{Dj}, thus:

$$\exists O_i \exists O_j \left\{ C_{Di} \subset O_i, C_{Dj} \subset O_j \middle| C_{Di} = C_{Dj} \right\}.$$

The equivalent C_{Di} and C_{Dj} show the probability of a semantic negotiation N between O_i and O_j, being handled in a common ontology context O_n:

$$\exists O_i \exists O_j \left(\left\{ C_{Di} \subset O_i, C_{Dj} \subset O_j \middle| C_{Di} = C_{Dj} \right\} \Rightarrow \exists N \exists O_n \left\{ N = \left\langle O_i, O_j, O_n \right\rangle \right\} \right).$$

6.3.3 Semantic negotiations for achieving pragmatic interoperability

Kavouras and Kokla (2008, pp. 207–210) discussed four commonly used architectures for the negotiation and/or integration of ontologies: alignment, partial compatibility, unification, and true integration. If there are two local ontologies, then alignment means that a set of link nodes are made between similar parts of both local ontologies, while both original ontologies keep their autonomy and no common ontologies are generated. Partial compatibility means that similar parts of two local ontologies are merged and then the merged parts are joined with the remaining parts of both local ontologies to generate a common ontology. Meanwhile, the merged parts will replace the related original parts in both local ontologies to make updates. Unification means decomposing and synthesizing both local ontologies to generate a common ontology, and this common ontology will completely replace both original local ontologies. True integration means a common ontology is generated by integrating two local ontologies, but both original local ontologies keep their autonomy and no changes will be made.

The alignment architecture is an extremely flexible approach, while the unification architecture is an extremely standardized approach. As discussed in section 2, the alignment architecture will lead to tedious negotiations between multiple geodata contexts. The unification architecture will result in less negotiations but it will restrict the flexibility of local ontologies. However, for the NMRA, flexibility is necessary in every geodata context because each provincial sub-project deals with different geological background and mineral occurrences, and a part of these differences should be reflected in the local ontologies and local databases. Therefore, both the alignment and unification architectures are not appropriate for the NMRA project in practice. The true integration architecture can generate a common ontology but, because the original local ontologies still keep their autonomy and are unchanged, the function of the common ontology is limited in issues of geodata interoperability. Finally, the partial compatibility architecture seems a reasonable choice for the NMRA project, because it keeps a balance between flexibility and standardization for local ontologies and databases. With adaptations to the partial compatibility architecture, further inferences can be made for achieving pragmatic interoperability between S_i and S_j based on the aforementioned assumptions for two local ontology contexts O_i and O_j.

It is assumed that there is a semantic negotiation N by which two local ontology contexts O_i and O_j can be manipulated together with a common ontology context O_n . The O_n includes objective facts and subjective dimensions, which in turn include people O_{Pn}^{O} , time O_{Tn}^{O} , location O_{Ln}^{O} , intention O_{In}^{S} , procedure O_{Pn}^{S} , and consequence O_{Cn}^{S} . Relationships between objective facts of O_n and those of O_i and O_j are relatively transparent. The O_{Pn}^{O} includes people O_{Pi}^{O} and O_{Pj}^{O} , and is a subset of the human information agent I_{Hn}^{A} of O_n . The O_{Ln}^{O} includes locations O_{Li}^{O} and O_{Lj}^{O} , and is a subset of the nature information agent I_{Nn}^{A} of O_n . Thus,

$N = \left\langle O_i, O_j, O_n \right\rangle$;

$O_n = \left\langle O_{Pn}^{O}, O_{Tn}^{O}, O_{Ln}^{O}, O_{In}^{S}, O_{Pn}^{S}, O_{Cn}^{S} \right\rangle$.

$O_{Pn}^{O} = O_{Pi}^{O} \bigcup O_{Pj}^{O}$; $O_{Pn}^{O} \subset I_{Hn}^{A}$; and

$O_{Ln}^{O} = O_{Li}^{O} \bigcup O_{Lj}^{O}$; $O_{Ln}^{O} \subset I_{Nn}^{A}$;

Nevertheless, relationships between the subjective dimensions of O_n and those of O_i and O_j are not as transparent as the objective facts and, thus, should be investigated further. The intention O_{In}^{S} of O_n is to capture the domain and local knowledge of both O_i and O_j . O_{In}^{S} is achieved by the procedure O_{Pn}^{S} and the consequence is a common ontology O_{Cn}^{S} . The O_{Pn}^{S} is a bottom-up procedure based on O_{Ci}^{S} and O_{Cj}^{S} , but it is noteworthy that O_{Cn}^{S} is neither $O_{Ci}^{S} \bigcup O_{Cj}^{S}$ nor $O_{Ci}^{S} \bigcap O_{Cj}^{S}$, but is an adaption of both O_{Ci}^{S} and O_{Cj}^{S} resulting from the semantic negotiation between O_i and O_j . O_{Cn}^{S} includes concept component C_{Cn} and relationship component C_{Rn} . Thus,

$O_{In}^{S} = C_{Dn} \bigcup C_{Ln}$; $C_{Dn} = C_{Di} = C_{Dj}$; $C_{Ln} = C_{Li} \bigcup C_{Lj}$; $\left\langle C_{Dn}, C_{Ln} \right\rangle \subset I_{Hn}^{A}$;

$O^S_{Pn} = \left\langle O^O_{Pn}, O^O_{Tn}, O^O_{Ln}, f\left(C_{Dn}, C_{Ln}\right)\right\rangle$; and

$O^S_{Cn} = \left\langle C_{Cn}, C_{Rn}\right\rangle$; $O^S_{Cn} \subset I^A_{Mn}$.

As discussed in section 2, all ontologies evolve. Correspondingly, it is assumed that the semantic negotiation N between O_i and O_j is a continuous procedure, in which N, O^S_{Ci}, O^S_{Cj} and O^S_{Cn} all have evolving versions. For example, the first version of them can be $N.v1$, $O^S_{Ci}.v1$, $O^S_{Cj}.v1$ and $O^S_{Cn}.v1$ respectively:

$O^S_{Ij}.v1 = \left\langle C_{Di}.v1, C_{Li}.v1\right\rangle$; $\left\langle C_{Di}.v1, C_{Li}.v1\right\rangle \subset I^A_{Hi}.v1$;

$O^S_{Ij}.v1 = \left\langle C_{Dj}.v1, C_{Lj}.v1\right\rangle$; $\left\langle C_{Dj}.v1, C_{Lj}.v1\right\rangle \subset I^A_{Hj}.v1$;

$O^S_{Pi}.v1 = \left\langle O^O_{Pi}.v1, O^O_{Ti}.v1, O^O_{Li}.v1, f\left(C_{Di}.v1, C_{Li}.v1\right)\right\rangle$;

$O^S_{Pj}.v1 = \left\langle O^O_{Pj}.v1, O^O_{Tj}.v1, O^O_{Lj}.v1, f\left(C_{Dj}.v1, C_{Lj}.v1\right)\right\rangle$;

$O^S_{Ci}.v1 = \left\langle C_{Ci}.v1, C_{Ri}.v1\right\rangle$;

$O^S_{Cj}.v1 = \left\langle C_{Cj}.v1, C_{Rj}.v1\right\rangle$;

$N.v1 = \left\langle O^S_{Ci}.v1, O^S_{Cj}.v1, O^S_{Cn}.v1\right\rangle$.

$O^S_{Ci}.v1$ includes concept component $C_{Ci}.v1$ and relationship component $C_{Ri}.v1$. Similarly, $O^S_{Cj}.v1$ includes $C_{Cj}.v1$ and $C_{Rj}.v1$. It is assumed that $C_{Ci}.v1$ and $O^S_{Cj}.v1$ both include a finite number (e.g., to give a simplified example, the number is assumed to be four) of concepts, and $C_{Ri}.v1$ and $C_{Rj}.v1$ both include a finite number (e.g., five) of relationships:

$C_{Ci}.v1 = \left\langle C_{Ci1}, C_{Ci2}, C_{Ci3}, C_{Ci4}\right\rangle$; $C_{Ri}.v1 = \left\langle C_{Ri1}, C_{Ri2}, C_{Ri3}, C_{Ri4}, C_{Ri5}\right\rangle$;

$C_{Cj}.v1 = \left\langle C_{Cj1}, C_{Cj2}, C_{Cj3}, C_{Cj4}\right\rangle$; $C_{Rj}.v1 = \left\langle C_{Rj1}, C_{Rj2}, C_{Rj3}, C_{Rj4}, C_{Rj5}\right\rangle$.

To agree on concept component $C_{Cn}.v1$ and relationship component $C_{Rn}.v1$ in the first version of the common ontology $O_{Cn}^{S}.v1$, there will be a practical procedure of comparison, adoption and/or adaption happening between $O_{Ci}^{S}.v1$ and $O_{Cj}^{S}.v1$, thus:

$O_{In}^{S}.v1 = C_{Dn}.v1 \bigcup C_{Ln}.v1$; $C_{Dn}.v1 = C_{Di}.v1 = C_{Dj}.v1$;

$C_{Ln}.v1 = C_{Li}.v1 \bigcup C_{Lj}.v1$;

$O_{Pn}^{O}.v1 = O_{Pi}^{O}.v1 \bigcup O_{Pj}^{O}.v1$; $O_{Ln}^{O}.v1 = O_{Li}^{O}.v1 \bigcup O_{Lj}^{O}.v1$; and

$O_{Pn}^{S}.v1 = \left\langle O_{Pn}^{O}.v1, O_{Tn}^{O}.v1, O_{Ln}^{O}.v1, f\left(C_{Dn}.v1, C_{Ln}.v1\right)\right\rangle$.

For instance, the negotiation between concepts in these two local ontologies may happen in the following way:

1. $C_{Ci1}.C_N = C_{Cj1}.C_N \wedge C_{Cj1}.C_M = C_{Cj1}.C_M$ (i.e., C_{Ci1} and C_{Cj1} have same name and same meaning), and people participating in the negotiation (i.e., $O_{Pn}^{O}.v1$) agree to make no changes;
2. $C_{Ci2}.C_N \neq C_{Cj2}.C_N \wedge C_{Ci2}.C_M \neq C_{Cj2}.C_M$ (i.e., C_{Ci2} and C_{Cj2} are two different concepts), and $O_{Pn}^{O}.v1$ agree to keep both C_{Ci2} and C_{Cj2} ;
3. $C_{Ci3}.C_N \neq C_{Ci3}.C_N \wedge C_{Cj3}.C_M = C_{Cj3}.C_M$ (i.e., C_{Ci3} and C_{Cj3} are synonyms), and $O_{Pn}^{O}.v1$ agree to abandon C_{Ci3} but keep C_{Cj3};
4. $C_{Ci4}.C_N = C_{Cj4}.C_N \wedge C_{Ci4}.C_M \neq C_{Cj4}.C_M$ (i.e., C_{Ci4} and C_{Cj4} are homonyms), and $O_{Pn}^{O}.v1$ agree to keep C_{Ci4}, and update C_{Cj4} to a new version C_{Cj4}' by changing its name from $C_{Cj4}.C_N$ to $C_{Cj4}'.C_N'$.

Meanwhile, negotiations between relationships in these two local ontologies may happen in this way:

1. $C_{Ri1} = C_{Rj1}$, and $O_{Pn}^{O}.v1$ agree to make no changes;
2. $C_{Ri2} \neq C_{Rj2}$, and $O_{Pn}^{O}.v1$ agree to keep both C_{Ri2} and C_{Rj2} ;

3. $C_{Ri3} = C_{Rj3}$, and $O^O_{Pn}.v1$ agree to make no changes;
4. $C_{Ri4} \neq C_{Rj4}$, but $O^O_{Pn}.v1$ do not agree on either relationship, so they create a new relationship C_{Ri4}' to replace both C_{Ri4} and C_{Rj4}.
5. $C_{Ri5} = C_{Rj5}$, but $O^O_{Pn}.v1$ think both relationships should be revised, so they agree to use C_{Rj5}' to replace both C_{Ri5} and C_{Rj5}.

Consequently, the first version of a common ontology $O^S_{Cn}.v1$ is agreed on, thus:

$O^S_{Cn}.v1 = \left\langle C_{Cn}.v1, C_{Rn}.v1 \right\rangle$;

$C_{Cn}.v1 = \left\langle C_{Ci1}, C_{Ci2}, C_{Cj2}, C_{Cj3}, C_{Ci4}, C_{Cj4}' \right\rangle$; and

$C_{Rn}.v1 = \left\langle C_{Ri1}, C_{Ri2}, C_{Rj2}, C_{Rj3}, C_{Ri4}', C_{Rj5}' \right\rangle$.

Then, following the procedure of the partial compatibility architecture, $O^S_{Cn}.v1$ is used to update two local ontologies $O^S_{Ci}.v1$ and $O^S_{Cj}.v1$ into $O^S_{Ci}.v1'$ and $O^S_{Cj}.v1'$ respectively. The updating procedures in both O_i and O_j follow a top–down approach because they can refer to $O^S_{Cn}.v1$, thus:

$\left\langle C_{Di}.v1', C_{Li}.v1' \right\rangle \subset \left(I^A_{Hi}.v1' \cup O^S_{Cn}.v1 \right)$;

$\left\langle C_{Dj}.v1', C_{Lj}.v1' \right\rangle \subset \left(I^A_{Hj}.v1' \cup O^S_{Cn}.v1 \right)$;

$O^S_{Pi}.v1' = \left\langle O^O_{Pi}.v1', O^O_{Ti}.v1', O^O_{Li}.v1', f\left(C_{Di}.v1', C_{Li}.v1' \right) \right\rangle$; and

$O^S_{Pj}.v1' = \left\langle O^O_{Pj}.v1', O^O_{Tj}.v1', O^O_{Lj}.v1', f\left(C_{Dj}.v1', C_{Lj}.v1' \right) \right\rangle$.

Details of the concepts and relationships in $O^S_{Ci}.v1'$ and $O^S_{Cj}.v1'$ are:

$O^S_{Ci}.v1' = \left\langle C_{Ci}.v1', C_{Ri}.v1' \right\rangle$;

$C_{Ci}.v1' = \left\langle C_{Ci1}, C_{Ci2}, C_{Cj3}, C_{Ci4} \right\rangle$;

$C_{Ri}.v1' = \left\langle C_{Ri1}, C_{Ri2}, C_{Rj3}, C_{Ri4}', C_{Rj5}' \right\rangle$;

$O^S_{Cj}.v1' = \left\langle C_{Cj}.v1', C_{Rj}.v1' \right\rangle$;

$C_{Cj}.v1' = \left\langle C_{Ci1}, C_{Cj2}, C_{Cj3}, C_{Cj4}' \right\rangle$; and

$C_{Rj}.v1' = \left\langle C_{Ri1}, C_{Rj2}, C_{Rj3}, C_{Ri4}', C_{Rj5}' \right\rangle$.

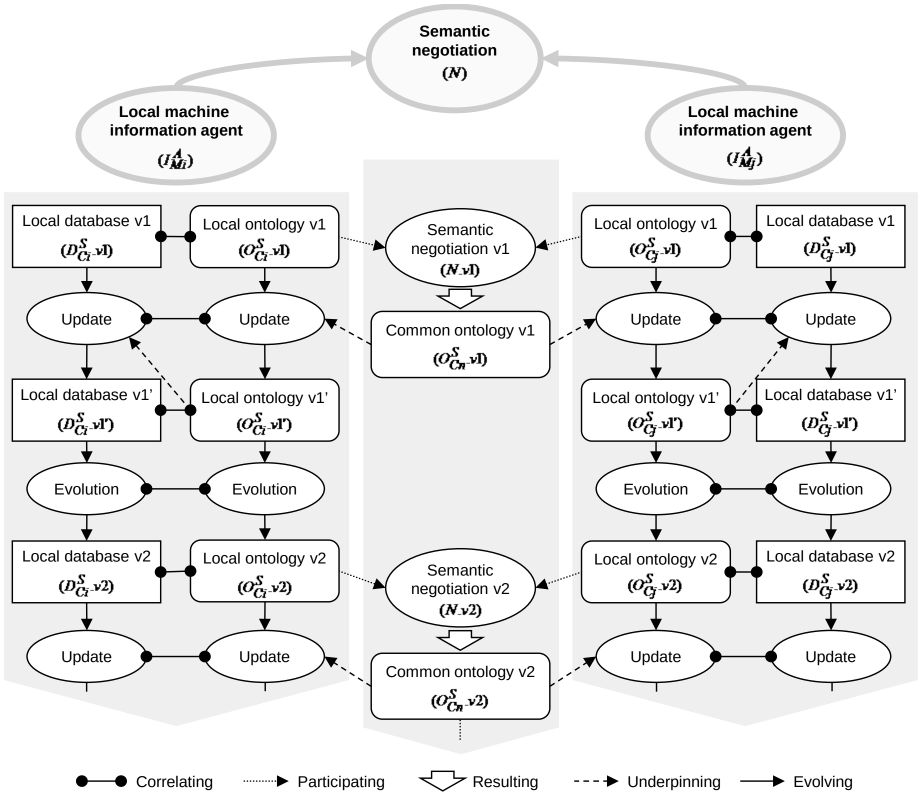


Fig. 6.2. Consequences of semantic negotiations between two evolving geodata contexts. The consequences are reflected in the common ontologies and the machine information agents of two local geodata contexts. See text for details of interactions between the human, nature and machine information agents of the two local geodata contexts.

Fig. 6.2 provides an overview of the afore-mentioned procedure of semantic negotiation, and the subsequent ontology-driven updates of local databases in both local contexts. $O^S_{Ci}.v1'$ and $O^S_{Cj}.v1'$ are used to update the two local databases $D^S_{Ci}.v1$ and $D^S_{Cj}.v1$ into $D^S_{Ci}.v1'$ and $D^S_{Cj}.v1'$, respectively. As discussed above, all ontologies are evolving, so there will be a second round of semantic negotiation $N.v2$, and then a third round $N.v3$, and so on. Correspondingly, new versions of common ontologies, local ontologies and local databases will emerge. For example, $O_{Ci}^{S}.v1'$ and $O_{Cj}^{S}.v1'$ will evolve to be $O_{Ci}^{S}.v2$ and $O_{Cj}^{S}.v2$, respectively. There will be a second round of semantic negotiation between them, aiming for a second version of a common ontology $O_{Cn}^{S}.v2$:

$\left\langle C_{Dn}.v2, C_{Ln}.v2 \right\rangle \subset I_{Hn}^{A}.v2$;

$O_{Pn}^{S}.v2 = \left\langle O_{Pn}^{O}.v2, O_{Tn}^{O}.v2, O_{Ln}^{O}.v2, f\left(C_{Dn}.v2, C_{Ln}.v2\right) \right\rangle$.

The semantic negotiations comprise a procedure by which the two geodata contexts S_i and S_j can deepen their understanding of each other. By using the agreed common ontologies, local ontologies in S_i and S_j are semantically consistent although they evolve continuously. Because the local databases in S_i and S_j are underpinned by the respective local ontologies, the interoperability of geodata stored in them will also be improved. The improvements are not limited to the interoperability between machine information agents (i.e., ontologies and databases) of S_i and S_j only. The semantic negotiations create an environment in which people of S_i can manage to know the elements of S_j , and vice versa. Therefore, when the people in S_j receive shared geodata from S_i , they understand not only the meaning of that geodata but also the background of that geodata. Subsequently, they can use that geodata appropriately in order to generate desired results and/or effects. In this way, the pragmatic interoperability between S_i and S_j are achieved.

6.4 Applications in NMRA project and results

In the past decades, the research groups the author worked in have cooperated together with CGS for several projects of geodata interoperability (e.g., Wang et al., 1999; Wu et al., 2005; Ma et al., 2007). In these works the concentration was the standardization of geoscience terms and the building and management of geo-database schemas. In the NMRA project, the previous work of a controlled vocabulary (Ma et al., 2005; Ma et al., 2007; Ma et al., 2010) was updated, and a program was developed to support the proposed procedure of semantic negotiations and the adaption and application of the resulting common ontology of these negotiations.

The program collected commonly used standards of geoscience terms in China (e.g., AQSIQ, 1988) and was used to generate draft conceptual schemas in semantic negotiations within the NMRA project. In the period of preparation of the NMRA project, staff from provincial and national sub-projects, CGS headquarters and some other research institutes organized periodic meetings and workshops, to negotiate and propose methodologies for the NMRA project. Latest studies (e.g., Wang, 1990; Zhai, 2003; Ye, 2004; Zhao, 2006; Chen, 2007; Cheng, 2007) on theories and technologies of regional mineral assessment had been synthesized by conversations and agreements in these workshops. Eventually, a general theoretical approach was agreed on[56]. The developed program was used to generate draft conceptual schemas for discussion and modification in these meetings and workshops, with an objective to make these schemas cover all subject domains of geodata involved in the general theoretical approach. The resulting conceptual schemas included definitions of 99 map themes, which comprised a general technical manual. Detailed items such as constitution of layers of each map theme, structure of attributes table of each layer, standard geoscience terms, map legends and metadata schemas, etc., were listed in the manual. Such efforts made the technical manual a common ontology for the 30 provincial sub-projects and 11 national sub-projects in the NMRA project.

After the finalization of the general technical manual, the developed program was updated to record all geoscience concepts and their relationships in the proposed 99 map themes. In the technical manual, each item (i.e., name of a map theme, layer name, attribute name, etc.) was a geoscience concept. Map themes, layers, and attributes table of layers comprised a hierarchical structure. The developed program recorded these concepts and their hierarchical relationships, and provided detailed definitive information (e.g., name, code, meaning, numerical restrictions, etc.) for each concept. Geospatial metadata elements (e.g., map title, producer, scale, map legends, projection systems, etc.) specified in the technical manual were also encoded as functional models in the program. Additionally, to promote flexibility, the technical manual specified a number of compulsory layers for each map theme, whereas other layers suggested in the same map theme were not mandatory. Such requirements were also included in the developed program.

---

[56] Technical approach – Introduction to NMRA project. http://imr.cags.ac.cn/qlpj/xiangmu_js/jishu_lx.html. [In Chinese, accessed April 24, 2011]

The developed program recorded the general technical manual as a common ontology for the NMRA project. Functions were developed in the program, such that it supported using the common ontology to update local ontologies and databases of provincial sub-projects. For example, researchers in a provincial sub-project can delete a non-compulsory layer in a map theme or add a new layer from another map theme according their needs. However, the structure of attributes table of each layer cannot be changed, although some columns in the table can be left empty if there is no data for them. A finished map theme can then be output as a file for a GIS program. In this file, both the standards of spatial data and the structure of attributes data are specified. Then, researchers can fill the file with local geodata and build up a digital map. The developed program provided flexibility of conceptual modeling for applications in sub-projects of NMRA, while it also kept standardization of compulsory map layers, attributes tables of layers, geoscience terms and geospatial metadata elements as specified in the general technical manual.

The provincial sub-projects used their geodata for mineral resources assessments, and transferred their original and result data to national sub-projects for integration and synthesis. Other functions were also developed in the program for national sub-projects to check the quality of collected maps. Researchers in a national sub-project can use the program to check compulsory map layers in each map theme. If a layer is missing, an attribute name is incorrect, or a geospatial metadata element is missing, etc., the program can recognize the problem and then show a notice to the researchers. If a layer in a collected map is not compulsory in the general technical manual, the program can still be used to check the completeness of the attributes table of the layer and the correctness of the geoscience terms inside. Moreover, the program can also be used like a vocabulary to check the definition of any geoscience term recorded in the general technical manual. Because the general technical manual made clear the structures and meanings of compulsory layers of each map theme to both provincial and national sub-projects, the developed program helped researchers in provincial and national sub-projects to control the quality of the compulsory layers. For those non-compulsory layers adapted by researchers in provincial sub-projects, the program also helped researchers in national sub-projects to understand their structures and meanings. After the checking and necessary refinements, geodata of provincial sub-projects were integrated and synthesized in national sub-projects, to generate the results of resources of different mineral deposits at the national scale.

In the middle of 2010, all provincial sub-projects had finished the assessment of potential resources of iron and aluminum deposits. Then, the provincial geo-databases were provided to several national sub-projects for integration and synthesis. Works of those national sub-projects were finished at the end of 2010. Results in the national projects showed that all compulsory map layers in provincial geo-databases were consistent with the general technical manual, and all those non-compulsory map layers adapted by provincial-projects were recognized by the developed program. The volumes of resources of iron and aluminum deposits assessed at the national level were also consistent with the accumulated results of provincial sub-projects.

6.5 Discussion

The developed program supported the procedure of semantic negotiations in the preparation period of the NMRA project, and helped researchers of this project to use the resulting general technical manual in provincial and national sub-projects to build consistent geo-databases. These works achieved consensus on understanding, use and potential result of geodata shared between provincial and national sub-projects and, thus, the pragmatic interoperability between them.

Context-dependent influences on building, understanding, and using of data (including geodata) have been increasingly studied recently. In the field of ubiquitous computing, context-awareness was discussed as an important characteristic of applications (e.g., those combined with mobile devices) because those applications have to adapt themselves to rapidly changing situations (Ranganathan and Campbell, 2003; Baldauf et al., 2007). In the field of data interoperability, representation of contexts was discussed as one of the key challenges for semantic interoperability issues (Kashyap and Sheth, 1996; Lee et al., 1996; Ouksel and Sheth, 1999). Recent studies (Singh, 2002; de Moor, 2005; Schoop et al., 2006) on the Pragmatic Web discussed that not only semantics but also pragmatics should be addressed in the sharing and using of data. The study of pragmatic interoperability described in this chapter shows that local ontologies and databases in geodata sources are both affected and featured by local contexts. Although the background of the study presented in this chapter is geosciences, the works discussed may also be useful for studies and applications in other fields.

Methods were developed to represent objective facts and subjective dimensions of a geodata context. Such representations considered both the common goal of the NMRA project and the diversity of actual works in sub-projects of NMRA. These methods are compatible to Frank's discussion

(2001) on the five tiers of ontology (i.e., human-independent reality, observation of physical world, objects with properties, social reality, subjective knowledge). The elements of objective facts and subjective dimensions in the studied model can also be compared with the basic model of “5W1H” (i.e., Who, When, Where, Why, What and How), which is widely used for information gathering in various fields of research (e.g., Shimazu et al., 2006; Hong et al., 2010). The “who, when, and where” can be fit into “people, time, and location” of objects facts, and “why, how, what” can be fit into “intention, procedure, and consequence” of subjective dimensions, respectively (Fig. 6.3).

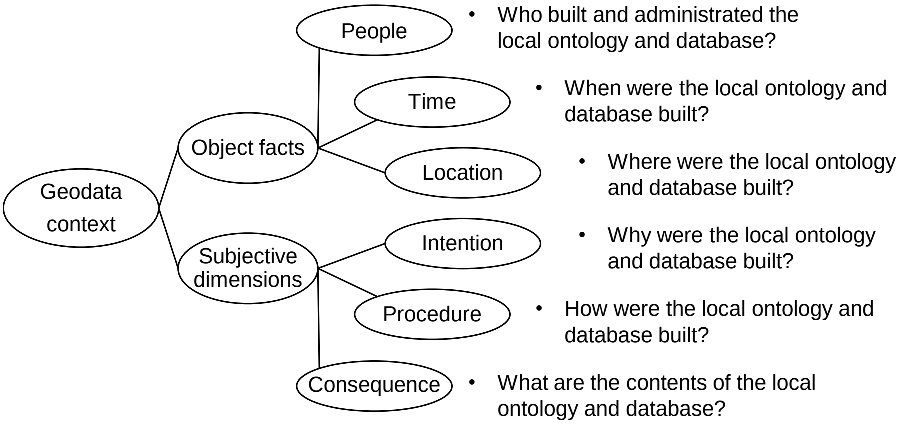


Fig. 6.3. Comparing pragmatic elements of a geodata context with questions of “5W1H” used for information gathering.

Negotiations and collaborations are regarded as a general approach for pragmatic interoperability. Several researchers (Bouquet et al., 2004; de Moor, 2005; Smith et al., 2007; Bobillo et al., 2008) discussed that negotiations and collaborations can smooth conflicts among local contextual ontologies and the agreed contents can be globalized to form a common ontology. Recent studies on participatory modeling (Voinov and Bousquet, 2010) show that a process of collaborative learning can enhance the stakeholders’ knowledge and understanding of a system and its dynamics under various conditions, and then help stakeholders to exchange information and build a common set of views and shared understanding about the system. In the study presented in this chapter, the proposed procedure of semantic negotiations involves all the pragmatic elements of geodata contexts, such that not only the local ontologies are compared and discussed, but also the respective geodata sources can improve their understanding of each other. Semantic negotiations can lead to a common ontology. However, geodata sources do not need to use the common ontology resulting from the proposed semantic negotiations to replace their local ontologies completely. Instead, they can use the common ontology to update their local ontologies, and in turn the updated local ontologies will be used to update local databases. In this way, the procedure of semantic negotiations keeps a balance between standardization and flexibility of distributed geodata, and improves the pragmatic interoperability among them.

Recently, approaches of semantic negotiations for pragmatic interoperability have been put into practices in several geoscience projects. For example, the OneGeology project (Jackson, 2007) aimed at making geological map data of the world accessible via the web. Firstly, experts from several countries worked together and proposed the GeoSciML (Sen and Duffy, 2005) as a common conceptual schema of geological maps. Then, GeoSciML was used to mediate geological map databases provided by different countries. Finally, users can retrieve data from these distributed databases and browse them via a unified web portal. The GeoSciML is continuously updating, which also leads to updates of services in the OneGeology project. In the OneGeology-Europe project (Laxton et al., 2010) there is a similar procedure of negotiations and collaborations when a common geoscience vocabulary was proposed and built by participants from several European countries (Asch, 2010). Another example is the U.S. National Geologic Map Database (NGMDB) project (Soller and Berg, 2005), which takes the North American Geologic Map Data Model (NADM) (NADM Steering Committee, 2004) as a general conceptual schema for mediating and integrating geological map databases of different states in US. The procedure of semantic negotiations discussed in this chapter addresses a similar point of view that negotiations can reduce heterogeneities and improve interoperability among geodata sources, but it also pointed out that the flexibility of local ontologies should be remained, though they can improve their interoperability by referring to the output (e.g., a common ontology) of negotiations.

This work suggests a few directions for further studies and practices. One is that the proposed model for geodata contexts can be enriched. Although the proposed pragmatic elements of information agents, objective facts and subjective dimensions cover essential elements of a geodata context, in the study presented in this chapter the emphasis was put on subdivisions and inferences of elements in subjective dimensions. As the representation of a context is a broad domain (cf. Grudin, 2001), more elements can be collected or current elements in the model can be subdivided, such that these new elements can be used to simulate potential situations (e.g., conflicts caused by people, time or location, etc.) in geodata interoperability and to propose solutions. Another direction is considering the organizational rules together with technical issues for improving geodata interoperability. Although addressing issues of systems, syntaxes, schemas, semantics and pragmatics can improve the interoperability of geodata from the technical side, issues from the organizational side are also necessary when these technical issues are put into practice (cf. Tolk, 2003). In a large project like the NMRA, improving the organizational rules can be a nice complement to the technical issues for addressing geodata interoperability.

6.6 Conclusions

Geological data are related to and affected by its surrounding context. Pragmatic elements in a geological data context include information agents, objective facts and subject dimensions, etc. Semantics cannot fully solve the pragmatic interoperability issues between geological data contexts because semantics concentrates on machine information agents of geological data. In the study presented in this chapter, a model of geological data contexts was demonstrated and it was used to interpret a procedure of semantic negotiations between geological data contexts. These works were applied in the National Mineral Resources Assessment project of China by developing and using a computer program. Applications and results proved that this program helped to achieve consensus on understanding, use and potential result of shared geological data. This study shows that awareness of contexts and a procedure of semantic negotiations are useful to improve the pragmatic interoperability of distributed geological data in a long-term perspective. Chapters 2–6 have described and discussed methods to answer the five research questions stated in Chapter 1. The following chapter 7 summarizes the results of the studies described in preceding chapters and presents answers to the research questions.

Chapter 7

Synthesis

Geological data are important not only for studies of the earth but also for approaches coping with societal affairs in economic development, environmental protection and hazard mitigation, etc. Interoperability of geological data is the focus of considerable attention in recent years, and different types of ontologies (i.e., an ontology spectrum) have been studied to address this challenge. Accordingly, the background of this study (see Chapter 1) is that:

> *Substantial progress has been made in developing geological ontologies and using them to mediate heterogeneous geological data, in which the capability of ontologies for promoting geological data interoperability is commonly acknowledged.*

Although this statement indicates that applying an ontology spectrum is a mainstream approach for problem-solving in the field of geological data interoperability, current and potential practices of this mainstream approach are faced with the following key challenges, which are addressed in this dissertation:

> *(1) Modeling and encoding of ontologies;*
> *(2) Multilinguality of geological data and ontologies;*
> *(3) Flexibility and usefulness of ontology-based applications; and*
> *(4) Mediation and evolution of geological data and ontologies.*

In respect of the background and the challenges, the objective of this study is defined as:

> *To explore approaches to address the aforementioned key challenges in applying an ontology spectrum to promote geological data interoperability, and to answer five research questions stated in Chapter 1.*

The studies presented in Chapters 2–6 described and discussed solutions or methods that were investigated to answer the five research questions, respectively. The reminder of this chapter first summarizes the results discussed in Chapters 2–6 and their inter-relationships, and then provides answers to the research questions and draws main conclusions. The main contributions of the research studies described in this dissertation are discussed relevant to the aforementioned four key challenges. This chapter ends with recommendations for further work.

7.1 Summary of results and their inter-relationships

(1) *Modeling and encoding of ontologies to address both internal and external aspects of local geological data interoperability.*

In Chapter 2, a pure hierarchical structure (Fig. 2.3) was used in a controlled vocabulary to organize professional terms of 27 subjects in geological data of mining projects. This structure is simple but is functional enough, because the vocabulary was mainly used to promote standard terms in local contexts of mining projects. The vocabulary was encoded as spreadsheets (Figs. 2.7a–f) to support applications in relational databases. Each term in the vocabulary was labeled with both Chinese and English names and was tagged with a unique alphanumerical code. Moreover, several national standards were adapted in the vocabulary to improve the interoperability of the vocabulary and the geological data standardized by it.

Distributed geological data sources used in the studies described in Chapters 3 and 5 are published online and thus are within a regional/global environment, but they are maintained within their local contexts. Works in these two chapters show that if distributed local geological data sources do not address standards strictly (e.g., using synonyms of standard geological terms), then the ontologies being used to mediate distributed geological data sources should be comprehensive (e.g., collecting synonyms as many as possible).

Compared to the 27 subjects covered by the controlled vocabulary in Chapter 2, the thesaurus described in Chapter 3 is much smaller because it focused only on the subject of geological time. However, the structure of this thesaurus is more complex, because both hierarchical and ordinal relationships (Fig. 3.2b) between geological time concepts were represented in the thesaurus, and annotations were collected to define those concepts. The ordinal hierarchical structure and the annotations of the thesaurus were encoded with an extended SKOS model (Fig. 3.2a), and were used in a pilot system to explain the meanings of geological time concepts to users. Similar to the operations of adapting national standards in a controlled vocabulary in

Chapter 2, the thesaurus in Chapter 3 adopted the International Stratigraphic Chart as the basic conceptual reference, and it collected many synonyms of geological time terms to mediate heterogeneous geological data.

The RDF/OWL-based ontology of geological time scale in Chapter 5 refines the thesaurus in Chapter 3. This ontology defined chronostratigraphic units (i.e., Eonothem, Erathem, System, Series and Stage) as classes, and it then defined all geological time concepts as instances of these classes (Fig. 5.1c). It refined the ordinal hierarchical structure by replacing the relationships “skos:broader” and “skos:narrower” in the thesaurus (Fig. 3.2b) with “gts:subsetOf” and “gts:supersetOf”, respectively (Fig. 5.1c). The ontology also includes annotations of geological time concepts from international standards and commonly used glossaries. Those annotations were used to develop featured functions with online geological maps in a pilot system.

(2) *Using multilingual ontologies to alleviate linguistic barriers of online geological data in a regional/global environment.*

While inaccuracy, inconsistency and incompleteness are challenges faced by multilingual thesauri of many subjects in geology, the thesaurus in Chapter 3 concentrates on the subject of geological time scale. It did not follow an alphabetical or alphanumerical sequence in the arrangements of terms. Instead, it took the International Stratigraphic Chart as a basic conceptual reference and arranged chronostratigraphic terms in an ordinal hierarchical structure (Fig. 3.1). Because the International Stratigraphic Chart is in English, the thesaurus used English terms to set up an initial structure and then labeled terms in six other languages. In this way, the accuracy and consistency of the thesaurus were achieved, and the completeness was partly achieved because there were synonyms of these multilingual terms to be collected. The thesaurus used an extended SKOS model for encodings, which not only represented the ordinal hierarchical structure but also provided properties to encode multilingual chronostratigraphic terms as preferred labels and multilingual geochronological terms and synonyms as alternative labels. A pilot system (Fig. 3.5) was set up to recognize and translate geological time terms from online geological maps. Results show that properly deployed multilingual geological thesauri are functional for alleviating linguistic barriers of online geological data and improving their interoperability.

Although the RDF/OWL-based ontology in Chapter 5 updated both object and datatype properties from the thesaurus in Chapter 3 to make an enriched representation of geological time scale, its multilingual labels (Fig. 5.1) were inherited from the thesaurus in Chapter 3. The difference is that the multilingual labels in the ontology of Chapter 5 were not used for translation, but for the recognition of geological time concepts regardless in which of the seven languages (i.e., those covered by the ontology) they were encoded. The pilot system (Fig. 5.9) in Chapter 5 developed a function of annotations by quoting the ontology, but currently these detailed annotations are in English only.

The size of the controlled vocabulary in Chapter 2 is much bigger than that of the thesaurus in Chapter 3. The developed controlled vocabulary adapted several national standards in China and provided terms in Chinese and English. In case studies in a mining group, the controlled vocabulary was used to standardize geo-databases for mine exploration applications. Records in a standardized database were translated from Chinese into English easily by using the controlled vocabulary.

(3) *Integrating inter-complementary methods of conceptual analysis to develop standard-compatible conceptual schemas for problem-solving in geology*.

In Chapter 4, the data-flow models (Figs. 4.2, 4.3 and 4.7) and object-oriented models (Fig. 4.8) define diverse relationships between classes, sets and instances of borehole metal-grade intervals and composites. These models have more complex structures than the controlled vocabulary in Chapter 2 and the thesaurus in Chapter 3. Data-flow models show progressing steps in the composting of borehole metal-grade intervals. Object-oriented models give a clear expression of objects and their relationships in the compositing procedure. The two types of models complement each other, however. The data-flow models were drawn as flow charts and the object-oriented models were encoded with UML schemas. Finally, the designed models were transformed into computer programs of a pilot system.

The two key attributes – metal-grade and length – of borehole intervals and composites in the models relate to profitability and minability of ore bodies outlined using the borehole composites. The two attributes were used in the compositing, dilution and classification of waste composites, economic but unminable composites and economic and minable composites. Such attributes and classes are compatible with commonly used standards for mineral resources estimation and, thus, improve the interoperability of the developed models and the results of compositing borehole metal-grade intervals.

In Chapter 4, annotations of attributes, sets and classes in the models were collected, and used in the user interface (Fig. 4.9) of the pilot system to explain the meanings of symbols, but they were not included in the flow charts or UML schemas due to the limited space of graphic views in these models. The annotations of geological time concepts collected in Chapters 3 and 5 were encoded as source codes in the SKOS-based thesaurus and the RDF/OWL-based ontology, respectively. They were retrieved and shown on the user interfaces of pilot systems (Figs 3.5, 5.9) dynamically following the operations of users.

(4) *Incorporating ontologies into state-of-the-art technologies in geo-information science to promote interoperability of online geological data for both geologists and non-geologists*.

In Chapter 5, a RDF/OWL-based ontology of geological time scale was built and several functions underpinned by the ontology were developed. One featured function based on the ontology is automatic annotations (Figs. 5.1a and 5.9) for geological time concepts recognized from online geological maps (i.e., WMS layers). Besides this, a Flash animation of geological time scale (Fig. 5.2) was built to visualize the ontology, and then more interactive functions were developed based on the Flash animation. One of them is changing the layout of the animation automatically following input queries of geological time concepts recognized from online maps (Fig. 5.4). Other interactive functions include using the animation to show legends of geological time features of online maps (Fig. 5.6), and using the legends as operation panels to filter out and generalize geological time features in the maps (Figs. 5.7 and 5.11). Several cases of conceptual mappings, such as “Lower Cambrian = Terreneuvian + Series 2” and “Middle Cambrian = Series 3”, were addressed in these functions. Because the ontology, the Flash animation and the WMS map layers were all developed with Web-compatible formats, it was efficient to combine them together in the Web environment. The aforementioned functions were included in a pilot system and tested by participants in a user-survey (Tables 5.1 and 5.2). Results show that these functions are helpful for both geologists and non-geologists to understand geological time contents in online geological maps and to explore more information from the maps.

Although the thesaurus in Chapter 3 relates to a same topic – geological time scale – as that of the ontology in Chapter 5, the works in Chapter 3 highlighted the functions developed for multilingual translations of geological time concepts recognized from online geological maps. These functions translate not only the geological time terms but also short annotations of the terms and the operational instructions on the user interface (Fig. 3.5). Another group of functions described in Chapter 3 is the methods of characteristic-oriented term retrieval (Fig. 3.4) developed in JavaScript programs for recognizing geological time concepts. The accuracy and functionality of these functions were proved in a pilot system.

The controlled vocabulary in Chapter 2, although not used online, were functional for standardizing geo-databases used in mine exploration projects and improved their interoperability with external projects. With the standardized geo-databases, some efficient functions were developed. One example function described in Chapter 2 is the automatic mapping of borehole logs (Fig. 2.9).

In Chapter 4, the object-oriented and data-flow models were regarded as ontologies resulting from conceptual analyses. The programs developed with C++ were based on these models, and in the pilot system they helped people to conduct compositing works and to obtain standard-compatible metal-grade composites in the results (Fig. 4.9).

(5) *Representing contexts of geological data and achieving pragmatic interoperability of distributed and evolving geological data in a long-term perspective*.

In Chapter 6, it was discussed that both geological data and ontologies are usually affected by contexts in which they are built and organized. The issues of geological data interoperability were considered in a flowing environment and one-station-stop approaches are not suitable for solving challenges in these issues. In this regards, the topic of pragmatic interoperability was proposed in Chapter 6. A model of information agents, object facts and subjective dimensions was designed to represent the contexts of geological data sources. Second-order logic statements were used to encoded objects and relationships in this model, based on which a procedure of semantic negotiations was conducted for approaching pragmatic interoperability among distributed geological data and ontologies (Fig. 6.2). The methods discussed in Chapter 6 have been used in the National Mineral Resources Assessment project of China and the results prove that these methods are effective for improving geological data interoperability among the sub-projects of this project.

Although the main topic of Chapter 2 was different from that of Chapter 6, it was also proposed that a controlled vocabulary should have an open structure in order that new terms can be added when they are found in actual geological works. In addition, Chapter 2 discussed that negotiations and collaborations among stakeholders in the same or related knowledge domains are helpful for promoting wider acceptability and interoperability of a controlled vocabulary.

The issues of collaborative modeling and pragmatic interoperability were not discussed in Chapter 3, but this chapter mentioned the basal time boundary of Quaternary as a notable case of the evolving meanings of geological time concepts. This case shows that properties and/or meanings of concepts in geological ontologies (and also geological data) may change in a long-term perspective. If such issues are not addressed properly, confusions and misunderstandings may arise between mismatched geological ontologies and data.

7.2 Answers to research questions and main conclusions

Based on the summary of the results, answers can be given to the five research questions specified in Chapter 1:

(1) *How can ontologies be modeled and encoded, so that the resulting ontologies are not only efficient for harmonizing local geological data but also function to improve the interoperability of local or internal geological data with extramural or external projects?*

A strategy of “global thoughts and local actions” should be deployed. Ontologies can be encoded in different formats (i.e., syntactic variability) according to the requirements of local actions, whereas concise or precise definitions (i.e., semantic variability) of concepts and their inter-relationships in ontologies depend on the works of modeling, in which the global thoughts of interoperability should be addressed if local geological data are underpinned by the built ontologies and are to be shared with extramural projects.

(2) *In a regional/global environment, how can linguistic barriers of online geological data be alleviated by building and using multilingual ontologies?*

There are three core requirements for a multilingual geological ontology: accurate, consistent, and complete. The multilingual labeling method is easier for use than the interlingual mapping method in building ontologies, whereas the former also requires that terms in different language share a common conceptual structure. Geological data are increasingly shared online and, thus, are brought into regional/global environments. If the concepts in those geological data can be recognized by multilingual ontologies, they can be translated in to different languages by retrieving multilingual labels in the ontologies. In this way, multilingual geological ontologies alleviate linguistic barriers of online geological data.

(3) *How can different methods of conceptual analysis be integrated to develop thematic conceptual schemas that are efficient for problem-solving and are compatible with commonly used standards in the field of geology?*

Different methods of conceptual analysis can be used to build conceptual schemas for a thematic work in geology. Each method has its own feature and can address the challenges of a work from a special aspect. For example, the data-flow method can analyze steps in problem-solving, whereas the object-oriented method can analyze objects and relationships in the same work. A more effective way is, however, to apply inter-complementary methods in conceptual analysis of a work and, thus, obtain a comprehensive understanding of the problem and the solution. Some logic theories, such as second-order logic and description logic, can be applied as a medium to integrate different methods in conceptual analysis. In addition, commonly used attributes and classes in a domain should be adopted in the conceptual analysis to improve the interoperability of the resulting conceptual schemas.

(4) *How can ontology-based tools be developed to improve the interoperability of online geological data, so as to help both geologists and non-geologists to retrieve geological information and discover geological knowledge?*

Only very few people are familiar with the background knowledge of geology. Geological ontologies store conceptualizations of domain knowledge in geology and can be used as knowledge references to explain the meaning of geological data. Ontologies can be incorporated into state-of-the-art technologies in geo-information science to develop applications for improving geological data interoperability. The contexts of the geological data should be considered in the developments of these applications. The usefulness of these applications – their capabilities to help people to access, decode, understand and appropriately use data – should be evaluated according to the objectives of data interoperability in a certain context.

(5) *What are the context-caused challenges for geological data interoperability, and how can these challenges be addressed in a long-term perspective?*

People's understanding of the earth is evolving and their conceptualizations of geological knowledge are not fixed. Such evolutions in ontologies also affect the contents of geological data if the data are underpinned by the ontologies. While the diversity of geological studies and conceptualizations in local contexts should be allowed and protected, it is also desirable that local geological data and ontologies can be understood and used by people in other contexts. Regular semantic negotiations and updates of a common ontology among local contexts in the same domain is an effective paradigm to achieve this objective.

The main conclusions of this study are:

Geological ontologies and geological data are evolving in a long-term perspective. Common understanding of subjects in geology requires semantic negotiations among stakeholders and domain experts. Different types of ontologies in an ontology spectrum can be used to encode the commonly agreed models of subjects in geology. By using ontologies, innovative applications can be developed to promote geological data interoperability at local, regional and global levels.

7.3 Main contributions

The main contributions of the research studies described in this dissertation are:

The importance of distinguishing and at the same time bridging between modeling and encoding geological ontologies has been addressed. The practices of "global thoughts and local actions" in modeling and encoding of various types of geological ontologies considered in this study provide experiences and lessons for other ongoing or in-preparation works of geological ontologies, as well as ontologies of non-geological domains.

The method of multilingual labeling for building accurate, consistent and complete multilingual geological ontologies has been explored and developed. The case study of multilingual thesaurus of geological time scale uses English terms as the basic reference while others as labels. In actual works, terms in any language can be used as the basic reference if the terms in different languages share a common conceptual structure. It has been shown in this dissertation that multilingual geological ontologies encoded in Web-compatible formats have great potential to improve the interoperability of online geological data.

Flexible functions underpinned by ontologies, such as automatic mapping of borehole logs, annotations and animations showing details of geological time concepts, conceptual mappings between geological terms, filtrations and generalizations of spatial features in online geological maps, etc., have been examined and developed. These functions are proven useful to help both geologists and non-geologists to understand contents in geological maps and conduct further operations. With these works, this dissertation expresses an opinion that users of geological data consist of not only geologists and earth scientists, but also other scientists and the general public, and there is still a lot of work to do if stakeholders want to improve the understandability of their geological data for a broader community of users.

The pragmatic interoperability of geological data in actual works has been discussed and a method of semantic negotiations between geological data contexts as an approach to address this challenge has been proposed. This proposed method puts the issues of geological data interoperability in a long-term perspective and suggests that those issues should be addressed sustainably.

As a whole, the studies described in this dissertation provide a route map for stakeholders who are willing to use ontologies to promote geological data interoperability at local, regional and global levels.

7.4 Recommendations for further work

Based on the outcomes of the research described in this dissertation, several recommendations for further studies can be provided.

(1) Enrichment of semantic expressions of developed ontologies. The developed RDF/OWL-based ontology of geological time scale can be enriched. Specifications of relationships between geological time concepts can be refined. Terms and annotations in more languages can be labeled, and more synonyms and conceptual mapping examples can be collected. Following the explored methods, ontologies of other geological subjects, such as rock types, mineral types and fossils, etc., can also be developed.

(2) Development of methods for conceptual mapping/matching between geological ontologies. Conceptual mapping/matching between ontologies described in this dissertation and ontologies developed by other researchers can be further explored. Mapping between ontologies in the same knowledge domain can potentially provide efficient approaches to address issues of geological data interoperability.

(3) Application of fuzzy logic and faceted classifications to build ontologies for subjects in geology. The ontologies described in this dissertation follow the crisp logic, such as second-order logic and fixed hierarchical classifications. Fuzzy logic and faceted classifications can be studied for building ontologies of some subjects in geology.

(4) Further development of innovative ontology-based applications. Applications of ontologies with latest geo-information technologies need to be developed and tested further for the field of geology. Geo-information technologies are fast evolving and many new methods and technologies are being developed. In this study, applications of ontologies with Web Map Service (WMS) of geological data have been developed. Applications of ontologies with Web Feature Service (WFS), Web Processing Service (WPS) and other technologies can be further explored.

(5) Refinement of models of geological data contexts. New elements or subdivisions of existing elements must be examined to enrich the current model of a geological data context. Fuzzy logic and other methods can be tested to rebuild the procedure of semantic negotiations, which is currently based on second-order logic.

References

Agterberg, F.P., 1989. Computer programs for mineral exploration. Science 245 (4913), 76–81.

Agarwal, P., 2005. Ontological considerations in GIScience. International Journal of Geographical Information Science 19 (5), 501–536.

Allison, M., Gundersen, L.C., Richard, S.M., Dickinson, T.L., 2008. Geosciences Information Network (GIN): a modular, distributed, interoperable data network for the geosciences. Eos Transactions 89 (53), American Geophysical Union Fall Meeting 2008, Abstract No. IN2013A-1073.

Anandakrishnan, S., Blankenship, D.D., Alley, R.B., Stoffa, P.L., 1998. Influence of subglacial geology on the position of a West Antarctic ice stream from seismic observations. Nature 394 (6688), 62–65.

Annels, A.E., 1991. Mineral Deposit Evaluation: A Practical Approach. Chapman & Hall, London, UK, 436 pp.

ANSI/NISO, 2005. ANSI/NISO Z39.19-2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. American National Standards Institute/National Information Standards Organization (ANSI/NISO). National Information Standards Organization Press, Bethesda, MD, USA, 172 pp.

Antoniou, G., Franconi, E., van Harmelen, R., 2005. Introduction to semantic web ontology languages. In: Eisinger, N., Maluszynski, J. (Eds.), Reasoning Web, Lecture Notes in Computer Science, vol. 3564. Springer-Verlag, Berlin & Heidelberg, pp. 1–21.

AQSIQ, 1988. GB/T 9649-1988 The Terminology Classification Codes of Geology and Mineral Resources. General Administration of Quality Supervision, Inspection and Quarantine of P.R. China (AQSIQ). Standards Press of China, Beijing, China, 1937 pp. [In Chinese and English].

Asch, K., 2001. The IGME 5000: a means to pan-European geological standards (Or just another local and transient convention?) In: Proceedings of 2001 Annual Conference of the International Association for Mathematical Geology, Cancún, Mexico, 7 pp., http://www.kgs.ku.edu/Conferences/IAMG/Sessions/I/Papers/asch_o.pdf [Accessed February 07, 2011].

Asch, K., 2005. The new digital geological map of Europe and standardisation: consistency as the last refuge of the unimaginative?! In: Ostaficzuk, S. (Ed.), The Current Role of Geological Mapping in Geosciences. Springer Netherlands, pp. 1–8.

Asch, K., 2010. The challenge of building a vocabulary for the geology of Europe. IUGS-CGI and OneGeology-Europe International Geoscience Language Workshop (IGSL 2010), 25-26 August 2010, Berlin, Germany. (18 slides), http://www.bgr.bund.de/cln_144/nn_1960030/EN/Themen/GG__geol__Info/IGSL2010/Downloads/Asch__IGSL__2,templateId=raw,property=publicationFile.pdf/Asch_IGSL_2.pdf [Accessed February 07, 2011].

Asch, K., Jackson, I., 2006. Commission for the Management & Application of Geoscience Information (CGI). Episodes 29 (3), 231–233.

Asch, K., Laxton, J., Bavec, M., Bergman, S., Perez Cerdan, F., Declercq, P.Y., Janjou, D., Kacer, S., Klicker, M., Nironen, M., Pantaloni, M., Schubert, C., 2010. Explanatory Notes for the Vocabulary to Describe Spatial Geological Data in Europe at a 1 : 1 Million Scale — for the eContentPlus project OneGeology-Europe. 86 pp., http://www.bgr.bund.de/cln_109/nn_1960030/EN/Themen/GG__geol__Info/IGSL2010/Downloads/WP3__Explanatory__Notes__and__Vocabulary,templateId=raw,property=publicationFile.pdf/WP3_Explanatory_Notes_and_Vocabulary.pdf [Accessed January 11, 2011].

Ashburner, M., 2000. Gene ontology: tool for the unification of biology. Nature Genetics 25, 25–29.

Auerbach, S., Schaeben, H., 1990. Computer-aided geometric design of geologic surfaces and bodies. Mathematical Geology 22 (8), 957–987.

Babaie, H.A., Oldow, J.S., Babaei, A., Lallemant, H.G.A., Watkinson, A.J., 2006. Designing a modular architecturefor the structural geology ontology. In: Sinha, A.K. (Ed.), Geoinformatics: Data to Knowledge: Geological Society of America Special Paper 397. Geological Society of America, Boulder, CO, USA, pp. 269–282.

Baeza-Yates, R., Ribeiro-Neto, B., 2011. Modern Information Retrieval, 2nd Edition. Addison-Wesley, Harlow, England, 870 pp.

Baldauf, M., Dustdar, S., Rosenberg, F., 2007. A survey on context-aware systems. International Journal of Ad Hoc and Ubiquitous Computing 2 (4), 263–277.

Baru, C., Chandra, S., Lin, K., Memon, A., Youn, C., 2009. The GEON service-oriented architecture for Earth Science applications. International Journal of Digital Earth 2 (supp. 1), 62–78.

Batanov, D., Vongdoiwang, W., 2007. Using ontologies to create object model for object-oriented software engineering. In: Sharman, R., Kishore, R., Ramesh, R. (Eds.), Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems. Springer, New York, pp. 461–487.

Bell, F.G., 2003. Geological Hazards: Their assessment, Avoidance and Mitigation. Spon Press, London, 656 pp.

Bermudez, L., Piasecki, M., 2006. Metadata community profiles for the semantic web. GeoInformatica 10 (2), 159–176.

Berners-Lee, T., Hendler, J. and Lassila, O., 2001. The Semantic Web. Scientific American 284 (5), 34–43.

BGS, 2005. Britain Beneath Our Feet. British Geological Survey (BGS). CD-ROM.

Bibby, L., 2006. Establishing vocabularies for the exchange of geological map data? how to herd stray cats. ASEG Extended Abstracts 2006 (1), doi:10.1071/ASEG2006ab1015.

Bishr, Y., 1998. Overcoming the semantic and other barriers to GIS interoperability. International Journal of Geographical Information Science 12 (4), 299–314.

Bishr, Y., 2006. Geospatial Semantic Web. In: Rana, S., Sharma, J. (Eds.), Frontiers of Geographic Information Technology. Springer Berlin Heidelberg, pp. 139–154.

Bobillo, F., Delgado, M., Gómez-Romero, J., 2008. Representation of context-dependant knowledge in ontologies: a model and an application. Expert Systems with Applications 35 (4), 1899–1908.

Boguraev, B., Kennedy, C., 1997. Technical terminology for domain specification and content characterisation. In: Pazienza, M.T. (Ed.), Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, Lecture Notes in Computer Science, vol. 1299. Springer-Verlag, Berlin & Heidelberg, Germany, pp. 73–96.

Bond, C.E., Shipton, Z.K., Jones, R.R., Butler, R.W.H., Gibbs, A.D., 2007. Knowledge transfer in a digital world: field data acquisition, uncertainty, visualization, and data management. Geosphere 3 (6), 568–576.

Bonham-Carter, G.F., 1994. Geographic Information Systems for Geoscientists: Modeling with GIS. Pergamon, Elsevier Science Ltd., Kidlington, UK, 398 pp.

Booch, G., Maksimchuk, R., Engle, M., Young, B., Conallen, J., Houston, K. (Eds.), 2007. Object-Oriented Analysis and Design with Applications, 3rd Edition. Addison-Wesley Professional, Upper Saddle River, NJ, 720 pp.

Boolos, G.S., 1975. On Second-Order Logic. The Journal of Philosophy 72 (16), 509–527.

Borgo, S., Guarino, N., Vieu, L., 2005. Formal Ontology for Semanticists. Lecture notes of the 17th European Summer School in Logic, Language and Information (ESSLLI 2005), Edinburgh, Scotland, 12 pp., http://www.loa-cnr.it/Tutorials/ESSLLI1.pdf [Accessed February 15, 2008].

Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L., Stuckenschmidt, H., 2004. Contextualizing ontologies. Journal of Web Semantics: Science, Services and Agents on the World Wide Web 1 (4), 325–343.

Brodaric, B., 2004. The design of GSC FieldLog: ontology-based software for computer aided geological field mapping. Computers & Geosciences 30 (1), 5–20.

Brodaric, B., 2005. Representing Geo-pragmatics. PhD Dissertation, The Pennsylvania State University, University Park, PA, USA, 183 pp.

Brodaric, B., 2007. Geo-Pragmatics for the Geospatial Semantic Web. Transactions in GIS 11 (3), 453–477.

Brodaric, B., Boisvert, E., Patera, A., 2002. A set-theoretic technique and software for managing multiple-classification in geological hierarchies. In: Proceedings of the 2002 Annual Conference of International Association for Mathematical Geology, Berlin, Germany (on CD-ROM).

Brodaric, B., Gahegan, M., 2006. Representing geoscientific knowledge in cyberinfrastructure: challenges, approaches and implementations. In: Sinha, A.K. (Ed.), Geoinformatics: Data to Knowledge: Geological Society of America Special Paper 397. Geological Society of America, Boulder, CO, USA, pp. 1–20.

Brodaric, B., Probst, F., 2009. Enabling cross-disciplinary e-science by integrating geoscience ontologies with Dolce. IEEE Intelligent Systems 24 (1), 66–77.

Brodaric, B., Sharpe, D., Boisvert, E., 2009. Groundwater Information Network: enabling online access and analysis of Canadian groundwater information. Eos Transactions 90 (22), American Geophysical Union Joint Assembly 2009, Abstract No. IA2022A-2006.

Bronowski, J., Bellugi, U., 1970. Language, name, and concept. Science 168 (3932), 669–673.

Broome, J., 2005. The future of geoscience in the 21st century: art, science, or resource? In: Ostaficzuk, S.R. (Ed.), The Current Role of Geological Mapping in Geosciences. Springer Netherlands, Amsterdam, the Netherlands, pp. 51–58.

Broughton, V., 2006. The need for a faceted methods of information retrieval. Aslib Proceedings 58 (1-2), 49–72.

Buccella, A., Cechich, A., Fillottrani, P., 2009. Ontology-driven geographic information integration: a survey of current approaches. Computers & Geosciences 35 (4), 710–723.

Camon, E., Magrane, M., Barrell, D., Binns, D., Fleischmann, W., Kersey, P., Mulder, N., Oinn, T., Maslen, J., Cox, A., Apweiler, R., 2003. The Gene Ontology Annotation (GOA) Project: Implementation of GO in SWISS-PROT, TrEMBL, and InterPro. Genome Research 13 (4), 662–672.

Carranza, E.J.M., 2009. Geochemical Anomaly and Mineral Prospectivity Mapping in GIS. Elsevier, Amsterdam, 366 pp.

Carranza, E.J.M., Radwan, M., Hale, M., 2004. GDI for mineral resources development. GIS Development: The Global Geospatial Magazine 8 (9), 44–47.

CCOP and CIFEG (Eds.), 2006. Asian Multilingual Thesaurus of Geosciences. Coordinating Committee for Geoscience Programmes in East and Southeast Asia (CCOP) & Centre International pour la Formation et les Echanges en Géosciences (CIFEG), 563 pp. http://www.ccop.or.th/download/pub/AMTG_2006.pdf [Accessed February 07, 2011].

CGI-IUGS, 2008. GeoSciML Cookbook—How to Map Data to GeoSciML Version 2, Commission for the Management and Application of Geoscience Information, International Union of Geological Sciences (CGI-IUGS), 44 pp., http://www.onegeology.org/docs/technical/GeoSciML_Data_CookBook_V2.pdf [Accessed April 10, 2009].

CGMW, CPRM and DNPM, 2003. 1:5,500,000 Geological Map of South America. Commission for the Geological Map of the World (CGMW), Geological Survey of Brazil (CPRM), National Department of Mineral Production, Brazil (DNPM). CD-ROM.

Chen, A., Di, L., Bai, Y., Wei, Y., Liu, Y., 2010. Grid computing enhances standards-compatible geospatial catalogue service. Computers & Geosciences 36 (4), 411–421.

Chen, T., Guo, Y., Liu, Y.C., 2003. Chinese field matching algorithm. Computer Engineer 29 (13), 118–124. [In Chinese with English abstract]

Chen, Y.C. (Ed.), 2007. Mineralization System of China and Regional Mineralization Assessment. Geological Publishing House, Beijing, 1021 pp. [In Chinese]

Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geology Reviews 32 (1-2), 314–324.

Cimino, J.J., 1998. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of Information in Medicine 37 (4-5), 394–403.

Clegg, P., Bruciatelli, L., Domingos, F., Jones, R.R., De Donatis, M., Wilson, R.W., 2006. Digital geological mapping with table PC and PDA: a comparison. Computers & Geosciences 32 (10), 1682-1698.

Coleman, A.S., Bracke, P., 2006. Controlled vocabularies as a sphere of influence. In: Raghavan, K.S., Prasad, K.N. (Eds.), Knowledge Organization, Information Systems and Other Essays: Professor A. Neelameghan Festschrift. Ess Ess Publications, New Delhi, India, pp. 89–110.

Cox, S.J.D., Richard, S.M., 2005. A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards. Geosphere 1 (3), 119–137.

Crasnow, S.L., 2000. How natural can ontology be? Philosophy of Science 67 (1), 114–132.

Culshaw, M.G., Reeves, H.J., Jefferson, I., Spink, T.W. (Eds.), 2009. Engineering Geology for Tomorrow's Cities. The Geological Society, London, 315 pp.

D'Ambros, M., Lanza, M., Lungu, M., Robbes, R., 2010. On porting software visualization tools to the web. International Journal on Software Tools for Technology Transfer (STTT), doi: 10.1007/s10009-10010-10171-10009.

Dai, F.C., Lee, C.F., Zhang, X.H., 2001. GIS-based geo-environmental evaluation for urban land-use planning: a case study. Engineering Geology 61 (4), 257–271.

Davies, J., Fensel, D., van Harmelen, F., 2003. Towards the Semantic Web: Ontology-driven Knowledge Management. John Wiley & Sons, Chichester, England, 328 pp.

de Moor, A., 2005. Patterns for the Pragmatic Web. In: Dau, F., Mugnier, M.-L., Stumme, G. (Eds.), Proceedings of the 13th International Conference on Conceptual Structures (ICCS2005), Lecture Notes in Computer Science, vol. 3596. Springer-Verlag, Berlin & Heidelberg, pp. 1–18.

de Mulder, E.F.J., Nield, T., Derbyshire, E., 2006. The International Year of Planet Earth (2007-2009): earth sciences for society. Episodes 29, 82–86.

De Paor, D.G., Whitmeyer, S.J., 2011. Geological and geophysical modeling on virtual globes using KML, COLLADA, and Javascript. Computers & Geosciences 37 (1), 100–110.

Deissenboeck, F., Pizka, M., 2006. Concise and consistent naming. Software Quality Journal 14 (3), 261–282.

Delgado, L., Ortíz, G., Hernández, J.R., Cerdán, F.P., 2001. Building of trans-national geologic map databases--the Iberian Pyrite Belt case. In: Proceedings of 2001 Annual Conference of the International Association for Mathematical Geology, Cancún, Mexico, 10 pp., http://www.kgs.ku.edu/Conferences/IAMG/Sessions/I/Papers/Delgado/delgado.pdf [Accessed February 07, 2011].

Deliiska, B., 2007. Thesaurus and domain ontology of geoinformatics. Transactions in GIS 11 (4), 637–651.

Dicken, C.L., Nicholson, S.W., Horton, J.D., Kinney, S.A., Gunther, G., Foose, M.P., Mueller, J.A.L., 2008. Preliminary integrated geologic map databases for the United States: Delaware, Maryland, New York, Pennsylvania, and Virginia (Version 1.1), http://pubs.usgs.gov/of/2005/1325 [Accessed February 07, 2011].

Diering, J.A.C., 1992. ONE-D: a program for one-dimensional composite optimization. In: Annels, A.E. (Ed.), Case histories and methods in mineral resource evaluation; Geological Society, London, Special Publications, vol. 63, pp. 185–190.

Dimmer, E., Berardini, T.Z., Barrell, D., Camon, E., 2008. Methods for gene ontology annotation. In: Edwards, D. (Ed.), Plant Bioinformatics. Humana Press, pp. 495–520.

Din, J., Idris, S., 2009. Object-oriented design process model. International Journal of Computer Science and Network Security 9 (10), 71–79.

Ding, Y., Foo, S., 2002. Ontology research and development. part 2—a review of ontology mapping and evolving. Journal of Information Science 28 (5), 375–388.

Dominy, S.C., Noppé, M.A., Annels, A.E., 2002. Errors and uncertainty in mineral resource and ore reserve estimation: The importance of getting it right. Exploration and Mining Geology 11 (1-4), 77–98.

Dupplaw, D., Dasmahapatra, S., Hu, B., Lewis, P., Shadbolt, N., 2009. A distributed, service-based framework for knowledge applications with multimedia. ACM Transactions on Information Systems 27 (4), 1–29.

Durbha, S.S., King, R.L., Shah, V.P., Younan, N.H., 2009. A framework for semantic reconciliation of disparate earth observation thematic data. Computers & Geosciences 35 (4), 761–773.

Elst, P., Yard, T., Jacobs, S., Drol, W., 2006. Communication between Flash and the browser. In: Elst, P., Yard, T., Jacobs, S., Drol, W. (Eds.), Object-Oriented ActionScript for Flash 8. Apress, Berkeley, CA, USA, pp. 417–436.

Fang, Y.C., Huang, H.C., Chen, H.H., Juan, H.F., 2008. TCMGeneDIT: a database for associated traditional Chinese medicine, gene and disease information using text mining. BMC Complementary and Alternative Medicine 8. doi:10.1186/1472-6882-8-58.

FAO, 2010. AGROVOC—Multilingual Agricultural Thesaurus, Food and Agriculture Organization of the United Nations (FAO), 2 p., ftp://ftp.fao.org/gi/gil/gilws/aims/references/flyers/agrovoc_en.pdf [Accessed February 14, 2010].

Fils, D., Cervato, C., Reed, J., Diver, P., Tang, X., Bohling, G., Greer, D., 2009. CHRONOS architecture: experiences with an open-source services-oriented architecture for geoinformatics. Computers & Geosciences 35 (4), 774–782.

Fox, P., Hendler, J., 2011. Changing the equation on scientific data visualization. Science 331 (6018), 705–708.

Fox, P., McGuinness, D.L., Cinquini, L., West, P., Garcia, J., Benedict, J.L., Middleton, D., 2009. Ontology-supported scientific data frameworks: the Virtual Solar-Terrestrial Observatory experience. Computers & Geosciences 35 (4), 724–738.

Frank, A.U., 2001. Tiers of ontology and consistency constraints in geographical information systems. International Journal of Geographical Information Science 15 (7), 667–678.

Gahegan, M., Luo, J., Weaver, S.D., Pike, W., Banchuen, T., 2009. Connecting GEON: making sense of the myriad resources, researchers and concepts that comprise a geoscience cyberinfrastructure. Computers & Geosciences 35 (4), 836–854.

Galton, A., 2009. Spatial and temporal knowledge representation. Earth Science Informatics 2 (3), 169–187.

Garcia-Sanchez, F., Valencia-Garcia, R., Martinez-Bejar, R., Fernandez-Breis, J.T., 2009. An ontology, intelligent agent-based framework for the provision of semantic web services. Expert Systems with Applications 36 (2), 3167–3187.

Garrett, J.J., 2005. Ajax: a new approach to web applications. http://www.adaptivepath.com/ideas/essays/archives/000385.php [Accessed February 07, 2011].

Garruzzo, S., Rosaci, D., 2010. Ontology enrichment in multi agent systems through semantic negotiation. In: Meersman, R., Tari, Z. (Eds.), On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS, Lecture Notes in Computer Science, vol. 4803. Springer-Verlag, Berlin & Heidelberg, Germany, pp. 391–398.

Gebhardt, S., Wehrmann, T., Klinger, V., Schettler, I., Huth, J., Künzer, C., Dech, S., 2010. Improving data management and dissemination in web based information systems by semantic enrichment of descriptive data aspects. Computers & Geosciences 36 (10), 1362–1373.

Georgiadou, P.Y., 2006. SDI ontology and implications for research in the developing world: Editorial. International Journal of Spatial Data Infrastructures Research 1 (1), 51–64.

Gerhard, L.C., Harrison, W.E., Hanson, B.M. (Eds.), 2001. Geological Perspectives of Global Climate Change. American Association of Petroleum Geologists (AAPG), Tulsa, OK, USA, 372 pp.

Geroimenko, V., Geroimenko, L., 2006. Interactive interfaces for mapping E-Commerce ontologies. In: Geroimenko, V., Chen, C. (Eds.), Visualizing the Semantic Web. Springer London, pp. 199–209.

Gey, F.C., Kando, N., Peters, C., 2005. Cross-Language Information Retrieval: the way ahead. Information Processing & Management 41 (3), 415–431.

Gilchrist, A., 2003. Thesauri, taxonomies and ontologies—an etymological note. Journal of Documentation 59 (1), 7–18.

Gillespie, M.R., Styles, M.T., 1999. BGS Rock Classification Scheme, Volume1—Classification of Igneous Rocks, British Geological Survey Report, (2nd edition), RR 99-06. British Geological Survey, Keyworth, Nottingham, UK, 52 pp.

Gonçalves, B., Zamborlini, V., Guizzardi, G., Filho, J.G.P., 2009. An ontology-based application in heart electrophysiology: representation, reasoning and visualization on the web. In: Proceedings of the 2009 ACM symposium on Applied Computing, Honolulu, Hawaii, pp. 816–820.

Gravesteijn, J., Kortman, C., Potenza, R., Rassam, G.N. (Eds.), 1995. Multilingual Thesaurus of Geosciences, 2nd edition. Information Today, Inc., Medford, NJ, USA., 645 pp.

Gravesteijn, J., Rassam, G.N., 1990. International cooperation in geoscience information: a global challenge. Episodes 13 (3), 178–183.

Gray, J., Liu, D.T., Nieto-Santisteban, M., Szalay, A., DeWitt, D.J., Heber, G., 2005. Scientific data management in the coming decade. SIGMOD Rec. 34 (4), 34–41.

Green, D., Bossomaier, T., 2002. Online GIS and Spatial Metadata. Taylor & Francis, London, 222 pp.

Green, P., Rosemann, M., 2000. Integrated process modeling: An ontological evaluation. Information Systems 25 (2), 73–87.

Green, W.R., 1991. Exploration with a Computer: Geoscience Data Analysis Applications. Pergamon Press, Oxford, 225 pp.

Gruber, T.R., 1993. A translation approach to portable ontology specifications. Knowledge Acquisition 5 (2), 199–220.

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies 43 (5-6), 907–928.

Grudin, J., 2001. Desituating action: digital representation of context. Human-Computer Interaction 16 (2), 269–286.

Gruen, D., Rasmussen, J., Liu, J., Hupfer, S., Ross, S., 2008. Collaborative reasoning and collaborative ontology development in CRAFT. In: Proceedings of 2008 AAAI Spring Symposium, 8 pp., http://domino.research.ibm.com/cambridge/research.nsf/0/3c52b362957999d885257665006a1533/$FILE/SS07GruenD.pdf [Accessed February 28, 2011].

GSJ-AIST, 2009. Digital Geoscience Map Series G-16: Seamless Geological Map of Japan at a scale of 1:200,000. Geological Survey of Japan (GSJ) - National Institute of Advanced Industrial Science and Technology (AIST). CD-ROM.

Guarino, N., 1997a. Semantic matching: formal ontological distinctions for information organization, extraction, and integration. In: Pazienza, M.T. (Ed.), Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, Lecture Notes in Computer Science, vol. 1299. Springer-Verlag, Berlin & Heidelberg, Germany, pp. 139–170.

Guarino, N., 1997b. Understanding, building and using ontologies. International Journal of Human-Computer Studies 46 (2-3), 293–310.

Guarino, N., 1998. Formal ontology and information systems. In: Guarino, N. (Ed.), Formal Ontology in Information Systems. IOS Press, Amsterdam, the Netherlands, pp. 3–15.

Guha, R., McCool, R., 2003. TAP: a Semantic Web platform. Computer Networks 42 (5), 557–577.

Haase, P., Harmelen, F., Huang, Z., Stuckenschmidt, H., Sure, Y., 2005. A framework for handling inconsistency in changing ontologies. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (Eds.), The Semantic Web—ISWC 2005, Lecture Notes in Computer Science, vol. 3729. Springer-Verlag, Berlin & Heidelberg, Germany, pp. 353–367.

Haile, N.S., 1987. Time and age in geology: the use of upper/lower, late/early in stratigraphic nomenclature. Marine and Petroleum Geology 4 (3), 255–257.

Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., Riedemann, C., 1999. Semantic interoperability: a central issue for sharing geographic information. The Annals of Regional Science 33 (2), 213–232.

Haunert, J.-H., Wolff, A., 2010. Area aggregation in map generalisation by mixed-integer programming. International Journal of Geographical Information Science 24 (12), 1871–1897.

Helbig, H., 2006. Knowledge Representation and the Semantics of Natural Language. Springer-Verlag, Berlin & Heidelberg, Germany, 646 pp.

Hendler, J., 2003. Science and the Semantic Web. Science 299 (5606), 520–521.

Hepp, M., 2006. Products and services ontologies: A methodology for deriving OWL ontologies from industrial categorization standards. International Journal on Semantic Web and Information Systems 2 (1), 72–99.

Hey, T., Trefethen, A.E., 2005. Cyberinfrastructure for e-Science. Science 308 (5723), 817-821.

Hinman, P.G., 2005. Fundamentals of Mathematical Logic. A K Peters, Wellesley, MA, USA, 878 pp.

Holten, D., 2006. Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Transactions on Visualization and Computer Graphics 12 (5), 741–748.

Hong, D., Ahn, J., Yi, S., Park, H., Lee, N., Kim, D., 2010. Contents Database Structure based on 5W1H using Multi-Dimension and Agent. Information-an International Interdisciplinary Journal 14 (2), 297–312.

Hong, E.L., Balakrishnan, R., Dong, Q., Christie, K.R., Park, J., Binkley, G., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E., Hitz, B.C., Krieger, C.J., Livstone, M.S., Miyasato, S.R., Nash, R.S., Oughtred, R., Skrzypek, M.S., Weng, S., Wong, E.D., Zhu, K.K., Dolinski, K., Botstein, D., Cherry, J.M., 2008. Gene Ontology annotations at SGD: new data sources and annotation methods. Nucleic Acids Research 36 (suppl 1), D577–D581.

Howard, A.S., Hatton, B., Reitsma, F., Lawrie, K.I.G., 2009. Developing a geoscience knowledge framework for a national geological survey organisation. Computers & Geosciences 35 (4), 820–835.

Hubaux, A., 1973. A new geological tool – the data. Earth-Science Reviews, 9 (2), 159-196.

Huber, R., Klump, J., Gotz, S., 2003. A tree for rocks—hierarchies in stratigraphic databases. Computers & Geosciences 29 (7), 921–928.

Huff, W.D., 1998. Using internet resources in an introductory geology course. Computers & Geosciences 24 (7), 689-392.

Hustrulid, W., Kuchta, M., 2006. Open Pit Mine Planning and Design, 2nd Edition. Taylor & Francis, Leiden, The Netherlands, 836 pp.

Ipfelkofer, F., Lorenz, B., Ohlbach, H.J., 2006. Ontology driven visualisation of maps with SVG - an example for semantic programming. In: Banissi, E., Burkhard, R.A., Ursyn, A., Zhang, J.J., Bannatyne, M.W.M., Maple, C., Cowell, A.J., Tianm, G.Y., Hou, M. (Eds.), Information Visualization. IEEE Computer Society, Los Alamitos, pp. 424–429.

ISO, 2005. ISO/IEC 11179-5: Information technology—Metadata registries (MDR)—Part 5: Naming and identification principles. International Organization for Standardization (ISO), Geneva, 17 pp.

Jackson, I., 2007. OneGeology—Making geological map data for the earth accessible. Episodes 30 (1), 60–61.

Jackson, I., Wyborn, L., 2008. One planet: OneGeology? The Google Earth revolution and the geological data deficit. Environmental Geology 53 (6), 1377–1380.

Jaiswal, P., Avraham, S., Ilic, K., Kellogg, E.A., McCouch, S., Pujar, A., Reiser, L., Rhee, S.Y., Sachs, M.M., Schaeffer, M., Stein, L., Stevens, P., Vincent, L., Ware, D., Zapata, F., 2005. Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. Comparative and Functional Genomics 6 (7-8), 388–397.

Jupp, S., Stevens, R., Bechhofer, S., Yesilada, Y., Kostkova, P., 2008. Knowledge Representation for Web Navigation. In: Proceedings of Semantic Web Applications and Tools for the Life Sciences (SWAT4LS 2008), e-Science Institute, Edinburgh, Scotland, 15 pp., http://www.cs.man.ac.uk/~sjupp/publications/2008/swat4ls-Jupp.pdf [Accessed February 07, 2011].

Kapageridis, I.K., 2009. Current state of integrated software solutions for the mining industry. The Masterbuilder 11 (4), 78–82.

Kaplan, D., 1989. Demonstratives: an essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In: Almog, J., Perry, J., Wettstein, H. (Eds.), Themes from Kaplan. Oxford University Press, New York, pp. 481–564.

Kashyap, V., Sheth, A., 1996. Semantic and schematic similarities between database objects: a context-based approach. The Very Large Database Journal 5 (4), 276–304.

Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E., 2007. Ontology visualization methods—a survey. ACM Computing Surveys 39 (4), Article 10. doi: 10.1145/1287620.1287621.

Kavouras, M., Kokla, M., 2008. Theories of geographic concepts: ontological approaches to semantic integration. CRC Press, Taylor & Francis Group, Boca Raton, Florida, USA, 319 pp.

Kleppner, D., Sharp, P.A., 2009. Research data in the digital age. Science 325 (5939), 368.

Klien, E., 2007. A rule-based strategy for the semantic annotation of geodata. Transactions in GIS 11 (3), 437–452.

Koubarakis, M., Skiadopoulos, S., Tryfonopoulos, C., 2006. Logic and computational complexity for Boolean information retrieval. IEEE Transactions on Knowledge and Data Engineering 18 (12), 1659–1666.

Kraak, M.-J., 2004. The role of the map in a Web-GIS environment. Journal of Geographical Systems 6 (2), 83–93.

Kraak, M.-J., 2008. Exploratory visualization. In: Shekhar, S., Xiong, H. (Eds.), Encyclopedia of GIS. Springer, Berlin, pp. 301–307.

Krivov, S., Williams, R., Villa, F., 2007. GrOWL: A tool for visualization and editing of OWL ontologies. Web Semantics: Science, Services and Agents on the World Wide Web 5 (2), 54–57.

Kuhn, W., 2009. Semantic engineering. In: Navratil, G. (Ed.), Research Trends in Geographic Information Science, Lecture Notes in Geoinformation and Cartography. Springer, Berlin & Heidelberg, pp. 63–76.

Kuhn, W., 2010. Modeling vs encoding for the Semantic Web. Semantic Web 1 (1), 11–15.

Kukich, K., 1992. Techniques for automatically correcting words in text. Computing Surveys 24 (4), 377–439.

Kulawiak, M., Prospathopoulos, A., Perivoliotis, L., Łuba, M., Kioroglou, S., Stepnowski, A., 2010. Interactive visualization of marine pollution monitoring and forecasting data via a Web-based GIS. Computers & Geosciences 36 (8), 1069–1080.

Kulik, L., Duckham, M., Egenhofer, M., 2005. Ontology-driven map generalization. Journal of Visual Languages and Computing 16 (3), 245–267.

Lacasta, J., Nogueras-Iso, J., Béjar, R., Muro-Medrano, P.R., Zarazaga-Soria, F.J., 2007. A web ontology service to facilitate interoperability within a spatial data infrastructure: applicability to discovery. Data & Knowledge Engineering 63 (3), 947–971.

Lam-Adesina, A.M., Jones, G.J.F., 2006. Examining and improving the effectiveness of relevance feedback for retrieval of scanned text documents. Information Processing & Management 42 (3), 633–649.

Laxton, J., Serrano, J.-J., Tellez-Arenas, A., 2010. Geological applications using geospatial standards – an example from OneGeology-Europe and GeoSciML. International Journal of Digital Earth 3 (supp. 1), 31–49.

Lazarinis, F., Vilares, J., Tait, J., Efthimiadis, E., 2009. Current research issues and trends in non-English Web searching. Information Retrieval 12 (3), 230–250.

Lee, J., Madnick, S., Siegel, M., 1996. Conceptualizing semantic interoperability: a perspective from the knowledge level. International Journal of Cooperative Information Systems 5 (4), 367–393.

Lehmann, S., Schwanecke, U., Dörner, R., 2010. Interactive visualization for opportunistic exploration of large document collections. Information Systems 35 (2), 260–269.

Levinson, S.C., 1983. Pragmatics. Cambridge University Press, Cambridge, 434 pp.

Liebig, T., Noppens, O., 2005. OntoTrack: a semantic approach for ontology authoring. Web Semantics: Science, Services and Agents on the World Wide Web 3 (2-3), 116–131.

Linnarsson, R., 1989. The data dictionary—a controlled vocabulary for integrating clinical databases and medical knowledge bases. Methods of Information in Medicine 28 (2), 78–85.

Liu, G., Yuan, Y., Wu, C., 1999. Application of parametric design method in geological and mineral map computer-aided mapping. Geological Science and Technology Information 18 (1), 93–96. [In Chinese with English abstract]

Lloyd, J.J., 1973. Concept relations in a multilingual thesaurus. Geology 1 (4), 152.

Losee, R.M., 1997. Comparing Boolean and probabilistic information retrieval systems across queries and disciplines. Journal of the American Society for Information Science 48 (2), 143–156.

Loudon, T.V., 2000. Geoscience after IT: A View of the Present and Future Impact of Information Technology on Geoscience. Elsevier, Oxford, 142 pp.

Loudon, T.V., 2009. Four interacting aspects of a geological survey knowledge system. Computers & Geosciences 35 (4), 700–705.

Loudon, T.V., Laxton, J.L., 2007. Steps toward Grid-based geological survey: suggestions for a systems framework of models, ontologies, and workflows. Geosphere 3 (5), 319–336.

Lu, D., Jin, Y., 1997. Forward for Windows—An object-oriented well logging interpretation platform. Well Logging Technology 21 (1), 55–60. [In Chinese with English abstract]

Ludäscher, B., Lin, K., Bowers, S., Jaeger-Frank, E., Brodaric, B., Baru, C., 2006. Managing scientific data: from data integration to scientific workflows. In: Sinha, A.K. (Ed.), Geoinformatics: Data to Knowledge; Geological Society of America Special Paper 397. Geological Society of America, Boulder, Colorado, USA, pp. 109–130.

Ludäscher, B., Lin, K., Brodaric, B., Baru, C., 2003. GEON: toward a cyberinfrastructure for the geosciences—a prototype for geological map interoperability via domain ontologies. In: Soller, D.R. (Ed.), Digital Mapping Techniques '03—Workshop Proceedings, Millersville, PA, USA, pp. 223–229.

Lumb, L.I., Freemantle, J.R., Lederman, J.I., Aldridge, K.D., 2009. Annotation modeling with formal ontologies: implications for informal ontologies. Computers & Geosciences 35 (4), 855–861.

Lüscher, P., Weibel, R., Burghardt, D., 2009. Integrating ontological modelling and Bayesian inference for pattern classification in topographic vector data. Computers, Environment and Urban Systems 33 (5), 363–374.

Lutz, M., Sprado, J., Klien, E., Schubert, C., Christ, I., 2009. Overcoming semantic heterogeneity in spatial data infrastructures. Computers & Geosciences 35 (4), 739–752.

Ma, X., Carranza, E.J.M., van der Meer, F.D., Wu, C., Zhang, X., 2010a. Algorithms for multi-parameter constrained compositing of borehole assay intervals from economic aspects. Computers & Geosciences 36 (7), 945–952.

Ma, X., Carranza, E.J.M., Wu, C., van der Meer, F.D., Liu, G., A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps. Computers & Geosciences (2011), doi: 10.1016/j.cageo.2011.02.011.

Ma, X., Wang, X., Wu, C., Ju, F., 2007. Metadata hierarchy in integrated geoscientific database for regional mineral prospecting. Geo-Spatial Information Science 10 (3), 223–227.

Ma, X., Wu, C., Carranza, E.J.M., Schetselaar, E.M., van der Meer, F.D., Liu, G., Wang, X., Zhang, X., 2010b. Development of a controlled vocabulary for semantic interoperability of mineral exploration geodata for mining projects. Computers & Geosciences 36 (12), 1512–1522.

Ma, X., Wu, C., Wang, K., Zhang, X., 2007. Point-source database theory and its application in database architecture design of Digital Mine. In: Proceedings of the 2007 Annual Conference of International Association for Mathematical Geosciences: Geomathematics and GIS Analysis of Resources, Environment, and Hazards, Beijing, China, pp. 421–424.

Ma, X., Wu, C., Wang, X., Chen, Y., Wang, Y., 2005. Centralized management approach and database development of multisource geoscientific information. In: Proceedings of IAMG'05: GIS and Spatial Analysis, Toronto, Canada, pp. 1006–1011.

Manso, M.-Á., Wachowicz, M., Bernabé, M.-Á., 2009. Towards an integrated model of interoperability for spatial data infrastructures. Transactions in GIS 13 (1), 43–67.

Mascarelli, A.L., 2009. Quaternary geologists win timescale vote. Nature 459, 624–624.

McGuinness, D.L., 2003. Ontologies come of age. In: Fensel, D., Hendler, J., Lieberman, H., Wahlster, W. (Eds.), Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, Cambridge, MA, USA, pp. 171–196.

McGuinness, D.L., Fox, P., Brodaric, B., Kendall, E., 2009. The emerging field of semantic scientific knowledge integration. IEEE Intelligent Systems 24 (1), 25–26.

McGuinness, D.L., Fox, P., Sinha, A.K., Raskin, R., 2007. Semantic integration of heterogeneous volcanic and atmospheric data. In: Brady, S.R., Sinha, A.K., Gundersen, L.C. (Eds.), Proceedings of Geoinformatics 2007—Data to Knowledge, Reston, VA, USA, pp. 10–13.

McGuinness, D.L., Sinha, A.K., Fox, P., Raskin, R., Heiken, G., Barnes, C., Wohletz, K., Venezky, D., Lin, K., 2006. Towards a reference volcano ontology for semantic scientific data integration. In: Proceedings of the 2006 American Geophysical Union Joint Assembly, Baltimore, MD, USA, Abstract No. IN42A-03.

Mey, J.L., 2003. Context and (dis)ambiguity: a pragmatic view. Journal of Pragmatics 35 (3), 331–347.

Michael, A.J., Eberhart-Phillips, D., 1991. Relations among fault behavior, subsurface geology, and three-dimensional velocity models. Science 253 (5020), 651–654.

Michalak, J., 2005. Topological conceptual model of geological relative time scale for geoinformation systems. Computers & Geosciences 31 (7), 865–876.

Miles, A., Matthews, B., Wilson, M., 2001. RDF Encoding of Multilingual Thesauri, http://www.w3c.rl.ac.uk/SWAD/deliverables/8.3.html [Accessed February 07, 2011].

Miles, A., Pérez-Agüera, J.R., 2007. SKOS: simple knowledge organisation for the web. Cataloging & Classification Quarterly 43 (3), 69–83.

Min, C., Ebert, D., Hagen, H., Laramee, R.S., van Liere, R., Ma, K.L., Ribarsky, W., Scheuermann, G., Silver, D., 2009. Data, information, and knowledge in visualization. IEEE Computer Graphics and Applications 29 (1), 12–19.

Mori, A.R., 1995. Coding systems and controlled vocabularies for hospital information systems. International Journal of Bio-Medical Computing 39 (1), 93–98.

Morris, C.W., 1938. Foundations of the Theory of Signs. In: Neurath, O. (Ed.), International Encyclopedia of Unified Science, Vol. 1, No. 2. University of Chicago Press, Chicago, pp. 13–71.

Mylopoulos, J., Chung, L., Yu, E., 1999. From object-oriented to goal-oriented requirements analysis. Communications of the ACM 42 (1), 31–37.

NADM Steering Committee, 2004. NADM Conceptual Model 1.0—A conceptual model for geologic map information: U.S. Geological Survey Open-File Report 2004-1334, North American Geologic Map Data Model (NADM) Steering Committee, Reston, Virginia, 58 pp.

Nambiar, U., Ludäescher, B., Lin, K., Baru, C., 2006. The GEON portal: accelerating knowledge discovery in the geosciences. In: Proceedings of the 8th ACM International Workshop on Web Information and Data Management, Arlington, Virginia, USA, pp. 83–90.

Neuendorf, K.K.E., Mehl, J.J.P., Jackson, J.A. (Eds.), 2005. Glossary of Geology, 5th edition. American Geological Institute, Alexandria, VA, USA, 800 pp.

Neun, M., Burghardt, D., Weibel, R., 2008. Web service approaches for providing enriched data structures to generalisation operators. International Journal of Geographical Information Science 22 (2), 133–165.

Noy, N.F., 2009. Ontology mapping. In: Staab, S., Studer, R. (Eds.), Handbook on Ontologies. Springer, Berlin & Heidelberg, pp. 573–590.

Noy, N.F., Klein, M., 2004. Ontology evolution: not the same as schema evolution. Knowledge and Information Systems 6 (4), 428–440.

Obrst, L., 2003. Ontologies for semantically interoperable systems. In: Proceedings of the Twelfth International Conference on Information and Knowledge Management, New Orleans, LA, USA, pp. 366–369.

OGC, 2007a. OGC 05-078r4: Styled Layer Descriptor Profile of the Web Map Service Implementation Specification. Open Geospatial Consortium (OGC), 45 pp.

OGC, 2007b. OGC 07-006r1: OpenGIS® Catalogue Services Specification. Open Geospatial Consortium (OGC), 204 pp.

OGC/ISO, 2010. OGC 09-026r1 and ISO 19143:2010(E): Geographic information—Filter encoding. Open Geospatial Consortium (OGC) and International Organization for Standardization (ISO), 82 pp.

Ogg, J.G., 2009. International Stratigraphic Chart, International Commission on Stratigraphy, 1 p., http://www.stratigraphy.org/upload/ISChart2009.pdf [Accessed February 07, 2011].

OneGeology-Europe Consortium, 2010. One Europe One Geology: Applying Geoscience for Society. OneGeology-Europe Consortium, ISBN 978-0-85272-668-6, 84 pp.

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263 (5147), 641–646.

Ouksel, A.M., Sheth, A., 1999. Semantic interoperability in global information systems. SIGMOD Record 28 (1), 5–12.

Pan, J.Z., Horrocks, I., 2007. RDFS(FA): connecting RDF(S) and OWL DL. IEEE Transactions on Knowledge and Data Engineering 19 (2), 192–206.

Pastor-Sanchez, J.A., Mendez, F.J.M., Rodriguez-Munoz, J.V., 2009. Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. Information Research-an International Electronic Journal 14 (4), paper 422, http://informationr.net/ir/14-4/paper422.html [Accessed February 07, 2011].

Peng, Z.-R., Tsou, M.H., 2003. Internet GIS: Distributed Geographic Information services for the Internet and Wireless Networks. John Wiley & Sons, Inc., Hoboken, NJ, USA, 679 pp.

Pipkin, B.W., Trent, D.D., Hazlett, R., Bierman, P. (Eds.), 2008. Geology and the Environment, 5th Edition. Thomson Brooks/Cole, Belmont, CA, USA, 505 pp.

Podemski, M., 2005. The Polish Geological Institute activities in the cross-border geological mapping. Przeglad Geologiczny 53 (10/2), 894–898.

Radecki, T., 1983. Generalized boolean methods of information-retrieval. International Journal of Man-Machine Studies 18 (5), 407–439.

Ram, S., Park, J., 2004. Semantic conflict resolution ontology (SCROL): an ontology for detecting and resolving data and schema-level semantic conflicts. IEEE Transactions on Knowledge and Data Engineering 16 (2), 189–202.

Ranganathan, A., Campbell, R.H., 2003. An infrastructure for context-awareness based on first order logic. Personal and Ubiquitous Computing 7 (6), 353–364.

Ranta, D.E., Ward, A.D., Ganster, M.W., 1984. Ore zoning applied to geological reserve estimation of molybdenum deposits. In: Erickson, A.J. (Ed.), Applied Mining Geology. American Institute of Mining, Metallurgical, and Petroleum Engineers, New York, pp. 83–114.

Raskin, R.G., Pan, M.J., 2005. Knowledge representation in the semantic web for earth and environmental terminology (SWEET). Computers & Geosciences 31 (9), 1119–1125.

Rassam, G.N., Gravesteijn, J., Potenza, R. (Eds.), 1988. Multilingual Thesaurus of Geosciences. Pergamon Press, New York, USA, 516 pp.

Rector, A., Drummond, N., Horridge, M., Rogers, J., Knublauch, H., Stevens, R., Wang, H., Wroe, C., 2004. OWL pizzas: practical experience of teaching OWL-DL: common errors and common patterns. In: Motta, E., Shadbolt, N., Stutt, A., Gibbins, N. (Eds.), Engineering Knowledge in the Age of the Semantic Web, Lecture Notes in Computer Science, vol. 3257. Springer-Verlag, Berlin & Heidelberg, pp. 63–81.

Reitsma, F., 2010. Geoscience explanations: identifying what is needed for generating scientific narratives from data models. Environmental Modelling & Software 25 (1), 93–99.

Reitsma, F., Laxton, J., Ballard, S., Kuhn, W., Abdelmoty, A., 2009. Semantics, ontologies and eScience for the geosciences. Computers & Geosciences 35 (4), 706–709.

Rhee, S.Y., Wood, V., Dolinski, K., Draghici, S., 2008. Use and misuse of the gene ontology annotations. Nature Reviews Genetics 9 (7), 509–515.

Richard, S.M., 2006. Geoscience concept models. In: Sinha, A.K. (Ed.), Geoinformatics: Data to Knowledge: Geological Society of America Special Paper 397. Geological Society of America, Boulder, CO, USA, pp. 81–107.

Richard, S.M., Matti, J., Soller, D.R., 2003. Geoscience terminology development for the national geologic map database. In: Soller, D.R. (Ed.), Digital Mapping Techniques '03—Workshop Proceedings, Millersville, PA, USA, pp. 157–167.

Richard, S.M., Soller, D.R., 2008. Vocabularies for geoscience information interchange. In: Soller, D.R. (Ed.), Digital Mapping Techniques '08—Workshop Proceedings, Moscow, ID, USA, pp. 101–104.

Rock, N.M.S., 1991. Progress in 1988–1990 with computer applications in the “hard-rock” arena: geochemistry, mineralogy, petrology, and Volcanology. Computers & Geosciences 17 (8), 1067–1090.

Rossiter, D., 2007. Classification of urban and industrial soils in the World Reference Base for Soil Resources. Journal of Soils and Sediments 7 (2), 96–100.

Roy, P.S., Williams, R.J., Jones, A.R., Yassini, I., Gibbs, P.J., Coates, B., West, R.J., Scanes, P.R., Hudson, J.P., Nichol, S., 2001. Structure and function of south-east Australian estuaries. Estuarine, Coastal and Shelf Science 53 (3), 351–384.

Satkunas, J., Graniczny, M., 1997. Lithuanian Polish cross-border geoenvironmental mapping and its relevancy to spatial development. In: Marinos, P.G., Koukis, G.C., Tsiambaos, G.C., Stournaras, G.C. (Eds.), Engineering Geology and the Environment. A.A. Balkema, Rotterdam, Netherlands, pp. 1483–1486.

Satkunas, J., Graniczny, M., Lazauskiene, J., 2004. Geoscientific data, sustainable development and international borders. In: Proceedings of Thematic Session on “Sustainable Development of Geo-resources and Geo-environment”, 42nd Coordinating Committee for Geoscience Programmes in East and Southeast Asia (CCOP) Annual Session, Beijing, China, pp. 35–41.

Schäffer, B., Baranski, B., Foerster, T., 2010. Towards Spatial Data Infrastructures in the Clouds. In: Painho, M., Santos, M.Y., Pundt, H. (Eds.), Geospatial Thinking. Springer-Verlag, Berlin & Heidelberg, pp. 399–418.

Schofield, P.N., Bubela, T., Weaver, T., Portilla, L., Brown, S.D., Hancock, J.M., Einhorn, D., Tocchini-Valentini, G., Hrabe de Angelis, M., Rosenthal, N., 2009. Post-publication sharing of data and tools. Nature 461 (7261), 171–173.

Schokker, J., 2010. 1:600,000 Geologische overzichtskaart van Nederland. TNO Bouw en Ondergrond (Geological Survey of the Netherlands). CD-ROM.

Schoop, M., de Moor, A., Dietz, J.L.G., 2006. The pragmatic web: a manifesto. Communications of the ACM 49 (5), 75–76.

Sen, M., Duffy, T., 2005. GeoSciML: development of a generic GeoScience Markup Language. Computers & Geosciences 31 (9), 1095–1103.

Sharpe, W.E., Leibfried, V.G., Kimmel, W.G., Dewalle, D.R., 1987. The relationship of water-quality and fish occurrence to soils and geology in an area of high hydrogen and sulfate ion deposition. Water Resources Bulletin 23 (1), 37–46.

Sheth, A.P., 1999. Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In: Goodchild, M., Egenhofer, M., Fegeas, R., Kottman, C. (Eds.), Interoperating Geographic Information Systems. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 5–29.

Shimazu, K., Arisawa, T., Saito, I., 2006. Interdisciplinary contents management using 5W1H interface for metadata. In: Proceedings of 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Los Alamitos, pp. 909–912.

Sinclair, A.J., Blackwell, G.H., 2002. Applied Mineral Inventory Estimation, Cambridge, UK, 381 pp.

Singh, M.P., 2002. The Pragmatic Web. IEEE Internet Computing 6 (3), 4–5.

Sinha, A.K., Malik, Z., Rezgui, A., Fox, D.L., Barnes, C.G., Lin, K., Heiken, G., Thomas, W.A., Gundersen, L.C., Raskin, R., Jackson, I., Fox, P., McGuinness, D., Seber, D., Zimmerman, H., 2010. Geoinformatics: transforming data to knowledge for geosciences. GSA Today 20 (12), 4–10.

Sinha, A.K., McGuinness, D.L., Fox, P., Raskin, R., Condie, K., Stern, R., Hanan, B., Seber, D., 2007. Towards a reference plate tectonics and volcano ontology for semantic scientific data integration. In: Brady, S.R., Sinha, A.K., Gundersen, L.C. (Eds.), Proceedings of Geoinformatics 2007—Data to Knowledge, Reston, VA, USA, pp. 43–46.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S., 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology 25 (11), 1251–1255.

Smith, B., Kumar, A., 2004. Controlled vocabularies in bioinformatics: a case study in the gene ontology. Drug Discovery Today: BIOSILICO 2 (6), 246–252.

Smits, P.C., Friis-Christensen, A., 2007. Resource Discovery in a European Spatial Data Infrastructure. IEEE Transactions on Knowledge and Data Engineering 19 (1), 85–95.

Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., Katz, S., 2004. Reengineering thesauri for new applications: the AGROVOC example. Journal of Digital Information 4 (4), Online Journal, http://journals.tdl.org/jodi/rt/printerFriendly/112/111 [Accessed February 07, 2011].

Soller, D., Berg, T., 2005. The U.S. National Geologic Map Database project: overview & progress. In: Ostaficzuk, S.R. (Ed.), The Current Role of Geological Mapping in Geosciences. Springer, Dordrecht, The Netherlands, pp. 245–277.

Stafford, N., 2010. Science in the digital age. Nature 467 (7317), S19–S21.

Stalnaker, R., 1998. On the representation of context. Journal of Logic, Language and Information 7 (1), 3–19.

Sugumaran, V., Storey, V.C., 2006. The role of domain ontologies in database design: An ontology management and conceptual modeling environment. ACM Transactions on Database Systems 31 (3), 1064–1094.

Suh, B., Bederson, B.B., 2002. OZONE: a zoomable interface for navigating ontology. In: Proceedings of International Conference on Advanced Visual Interfaces (AVI 2002), Trento, Italy, pp. 139–143.

Sun, X.P., 2010. OSLN: An Object-Oriented Semantic Link Network language for complex object description and operation. Future Generation Computer Systems-the International Journal of Grid Computing-Theory Methods and Applications 26 (3), 389–399.

Tennis, J.T., Sutton, S.A., 2008. Extending the simple knowledge organization system for concept management in vocabulary development applications. Journal of the American Society for Information Science and Technology 59 (1), 25–37.

Tilmes, C., Yesha, Y., Halem, M., 2010. Tracking provenance of earth science data. Earth Science Informatics 3 (1-2), 59–65.

Tolk, A., 2003. Beyond technical interoperability—introducing a reference model for measures of merit for coalition interoperability. In: Proceedings of the 8th International Command and Control Research and Technology Symposium (ICCRTS), Washington, DC, USA, 22 pp. [On CD-ROM].

Tolk, A., Turnitsa, C.D., Diallo, S.Y., Winters, L.S., 2006. Composable M&S web services for net-centric applications. Journal of Defense Modeling and Simulation: Applications, Methodology, Technology 3 (1), 27–44.

Tripathi, A., Babaie, H.A., 2008. Developing a modular hydrogeology ontology by extending the SWEET upper-level ontologies. Computers & Geosciences 34 (9), 1022–1033.

Tsuruoka, Y., Tsujii, J., Ananiadou, S., 2008. FACTA: a text search engine for finding associated biomedical concepts. Bioinformatics 24 (21), 2559–2560.

Tutmez, B., 2007. An uncertainty oriented fuzzy methodology for grade estimation. Computers & Geosciences 33 (2), 280–288.

UNECE, 1997. United Nations International Framework Classification for Reserves/Resources: Solid Fuels and Mineral Commodities. United Nations Economic Commission for Europe (UNECE). UN Document Symbol: [E/ECE/]ENERGY/WP.1/R.70. Original: English, Geneva, 23 pp.

UNECE, 2001. Guidelines to the United Nations International Framework Classification for Reserves/Resources: Solid Fuels and Mineral Commodities. United Nations Economic Commission for Europe (UNECE). UN Document Symbol: [E/ECE/]ENERGY/2001/11. Original: English, Geneva, 37 pp.

UNSPSC, 2004. United Nations Standard Products and Services Code (UNSPSC)—Chinese Series 6, 438 pp., http://www.unspsc.org/Codeset_Documents/UNSPSC_Chinese.pdf [Accessed February 15, 2009].

U.S. Geological Survey Geologic Names Committee, 2010. Divisions of geologic time-major chronostratigraphic and geochronologic units: U.S. Geological Survey Fact Sheet 2010-3059, 2 pp., http://pubs.usgs.gov/fs/2010/3059/pdf/FS10-3059.pdf [Accessed Febru-ary 07, 2011].

Uschold, M., Gruninger, M., 2004. Ontologies and semantics for seamless connectivity. SIGMOD Record 33 (4), 58-64.

van Assem, M., Malaisé, V., Miles, A., Schreiber, G., 2006. A method to convert thesauri to SKOS. In: Sure, Y., Domingue, J. (Eds.), The Semantic Web: Research and Applications, Lecture Notes in Computer Science, vol. 4011. Springer-Verlag, Berlin & Heidelberg, Germany, pp. 95–109.

van Heijst, G., Schreiber, A.T., Wielinga, B.J., 1997. Using explicit ontologies in KBS development. International Journal of Human-Computer Studies 46 (2-3), 183–292.

Verheyden, P., De Bo, J., Meersman, R., 2005. Semantically unlocking database content through ontology-based mediation. In: Bussler, C., Tannen, V., Fundulaki, I. (Eds.), Semantic Web and Databases, Lecture Notes in Computer Science, vol. 3372. Springer-Verlag, Berlin & Heidelberg, Germany, pp. 109–126.

Villa, F., Athanasiadis, I.N., Rizzoli, A.E., 2009. Modelling with knowledge: A review of emerging semantic approaches to environmental modelling. Environmental Modelling & Software 24 (5), 577–587.

Villaescusa, E., 1998. Geotechnical design for dilution control in underground mining. In: Singhal, R.K. (Ed.), Mine Planning and Equipment Selection 1998. Balkema, Rotterdam, pp. 141–149.

Visser, U., Stuckenschmidt, H., Schuster, G., Vögele, T., 2002. Ontologies for geographic information processing. Computers & Geosciences 28 (1), 103–117.

Voinov, A., Bousquet, F., 2010. Modelling with stakeholders. Environmental Modelling & Software 25 (11), 1268–1281.

Voss, J., 2009. Encoding changing country codes for the Semantic Web with ISO 3166 and SKOS. In: Sicilia, M.-A., Lytras, M.D. (Eds.), Metadata and Semantics. Springer-US, New York, pp. 211–221.

Walker, J.D., Geissman, J.W., (compilers), 2009. Geologic Time Scale: Geological Society of America, 1 p., http://www.geosociety.org/science/timescale/timescl.pdf [Accessed February 07, 2011].

Wang, S.C. (Ed.), 1990. Assessment of Gold Mineral Resources with Integrated Information Jilin Sci. & Tec. Press, 441 pp. [In Chinese]

Wang, X., Liu, G., Yuan, Y., Han, Z., 1999. Application of "the terminology classification codes of geology and mineral resources" on geological and mineral resources point-source information system. Earth Science—Journal of China University of Geosciences 24 (5), 529–532. [In Chinese with English abstract]

Ware, J.M., Wilson, I.D., Ware, J.A., 2003. A knowledge based genetic algorithm approach to automating cartographic generalisation. Knowledge-Based Systems 16 (5-6), 295–303.

Welty, C., 2002. Ontology-driven conceptual modeling. In: Pidduck, A.B., Mylopoulos, J., Woo, C.C., Ozsu, M.T. (Eds.), Advanced Information Systems Engineering, Lecture Notes in Computer Science, vol. 2348. Springer-Verlag, Berlin & Heidelberg, pp. 3–3, Presentation notes: http://www.cs.toronto.edu/caise02/cwelty.pdf [Accessed February 15, 2008].

Willems, N., van Hage, W.R., de Vries, G., Janssens, J.H.M., Malaisé, V., 2010. An integrated approach for visual analysis of a multisource moving objects knowledge base. International Journal of Geographical Information Science 24 (10), 1543–1558.

Woodcock, R., Simons, B., Duclaux, G., Cox, S., 2010. AuScope's use of standards to deliver earth resource data. Geophysical Research Abstracts 12, European Geosciences Union General Assembly 2010, Abstract No. EGU2010-1556.

Wu, C., Liu, G., Tian, Y., Liu, J., 2005. Theory and method problems of informatization of geological surveying and mineral resource exploration. Earth Science - Journal of China University of Geosciences 30 (3), 365–369. [In Chinese with English abstract]

Xu, C., Dowd, P.A., 2003. Optimal construction and visualisation of geological structures. Computers & Geosciences 29 (6), 761–773.

Ye, T.Z. (Ed.), 2004. Methods and Techniques for Solid Minerals Predication and Assessment. China Land Press, Beijing, 351 pp. [In Chinese]

Youn, C., Kaiser, T., Seber, D., Santini, C., 2008. Web-based simulating system for modeling earthquake seismic wavefields on the grid. Computers & Geosciences 34 (12), 1936–1946.

Yue, P., Di, L., Yang, W., Yu, G., Zhao, P., Gong, J., 2009. Semantic Web Services-based process planning for earth science applications. International Journal of Geographical Information Science 23 (9), 1139–1163.

Zadeh, L.A., 1965. Fuzzy sets. Information and Control 8 (3), 338–353.

Zadeh, L.A., 2008. Is there a need for fuzzy logic? Information Sciences 178 (13), 2751–2779.

Zhai, Y.S. (Ed.), 2003. Methods for Regional Mineralization Research. China Land Press, Beijing, 183 pp. [In Chinese]

Zhang, C.R., Zhao, T., Li, W.D., 2010. The framework of a geospatial semantic web-based spatial decision support system for Digital Earth. International Journal of Digital Earth 3 (2), 111–134.

Zhang, Q., 2002. Reasons and solutions to garbled characters in databases. Net Administrator World (9), 34–35 [In Chinese].

Zhao, P., Di, L., Yu, G., Yue, P., Wei, Y., Yang, W., 2009. Semantic Web-based geospatial knowledge transformation. Computers & Geosciences 35 (4), 798–808.

Zhao, P.D. (Ed.), 2006. Theories and Methods for Mineral Exploration. China University of Geosciences Press, Wuhan, China, 334 pp. [In Chinese]

Zillner, S., Hauer, T., Rogulin, D., Tsymbal, A., Huber, M., Solomonides, T., 2008. Semantic visualization of patient information. In: Proceedings of 21st IEEE International Symposium on Computer-Based Medical Systems (CBMS '08), Jyväskylä, Finland, pp. 296–301.

Appendix: Programs and documents of pilot systems

Pilot systems were developed for chapters 3, 4 and 5 of this dissertation. Programs, source codes, documents and videos related to these pilot systems were put online and can be accessed through a webpage.

https://sites.google.com/site/xgmaitc/phdwork

Summary

Geological data are not only essential for studying the earth but also for addressing key societal challenges. Interoperability of geological data has long been a topic of concern in scientific works. Compared to the ongoing deluge of digital geological data, approaches for promoting effective geological data interoperability are underdeveloped.

Challenges regarding data interoperability can arise at different levels, such as systems (i.e., network and services), syntax (i.e., language and encoding), schemas (i.e., modeling and structure), semantics (i.e., content and meaning), and pragmatics (i.e., use and effect). Among the various finished and/or ongoing studies addressing geological data interoperability, ontology-based approaches have attracted increasing attentions in recent years. Ontologies in computer science are defined as shared conceptualizations of domain knowledge. It was discussed in the literature that, among various types of ontologies, there is an ontology spectrum, which covers ontology types with varying semantic richness.

Despite the progress in building and using different types of geological ontologies, the application of an ontology spectrum to promote geological data interoperability still faces various challenges, among which four key challenges are addressed in this dissertation: (1) modeling and encoding of ontologies; (2) multilinguality of geological data and ontologies; (3) flexibility and usefulness of ontology-based applications; and (4) mediation and evolution of geological data and ontologies.

The research described and discussed in this dissertation aimed to explore approaches to address the aforementioned key challenges, and thus to provide a route map for applying an ontology spectrum to promote geological data interoperability at local, regional and global levels. The dissertation was focused on the following five specific research questions.

(1) How can ontologies be modeled and encoded, so that the resulting ontologies are not only efficient for harmonizing local geological data but also function to improve the interoperability of local or internal geological data with extramural or external projects?

(2) In a regional/global environment, how can linguistic barriers of online geological data be alleviated by building and using multilingual ontologies?

(3) How can different methods of conceptual analysis be integrated to develop thematic conceptual schemas that are efficient for problem-

solving and are compatible with commonly used standards in the field of geology?

(4) How can ontology-based tools be developed to improve the interoperability of online geological data, so as to help both geologists and non-geologists to retrieve geological information and discover geological knowledge?

(5) What are the context-caused challenges for geological data interoperability, and how can these challenges be addressed in a long-term perspective?

To give insights into above-stated questions, case studies of geological data interoperability at local, regional and global levels were conducted in this research. Several types of ontologies such as taxonomies, thesauri, conceptual schemas and logical language-based ontologies were examined, developed and deployed, according to the context of each case study. Based on the results of these case studies, the dissertation answered each research question and presented findings addressing the aforementioned four key challenges: (1) Modeling and encoding should be distinguished while also bridged in building and using of geological ontologies; (2) Multilingual labeling is effective for build accurate, consistent and complete multilingual geological ontologies, which are functional to mediate multilingual geological data; (3) Ontology-based functions are useful to improve geological data interoperability and benefit both geologists and non-geologists who use the data; and (4) Geological data and ontologies are affected by their contexts and semantic negotiations can be used to achieve pragmatic interoperability between distributed geological data.

As a whole, the dissertation presents strategies and methods for deploying an ontology spectrum properly in practices to promote geological data interoperability. Geological ontologies and geological data are evolving in a long-term perspective. Common understanding of subjects in geology requires semantic negotiations among stakeholders and domain experts. Different types of ontologies in an ontology spectrum can be used to encode the commonly agreed models of subjects in geology. By using ontologies, innovative applications can be developed to promote geological data interoperability at local, regional and global levels.

Samenvatting

Geologische data zijn niet alleen onmisbaar voor het bestuderen van de aarde, maar ook bij het aangaan van belangrijke maatschappelijke uitdagingen. De uitwisselbaarheid van geologische data is al sinds lang onderwerp van zorg binnen de wetenschap. Want terwijl er sprake is van een ware stortvloed aan digitale geologische data, is de aanpak van een effectieve uitwisseling van geologisch data onderbelicht gebleven.

De uitwisselbaarheid van data kan een probleem zijn op verschillende niveaus, bijvoorbeeld op het niveau van systemen (netwerken en diensten), op een syntactisch niveau (taal en representatie in code), op het schematisch niveau (modellering en structuur), op het semantische niveau (inhoud en betekenis) of het pragmatische niveau (gebruik en gevolgen). In de afgelopen jaren is in studies naar de uitwisselbaarheid van geologische gegevens een toenemende aandacht ontstaan voor het gebruik van ontologieën. Ontologieën worden in de informatica gedefinieerd als conceptualiseringen van domeinkennis. Uit de literatuur ontstaat het beeld van een ontologie-spectrum, waarbinnen verschillende typen ontologieën met variërende semantische rijkdom bestaan.

Ondanks positieve ontwikkelingen in het maken en gebruiken van diverse typen geologische ontologieën blijven er nog verschillende uitdagingen over in het toepassen van een ontologie-spectrum ten behoeve van de uitwisselbaarheid van geologische gegevens. In deze dissertatie behandelen we vier belangrijke struikelblokken: (1) het modelleren en representeren van ontologieën; (2) de meertaligheid van geologische data en ontologieën; (3) de flexibiliteit en bruikbaarheid van ontologie-gebaseerde toepassingen; en (4) de afstemming en ontwikkeling van geologische data en ontologieën.

Het onderzoek dat in deze dissertatie wordt beschreven en bediscussieerd was gericht op de genoemde struikelblokken. Dit om de weg te bereiden voor de toepassing van een ontologie spectrum, om zo de interoperabiliteit van geologische data op lokaal, regionaal en globaal niveau te bevorderen. De dissertatie is specifiek gericht op de volgende vijf onderzoeksvragen:

(1) Hoe kunnen ontologieën worden gemodelleerd en gerepresenteerd op een zodanige wijze dat de resulterende ontologieën niet alleen bruikbaar zijn voor het harmoniseren van lokale geologische data, maar ook de uitwisselbaarheid kunnen bevorderen van die lokale of interne geologische data met extramurale of externe projecten?

(2) Hoe kan het bouwen en gebruiken van meertalige ontologieën de taalkundige barrières in online geologische data verminderen, in een regionale/globale context?

(3) Hoe kunnen verschillende conceptuele analysemethoden worden gecombineerd om thematische conceptuele schema's te ontwikkelen die bruikbaar zijn voor het oplossen van problemen en die compatibel zijn met veelgebruikte standaarden in de geologische wereld?
(4) Hoe kunnen op ontologieën gebaseerde gereedschappen worden ontwikkeld die de uitwisselbaarheid van online geologische data bevorderen en zo geologen en leken helpen bij het verzamelen van geologische informatie en het ontdekken van geologische kennis?
(5) Wat zijn de door de context veroorzaakte problemen bij de uitwisselbaarheid van geologische gegevens en hoe kunnen deze het best worden aangepakt op de lange termijn?

Om antwoord te krijgen op bovenvermelde vragen zijn case studies gedaan naar de uitwisselbaarheid van geologische data op lokaal, regionaal en globaal niveau. Verschillende typen ontologieën, zoals taxonomieën, thesauri, conceptuele schema's en ontologieën gebaseerd op logische talen, zijn onderzocht, ontwikkeld en gebruikt. Dit is gedaan in de context van de verschillende case studies, en gebaseerd op de uitkomsten geven we in deze dissertatie antwoord op de onderzoeksvragen en geven we oplossingen voor de eerder genoemde vier struikelblokken: (1) Bij het ontwikkelen en gebruiken van ontologieën moet er onderscheid gemaakt worden tussen de modellering en de representatie, maar tegelijkertijd moet deze twee ook op elkaar afgestemd worden; (2) Meertalige labels zijn bruikbaar voor het bouwen van nauwkeurige, consistente en complete meertalige geologische ontologieën, die op hun beurt weer bruikbaar zijn voor het op elkaar afstemmen van meertalige geologische datasets; (3) Op ontologieën gebaseerde functies zijn geschikt voor het verbeteren van uitwisselbaarheid van geologische gegevens en dat komt ten goede aan zowel geologen als leken die deze data gebruiken; en (4) Geologische data en ontologieën worden beïnvloed door hun context en door het onderling afstemmen van de semantiek kan op een pragmatische manier uitwisselbaarheid tussen gedistribueerde geologische data bereikt worden.

Samenvattend presenteert deze dissertatie strategieën en methoden om een ontologie-spectrum op de juiste wijze in te zetten om uitwisselbaarheid van geologische gegevens te bevorderen. Op de lange termijn evolueren zowel de geologische ontologieën als de geologische data. Voor een algemeen begrip van geologische onderwerpen is een afstemming tussen belanghebbenden en domeinexperts nodig. Om de algemeen aanvaarde modellen van geologische onderwerpen te representeren kunnen verschillende typen ontologieën in een ontologie-spectrum gebruikt worden. Door het gebruik van ontologieën kunnen innovatieve applicaties worden ontwikkeld die de uitwisselbaarheid van geologische data op lokaal, regionaal en globaal niveau bevorderen.

Publications related to the dissertation

Peer-reviewed & ISI-indexed journal papers

Ma, X., Carranza, E.J.M., Wang, X., Wu, C., van der Meer, F.D., Pragmatic interoperability approach for distributed geodata. Submitted.

Ma, X., Carranza, E.J.M., Wu, C., van der Meer, F.D., Ontology-aided annotation, visualization and generalization of geological time scale information from online geological map services. Computers & Geosciences (2011), doi:10.1016/j.cageo.2011.07.018

Ma, X., Carranza, E.J.M., Wu, C., van der Meer, F.D., Liu, G. 2011. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps. Computers & Geosciences 37 (10), 1602–1615.

Ma, X., Wu, C., Carranza, E.J.M., Schetselaar, E.M., van der Meer, F.D., Liu, G., Wang, X., Zhang, X., 2010. Development of a controlled vocabulary for semantic interoperability of mineral exploration geodata for mining projects. Computers & Geosciences 36 (12), 1512–1522.

Ma, X., Carranza, E.J.M., van der Meer, F.D., Wu, C., Zhang, X., 2010. Algorithms for multi-parameter constrained compositing of borehole assay intervals from economic aspects. Computers & Geosciences 36 (7), 945–952.

Peer-reviewed conference papers

Ma, X., Carranza, E.J.M., Wu, C., van der Meer, F.D., 2011. Practicing an ontology spectrum for geological data interoperability. In: Proceedings of IAMG'11, Salzburg, Austria, 14 pp.

Ma, X., Carranza, E.J.M., van der Meer, F.D., Wu, C., 2010. Integrating data-flow analysis and object-oriented analysis for compositing of borehole metal-grade intervals. In: Proceedings of IAMG'10, Budapest, Hungary, 9 pp.

Liu, G., Wu, C., **Ma, X.**, Wang, Y., Tian, F., 2009. Approach for interoperability of multi-source geological hazard data based on ontology and GeoSciML. In: Proceedings of Geoinformatics'09, Fairfax, USA, 7 pp.

Peer-reviewed conference abstracts

Ma, X., Carranza, E.J.M., Wu, C., van der Meer, F.D., 2011. Combining ontology and data visualization techniques to generate interactive map legends for online geological maps. Geophysical Research Abstracts 13, European Geosciences Union General Assembly 2011, Vienna, Austria, Abstract No. EGU2011-2691.

Ma, X., 2010. Modelling and approaching pragmatic interoperability of distributed geoscience data. Geophysical Research Abstracts 12, European Geosciences Union General Assembly 2010, Vienna, Austria, Abstract No. EGU2010-5180.

Ma, X., Wu, C., Liu, G., 2008. Application of 3D GIS to improve the effect of kriging method in ore reserves estimation and mining. In: Proceedings abstracts of the joint annual meeting GAC-MAC-SEG-SGA, Québec, Canada, p. 102.

Ma, X., Wu, C., van der Meer, F.D., Carranza, E.J.M., 2008. Standardization of data elements: a primary step for a geological and mineral ontology. In: Abstracts of the 33rd IGC International Geological Congress, Oslo, Norway, 1 p.

Professional publications & presentations

Ma, X., Carranza, E.J.M., van der Meer, F.D., Wu, C., 2011. Who/what bit my pie? Ontology-aided visualized interactions with an online rock age map. Oral Presentation at 2011 ITC PhD Day, Enschede, 30 slides.

Ma, X., Carranza, E.J.M., van der Meer, F.D., Wu, C., 2010. Development of a SKOS - based multilingual thesaurus for automatic translation of geological time scale terms in online geological maps. In: IUGS-CGI and OneGeology-Europe International Geoscience Language Workshop, Berlin, 54 slides. (Invited Lecture)

Ma, X., Carranza, E.J.M., van der Meer, F.D., Wu, C., 2010. Taking more gold out of rock: modelling the compositing of borehole metal-grade intervals. Oral Presentation at 2010 ITC PhD Day, Enschede, 32 slides.

Ma, X., 2009. Function of KML as a bridge between OneGeology and Google Earth. Informatization of Land and Resources, (5), 9–12. (In Chinese with English Abstract)

Biography

Xiaogang (Marshall) Ma was born in Tianmen, China on 30 December 1980. In 2002, he completed his undergraduate study at China University of Geosciences (Wuhan) and obtained a Bachelor of Engineering degree, majoring in Land Resources Management. Prior to his graduation he was awarded a scholarship to conduct a master study in the same university, which later was turned into a master-doctoral joint study, with concentration on geo-spatial databases modeling and applications. Between 2002 and 2007 he had participated in several research projects funded by governmental organizations and mining companies in China. In 2007, he was awarded a scholarship to take up a PhD study at International Institute for Geo-Information Science and Earth Observation (ITC) with affiliation to Utrecht University when his work at Wuhan was all-but-defence. He started his study at ITC in December, 2007. The defense of his work at Wuhan was held in December, 2009 and he was awarded a Doctor of Engineering degree for his dissertation “Data Elements Study in Geological and Mineral Data Integration for Digital Mine”. When he came back to Enschede on January 01, 2010, ITC was a part of University of Twente and has been called the Faculty of Geo-Information Science and Earth Observation (ITC) since then.

His research interests include geo-thesauri, geo-ontologies, geodata interoperability, geo-conceptual modeling, data visualization and geodata services with W3C® and OGC® standards. He is a member of the International Association for Mathematical Geosciences (IAMG), the Commission for the Management and Application of Geoscience Information of International Union of Geological Sciences (CGI-IUGS), and the European Geoscience Union (EGU). He is also affiliated with several commissions in the International Cartographic Association (ICA) and the International Society for Photogrammetry and Remote Sensing (ISPRS). He was awarded the IAMG Student Research Grants in 2006. He leaded the foundation and was elected the first chair of the IAMG student chapter at ITC, and he is nominated to be a candidate for the IAMG council (2012-2016). He served as the vice-chair for ITC Faculty Council and a board member for ITC PhD Student Committee in 2010.

ITC Dissertation List

http://www.itc.nl/Pub/research_programme/Graduate-programme/Graduate-programme-PhD_Graduates.html