A Formal Grammar for Geographic Information Metadata

by

Bradford G. Nickerson
and
Ying Teng

TR00-132, February 2000

Faculty of Computer Science
University of New Brunswick
Fredericton, N.B. E3B 5A3
Canada

Phone: (506) 453-4566
Fax: (506) 453-3566
E-mail: fcs@unb.ca
http://www.cs.unb.ca

Table of Contents

1. Introduction

A context free grammar (CFG) is a formal system that describes a language L by specifying how any legal sentence can be derived. Noam Chomsky [3] first investigated the application of a formal mathematical model to the structure of natural languages. Computer scientists immediately recognized the value of Chomsky's work for formally representing computer languages [e.g. 1, 2, 8]. A language L is a subset of the closure set of an alphabet. Valid sentences of a language are generated using a grammar $G = (V_N, V_T, S, \Phi)$, where

V_N = the set of nonterminal symbols,

V_T = the set of terminal symbols,
S = the starting nonterminal symbol, and
Φ = the set of productions.

Productions define rules on how proper sentences in the language can be derived. For example, the following grammar defines rules for the language of expressions in a large number of computer languages:

<E> ::= <T> | <E> {"+" | "-"} <T>
<T> ::= <F> | <T> {"*" | "/"} <F>
<F> ::= <I> | "-"<I>
<I> ::= "(" <E> ")" | <id>

The <id> nonterminal stands for any legal identifer. Note that <id> can be considered a terminal in practice as it is returned directly as a token from the lexical analyser. The notation used here is the Extended Backus Naur Form (EBNF), and is summarized in Table 1.

Table 1. Meaning of EBNF symbols (adapted from [7]).

Symbol	**Meaning**
::=	is defined to be
|	Alternatively
<text>	Nonterminal
"text"	Literal (terminal)
*	The preceding syntactic unit can be repeated zero or more times
+	The preceding syntactic unit can be repeated one or more times
{}	The enclosed syntactic units are grouped as a single syntactic unit
[]	The enclosed syntactic unit is optional (may occur zero or one time)

For a context free grammar, Φ contains only productions of the form $\alpha ::= \beta$, where α is a single element from V_N and β has one or more elements from $\{V_N, V_T\}$. The way in which context free grammars are used is powerful. For example, the way in which the above grammar is stated ensures that when the parse tree is constructed, multiplication and division have a higher priority than addition and subtraction, that unary minus has a higher priority than multiplication and division, and that expressions in parentheses are always evaluated first.

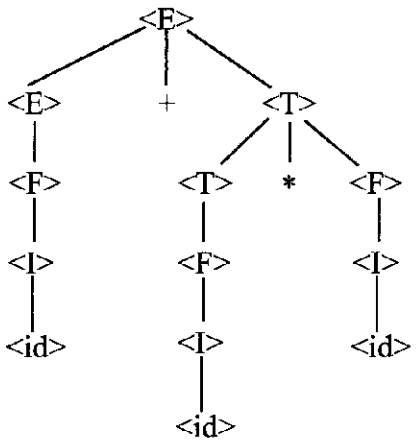
As an example of the use of the above grammar, the following is the leftmost derivation and parse tree for the sentence

<id> + <id> * <id>

derivation of <id> + <id> * <id>

<E> → <E> "+" <T>
→ <F> "+" <T>
→ <I> "+" <T>
→ <id> "+" <T>
→ <id> "+" <T> "*" <F>
→ <id> "+" <F> "*" <F>
→ <id> "+" <I> "*" <F>
→ <id> "+" <id> "*" <F>
→ <id> "+" <id> "*" <I>
→ <id> "+" <id> "*" <id>

parse tree for <id> + <id> * <id>



Due to the fact that the grammar above is unambiguous, there is only one possible way to construct the parse tree for the sentence <id> + <id> * <id>. Context free grammars are the formal basis of almost all modern day computer languages and computer representation schemes, including Java [4] and C++ [5, 9].

2. A context-free grammar for geographic information metadata

Annex B of the ISO geographic information metadata standard CD 19115 [6] contains 65 tables that describe the "data dictionary" of the metadata. For example, Table 2 below describes the metadata entity set information.

Table 2. Table B.2 metadata entity set information (from [6], p.29).

	Name / Role name	Short Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
1	MD_Metadata	Metadata	Information about the metadata	M	1	Class	Lines 2-22
2	fileIdentifier	fileID	Unique identifier for this metadata file	O	1	CharacterString	Free text
3	language	lang	Language used for documenting metadata	C / not defined by encoding?	1	Class	LanguageCode (ISO 639)
4	characterSet	charSet	Full name of the ISO character coding standard used for the metadata set	C/ISO 10646-2 not used?	1	Class	CharacterSetCode (ISO 10646-2 ISO 8859)
5	parentIdentifier	parID	Unique identifier of the parent metadata file	O	1	CharacterString	Free text
6	hierarchyLevel	hierLev	Scope to which the metadata applies (see informative Annex J for more information about metadata heirarchy levels)	C/ Scope is not equal to "dataset"?	1	Class	MD_Scope <<CodeList>>

	Name / Role name	Short Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
7	hierarchyLevelName	hierLevName	Name of the hierarchy level	C/ Scope is not equal to "dataset"?	1	CharacterString	Free text
8	contact	contact	Party responsible for the metadata information	O	1	Class	CI_ResponsibleParty <<DataType>>
9	date	date	Date that the metadata were created or last updated	O	1	Date	ISO 19108
10	metadataStandardName	mdStanName	Name of the metadata standard used	O	1	CharacterString	Free text
11	metadataStandardVersion	mdStanVer	Version of the metadata standard used	O	1	CharacterString	Free text
12	*Role name:* spatialRepresentationInfo	spatRepInfo	Digital mechanism used to represent spatial information in the dataset	O	N	Association	MD_SpatialRepresentation
13	*Role name:* referenceSystemInfo	refSysInfo	Description of the spatial and temporal reference systems used in the dataset	O	N	Association	*RS_ReferenceSystem* <<Abstract>>
14	*Role name:* metadataExtensionInfo	metExtensInf	Information describing metadata extensions	O	N	Association	MD_MetadataExtensionInformation
15	*Role name:* identificationInfo	idInfo	Basic information about the resource for which the matadata is about	M	N	Association	MD_Identification
16	*Role name:* featureCollection	featColl	A collection of geographic data to which metadata applices	M	N	Association	FT_FeatureCollection
17	*Role name:* featureCatalogueInfo	featCatInfo	Provides information about a catalogue which defines and describes the feature types, functions, attributes, and relationships, occurring in a set of geographic data	O	N	Association	FC_FeatureCatalogueDescription
18	*Role name:* distributionInfo	distInfo	Provides information about the distributor of and options for obtaining the dataset	O	1	Association	MD_Distribution
19	*Role name:* dataQualityInfo	dataQualInfo	Provides overall assessment of quality of data..	O	N	Association	DQ_DataQualityInformation (ISO 19113)
20	*Role name:* portrayalCatalogueInfo	portCatInfo	Provides information about the catalogue of rules defined for the portrayal of data.	O	N	Association	MD_PortrayalCatalogueRef
21	*Role name:* metadataConstraints	metConst	Provides restrictions on the acess and use of data	O	N	Association	MD_DataConstraints
22	*Role name:* applicationSchemaInfo	appSchInf	Provides informaiton about the conceptual schema of a dataset.			Association	MD_ApplicationSchemaInfo

	Name / Role name	Short Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
23	*Role name:* metadataMaintenance	metaMaint	Provides information about the frequency of metadata updates, and the scope of those updates.	O	1	Association	MD_MaintenanceInformation
24	*Role name:* propertyType	propTyp	Metadata is associated with the property of a feature.	O	N	Association	GF_PropertyType
25	*Role name:* featureType	featTyp	Metadata is associated with feature types.	O	N	Association	GF_FeatureType
26	*Role name:* featureAttribute	featAtt	Metadata is associated with the characteristic(s) of a feature.	O	N	Association	FT_FeatureAttribute
27	*Role name:* feature	feat	Metadata is associated with an abstraction of real world phenomena	O	N	Association	FT_Feature
28	*Role name:* aggregateDataset	aggDS	Metadata is associated with multiple datasets.	M	N	Association	DS_Aggregate

Note the 5th column that specifies O (optional), C (conditional) and M (mandatory). Table 3 shows the ISO definition of identification information.

Table 3. Table B.3 identification information (includes image identification) (from [6], p.31).

	Name / Role name	Short Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
29	MD_Identification	ID	Basic information about data	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 23-48
30	language	lang	Language(s) used within the dataset	M	N	Class	LanguageCode (ISO 639)
31	characterSet	charSet	Full name of the ISO character coding standard used for the data	C/ISO 10646-2 not used?	1	Class	CharacterSetCode (ISO 10646-2 ISO 8859-1)
32	abstract	abstract	Brief narrative summary of the content of the dataset	M	1	CharacterString	Free text
33	purpose	purpose	Summary of the intentions with which the dataset was developed	O	1	CharacterString	Free text
34	supplementalInformation	supplInfo	Other descriptive information about the dataset. Example; Data Model	O	1	CharacterString	Free text
35	credit	credit	Recognition of those who contributed to the dataset	O	1	CharacterString	Free text
36	status	status	Status of dataset	O	1	Class	MD_ProgressCode <<CodeList>>

	Name / Role name	Short Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
37	environment	envir	Description of the dataset in the producer's processing environment, including items such as the name of the software, the computer operating system, file name, and the dataset size	O	1	CharacterString	Free text
38	geographicBox	geoBox	Geographic areal domain of the dataset	C / used if geographicDescription is not used	N	Class	EX_GeographicBoundingBox
39	geographicDescription	geoDesc	Commonly used or well known name of a place, area or region which describes a spatial domain of the dataset	C / used if geographicBox is not used	N	Class	SI_LocationInstance
40	spatialResolution	spatRes	Factor which provides a general understanding of the density of spatial data in the dataset. Example: The denominator of the representative fraction or the mean ground sample distance	O	N	CharacterString	Free text
41	category	category	Keywords, describing a subject of a dataset	M	N	Class	MD_Category
42	datasetCitation	dsCitation	Recommended reference to be used for the dataset	M	1	Class	CI_Citation
43	datasetExtent	dsExt	Additional information about the bounding polygon, vertical, and temporal extent of the dataset	O	N	Class	EX_Extent
44	datasetPointOfContact	dsPOC	Identification of, and means of communication with, person(s) and organisations(s) associated with the dataset	O	N	Class	CI_ResponsibleParty <<DataType>>
45	*Role name:* datasetMaintenance	dsMaint	Provides information about the scope and frequency of updating	O	N	Association	MD_MaintenanceInformation
46	*Role name:* graphicOverview	graphOver	Provides a graphic that illustrates the dataset (should include a legend for the graphic)	O	N	Association	MD_BrowseGraphic
47	*Role name:* datasetFormat	dsFormat	Provides a description of the form of the data to be distributed	O	N	Association	MD_Format
48	*Role name:* descriptiveKeywords	descKey	Provides keywords, their type, and reference source	O	N	Association	MD_Keywords

	Name / Role name	Short Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
49	*Role name:* datasetSpecificUse	dsSpecUse	Provides basic information about specific application(s) for which the dataset has been or is being used by different users.	C/is use different than purpose?	N	Association	MD_Use
50	*Role name: datasetConstraints*	dsConst	Provides information about constraints which the dataset must fall under	O	N	Association	MD_DataConstraints
51	MD_ImageIdentification	ImageID	Information required identifying a series of images.	C/ Image series exists?	1	Specified Class (MD_Identification)	Lines 24-26
52	passSequenceIdentifier	passSeqID	Number that uniquely identifies the pass performed by a platform	M	1	Integer	Integer
53	imageOrbitalIdentifier	imagOrbID	Unique identifier for the orbital path of a platform and the row along an orbital path of a platform	M	1	CharacterString	Free text
54	orbitNumber	orbNum	Numer of the orbit in which the image was taken	M	1	Integer	Integer

Browse graphic information

55	MD_BrowseGraphic	BrowGraph	Graphic that provides an illustration of the dataset (should include a legend for the graphic)	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (MD_Identification)	Lines 49-52
56	fileName	fileName	Name of the file that contains a graphic that provides an illustration of the dataset	M	1	CharacterString	Free text
57	fileDescription	fileDesc	Text description of the illustration	O	1	CharacterString	Free text
58	fileType	fileType	Graphic file type of a related graphic file Examples: CGM, EPS, GIF, JPEG, PBM, PS, TIFF, XWD	O	1	CharacterString	Free text

Keyword information

59	MD_Keywords	Keywords	Keywords, their type and reference source	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (MD_Identification)	Lines 53-56
60	keyword	keyword	Common-use word(s) or phrase(s) used to describe the subject	M	N	CharacterString	Free text
61	type	type	Method used to group similar keywords	O	1	Class	MD_KeywordType <<CodeList>>

62	thesaurusName	thesaName	Name of the formally registered thesaurus or a similar authoritative source of keywords	O	1	CharacterString	Free text

Use information

63	MD_Use	Use	Brief description of ways in which the dataset is currently used.	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (MD_Identification)	Lines 57-61
64	specificUse	specUse	Brief description of the dataset and/or dataset series use	M	1	CharacterString	Free text
65	useDateTime	useDatTim	Date and time of the first occurrence or range of occurrences of the dataset and/or dataset series	O	1	DateTime	ISO 19108
66	userDefinedLimitations	usrDefLims	Applications for which the dataset and/or dataset series is not suitable	O	1	CharacterString	Free text
67	userContactInfo	usrContInfo	Identification of means of communicating with person(s) and organisation(s) using the dataset and/or dataset series	O	N	Class	CI_ResponsibleParty <<DataType>>

We believe it is important for any computer representation scheme such as the ISO geographic information metadata standard CD 19115 [6] to use a formal grammar. In fact, a formal grammar is the only way to clearly define the language of geographic information metadata. This allows the development of tools that can accurately parse the metadata information and that can supply good error messages back to the user defining the metadata files. Currently, the ISO standard [6] uses UML diagrams (see e.g. [10] as well as [6]) to define how the metadata objects relate to one another.

As a start at defining a formal grammar for the ISO geographic information metadata, we have translated the above two tables into a formal context free grammar. Figure 1 below shows the initial part of the grammar for table 2. The suggested complete grammar for tables 2 and 3 above is given in Appendices I and II.

```
(1) <MD_Metadata> ::= "<Metadata>"
                [ "<fileID>" <string_literal> "</fileID>" ]
                [ "<lang>" <LanguageCode> "</lang>" ]
                [ "<charSet>" <CharacterSetCode> "</charSet>" ]
                [ "<parID>" <string_literal> "</parID>" ]
                [ {"<hierLev>" "dataset" <MD_Scope> "</hierLev>"} |
                        {"<hierLev>" "notdataset"
                                "<hierLevName>" <string_literal> "</hierLevName>"
                        "</hierLev>"} ]
                [ "<contact>" <CI_ResponsibleParty> "</contact>" ]
                [ "<date>" <Date> "</date>" ]
                [ "<mdStanName>" <string_literal> "</mdStanName>" ]
                [ "<mdStanVer>" <string_literal> "</mdStanVer>" ]
                { "<spatRepInfo>" <MD_SpatialRepresentation> "</spatRepInfo>" }*
                { "<refSysInfo>" <RS_ReferenceSystem> "</refSysInfo>" }*
                { "<metExtensInf>" <MD_MetadataExtensionInformation> "</metExtensInf>" }*
                { "<idInfo>" <MD_Identification> "</idInfo>" }+
                { "<featColl>" <FT_FeatureCollection> "<featColl>" }+
                { "<featCatInfo>" <FC_FeatureCatalogueDescription> "</featCatInfo>" }*
                [ "<distInfo>" <MD_Distribution> "</distInfo>" ]
```

Figure 1. Initial part of the formal grammar for table 2 above.

Note that the production number is given in parentheses to the left of the production rule. An XML tag representation is chosen that uses the Annex B "Short Name" (see [6]) for the beginning and ending tag definition. We deliberately chose the XML syntax to simplify understanding the often long metadata text descriptions. Beginning and ending tags for each distinct metadata entity are extremely helpful for syntactic and semantic analysis of computer languages. Optional (O) items with maximum occurrence of 1 are given within square brackets []. Optional (O) items with maximum occurrence of N are given within curly braces followed by an * (i.e. {}*). Mandatory (M) items with maximum occurrence of N are given within curly braces followed by a + (i.e. {}+). All "free text" items are replaced by a <string_literal> nonterminal. All entities that refer to other domains use a nonterminal to perform the link (e.g. 12 spatialRepresentationInfo links to the <MD_SpatialRepresentation> nonterminal, 15 identificationInfo links to the <MD_Identification> nonterminal (rule 2 in Appendix II)).

Conditional items are problematic. For example, item 3 "language" from table 2 above is "Conditional (C) / not defined by encoding?". The meaning of this "obligation" is unclear, as is the meaning of the following item 4 "characterSet" "C/ISO 10646-2 not used?". For the time being, we have simply indicated that these items are optional (0 or 1 time). To be useful, the definition of these conditional items should be replaced as either Optional (O) or Mandatory (M). For example, we have changed the syntax slightly for entities 6 and 7 in table 2. Instead of having the obligation "C/ Scope is not equal to "dataset"?", we have the keyword "dataset" indicating that a <MD_Scope> nonterminal production follows. The keyword "notdataset" indicates that a "<hierLevName>" <string_literal> "</hierLevName>" entity follows that gives the name of the hierarchy level. This is one example that illustrates how a context free grammar

can precisely and clearly define how the metadata must be constructed.

Another instance where we have used the context free grammar to modify the syntax of [6] to make it precise is shown below in Figure 2.

```
(2) <MD_Identification> ::= "<ID>"
                { "<lang>" <LanguageCode> "</lang>" }+
                [ "<charSet>" <CharacterCodeSet> "</charSet>" ]
                "<abstract>" <string_literal> "</abstract>"
                [ "<purpose>" <string_literal> "</purpose>" ]
                [ "<suppInfo>" <string_literal> "</suppInfo>" ]
                .
                . /* stuff snipped */
                .
                { "<descKey>" <MD_Keywords> "</descKey>" }*
                { "<dsSpecUse>" <MD_Use> "</dsSpecUse>" }*
                [ "<dsConst>" <MD_DataConstraints> "</dsConst>" ]
                [ "<ImageID>" <MD_ImageIdentification>
                        "<passSeqID>" <integer_literal> "</passSeqID>"
                        "<imagOrbID>" <string_Literal> "</imagOrbID>"
                        "<orbNum>" <integer_literal> "</orbNum>"
                        "</ImageID>" ]
                "</ID>"
```

Figure 2. Part of the formal grammar for table 3 above.

Row 51 of table 3 above contains the "C/ Image series exists?" obligation. Our approach is to precisely define this in the grammar; i.e. if the "<ImageID>" ... "</ImageID>" tags exist (they are optional as indicating by the enclosing [ ]), then the pass sequence number, image orbital identifier and orbit number are all required to be given. There is no confusion about how the conditional obligation applies.

3. Conclusions

We have made a first attempt at defining a formal grammar for geographic information metadata. This formal grammar uses XML tags to precisely define the limits or each metadata entity. A formal context free grammar representation is required to precisely define how the geographic information metadata sentences are correctly constructed.

Some open questions and problems remain. For example, a formal grammar defines production rules in a specific order. This means that for the example grammar given in Appendices I and II, the metadata tags must be given in the same order as they are specified in the grammar. Conditional entities in the metadata data dictionary are problematic. As mentioned in Appendix III, it is important for these tags to be changed to either optional (O) or mandatory (M), and moved to their appropriate place within the other parts of the data dictionary

definitions. This will enable accurate and meaningful error messages to the end user from toolkits used to parse metadata files.

We have found an inconsistency; in table 2 above, featureCollection (item 16) is listed as mandatory, but in the UML model A.2 of Metadata schemas (Annex A of [6]) it is not displayed. In addition, it would be very helpful if XML type tags were mandatory (and not optional as explained on p.27 of [6]) to make parsing the metadata easier and more consistent with metadata representation in other domains (e.g. medical and manufacturing electronic data interchange). A formal presentation of comments for the ISO/TC 211 Secretariat is given in Appendix IV.

A good start has been made in formally defining geographic information metadata. It is important that work continue to formally define the standard [6] to benefit from the modern approaches to representing a language (e.g. that used in [7]).

4. References

1. Aho, A.V., Sethi, R. and Ullman, J.D. *Compilers Principles, Techniques, and Tools*, Addison-Wesley, Reading, MA, 1986.
2. Appel, A.W. *Modern Compiler Implementation in Java*, Cambridge University Press, New York, NY, 1998.
3. Chomsky, N. "Three models for the description of language", *IRE Trans. on Information Theory*, IT-2:3, 1956, pp.113-124.
4. Gosling, J., Joy, B. and Steele, G. *The Java Language Specification*, Addison-Wesley, Reading, MA, 1996, also available at http://java.sun.com/docs/books/jls/
5. ISO/IEC 14882, *Programming Languages - C++*, International Organization for Standardization/International Electrotechnical Commission, Sept. 15, 1998, 732 pages, available for download from http://www.cssinfo.com/ncitsgate.html
6. ISO/TC 211 Secretariat, "ISO/TC 211 Geographic information/Geomatics 2. CD 19115, Geographic information - Metadata", Committee draft for comments, ISO/TC 211 N831, available from Norwegian Technology Standards Institution, Oslo, Norway, http://www.statkart.no/isotc211/, November 30th, 1999, 118 pages.
7. Object Management Group, "The Common Object Request Broker: Architecture and Specification", minor revision 2.3.1, October, 1999, refer specifically to Chapter 3-OMG IDL Syntax and Semantics, http://www.omg.org/corba/cichpter.html, 58 pages.
8. Sebasta, R.W. *Concepts of Programming Languages*, third edition, Addison-Wesley, Reading, MA, 1996.
9. Stroustrup, B. *The C++ Programming Language*, third edition, Addison-Wesley, Reading, MA, 1997.
10. Teng, Y. and Nickerson, B.G. "XML (eXtensible Markup Language) for Geospatial Metadata", UNB Faculty of Computer Science Technical Report TR00-131, Jan. 2000, 34 pages.

Appendix I - Suggested formal grammar representation for B.2 Metadata entity set information

Note that this formal grammar relies on the existence of other lexical conventions, alphabetic characters, digit characters, graphic characters, and on the existence of integer, character, floating point and string literals. For a complete description of the formal grammar representation of geographic information metadata, a description similar to that provided for the OMG IDL Syntax and Semantics [7] is required.

(1) <MD_Metadata> ::= "<Metadata>"
["<fileID>" <string_literal> "</fileID>"]
["<lang>" <LanguageCode> "</lang>"]
["<charSet>" <CharacterSetCode> "</charSet>"]
["<parID>" <string_literal> "</parID>"]
[{"<hierLev>" "dataset" <MD_Scope> "</hierLev>"} |
{"<hierLev>" "notdataset"
"<hierLevName>" <string_literal> "</hierLevName>"
"</hierLev>"}]
["<contact>" <CI_ResponsibleParty> "</contact>"]
["<date>" <Date> "</date>"]
["<mdStanName>" <string_literal> "</mdStanName>"]
["<mdStanVer>" <string_literal> "</mdStanVer>"]
{ "<spatRepInfo>" <MD_SpatialRepresentation> "</spatRepInfo>" }*
{ "<refSysInfo>" <RS_ReferenceSystem> "</refSysInfo>" }*
{ "<metExtensInf>" <MD_MetadataExtensionInformation> "</metExtensInf>" }*
{ "<idInfo>" <MD_Identification> "</idInfo>" }+
{ "<featColl>" <FT_FeatureCollection> "<featColl>" }+
{ "<featCatInfo>" <FC_FeatureCatalogueDescription> "</featCatInfo>" }*
["<distInfo>" <MD_Distribution> "</distInfo>"]
{ "<dataQualInfo>" <DQ_DataQualityInformation> "</dataQualInfo>" }*
{ "<portCatInfo>" <MD_PortrayalcatalogueRef> "</portCatInfo>" }*
{ "<metConst>" <MD_DataConstraints> "</metConst>" }*
{ "<appSchInf>" <MD_ApplicationSchemaInfo> "<appSchInf>" }*
["<metaMaint>" <MD_MaintenanceInformation> "</metaMaint>"]
{ "<PropTyp>" <GF_PropertyType> "</PropTyp>" }*
{ "<featTyp>" <GF_FeatureType> "</featTyp>" }*
{ "<featAtt>" <FT_FeatureAttribute> "</featAtt>" }*
{ "<feat>" <FT_Feature> "</feat>" }*
{ "<aggDS>" <DS_Aggregate> "</aggDS>" }+
"</Metadata>"

Appendix II - Suggested formal grammar representation for B.3 Identification information

(2) <MD_Identification> ::= "<ID>"
{ "<lang>" <LanguageCode> "</lang>" }+
["<charSet>" <CharacterCodeSet> "</charSet>"]
"<abstract>" <string_literal> "</abstract>"
["<purpose>" <string_literal> "</purpose>"]
["<suppInfo>" <string_literal> "</suppInfo>"]
["<credit>" <string_literal> "</credit>"]
["<status>" <MD_ProgressCode> "</status>"]
["<envir>" <string_literal> "</envir>"]
{ { "<geoBox>" <EX_GeographicBoundingBox> "</geoBox>" } |
{ "<geoDesc>" <SI_LocationInstance> "</geoDesc>" } }
{ "<spatRes>" <string_literal> "</spatRes>" }*
{ "<category>" <MD_Category> "</category>" }+
{ "<dsCitation>" <CI_Citation> "</dsCitation>" }+
{ "<dsExt>" <EX_Extent> "</dsExt>" }*
{ "<dsPOC>" <CI_ResponsibleParty> "</dsPOC>" }*
{ "<dsMaint>" <MD_MaintenanceInformation> "</dsMaint>" }*
{ "<graphOver>" <MD_BrowseGraphic> "</graphOver>" }*
{ "<dsFormat>" <MD_Format> "</dsFormat>" }*
{ "<descKey>" <MD_Keywords> "</descKey>" }*
{ "<dsSpecUse>" <MD_Use> "</dsSpecUse>" }*
["<dsConst>" <MD_DataConstraints> "</dsConst>"]
["<ImageID>" <MD_ImageIdentification>
"<passSeqID>" <integer_literal> "</passSeqID>"
"<imagOrbID>" <string_Literal> "</imagOrbID>"
"<orbNum>" <integer_literal> "</orbNum>"
"</ImageID>"]
"</ID>"

(3) <MD_BrowseGraphic> ::= "<BrowGraph>"
"<fileName>" <string_literal> "</fileName>"
["<fileDesc>" <string_literal> "</fileDesc>"]
["<fileType>" <string_literal> "</fileType>"]
"</BrowGraph>"

(4) <MD_Keywords> ::= "<Keywords>"
{ "<keyword>" <string_literal> "</keyword>" }+
["<type>" <MD_KeywordType> "</type>"]
["<thesaName>" <string_literal> "</thesaName>"]
"</Keywords>"

(5) <MD_Use> ::= "<Use>"
"<specUse>" <string_literal> "</specUse>"
["<useDatTim>" <DateTime> "</useDatTim>"]

[“<usrDefLims>” <string_literal> “</usrDefLims>”]
{ “<usrContInfo>” <CI_ResponsibleParty> “</usrContInfo>” }*
“</Use>”

Appendix III - E-mail correspondence between Yves Hudon, Brad Nickerson and Jakes Knopper

From bgn@unb.ca Tue Feb 15 09:05:31 2000
Date: Tue, 8 Feb 2000 16:24:22 -0400 (AST)
From: Brad Nickerson <bgn@unb.ca>
To: Yves Hudon <Yves.Hudon@sct.gouv.qc.ca>
Cc: Jake Knoppers <mpereira@istar.ca>, Ray Gates <Ray_Gates@manulife.com>,
TENG YING <n74v9@unb.ca>,
"[iso-8859-1] Bédard, Yvan" <yvan.bedard@scg.ulaval.ca>
Subject: Re: ISO TC/211 N 831 CD 19115, Geo.inf. - Metadata

[The following text is in the "X-UNKNOWN" character set.]

[Your display is set for the "US-ASCII" character set.]

[Some characters may be displayed incorrectly.]

Yves:

Ying Teng and I have made a first pass at representing tables B.2 (p.29) and B.3 (p.31) using formal context free grammar notation. We have the following comments:

1. C (conditional) obligations (items 3, 4, 6, 7, 31, 49, 51) refer to other standards (e.g. ISO 639, 10646-2, 8859, 19108) that might or might not have formal grammars defining them. This complicates defining clearly the formal grammar for CD 19115.

   In addition, these items imply a dependence on the existence and syntax used in other nonterminals appearing much later in the grammar. This makes writing meaningful productions that have a "C" obligation/condition almost impossible. One can only write it as if it were optional, which means that error detection and reporting is impeded. If possible, I would remove all type "C" obligation/condition codes, and replace them with either "O" or "M".

   One can handle the dependency in a different way, if necessary. For example, items 6 and 7 refer to "Scope is not equal to "dataset"?". These conditional tems can be moved out of the tables B2 and B3 and be defined in the tables (non-terminals) they reference (i.e. table B5.1 p.35 in this example).

This is the proper way to do it to give the compiler writer maximum opportunity to detect and report errors.
Items 38 and 39 are an exception. They depend directly on one another, so the context free grammar (CFG) syntax simply states that either geoBox or geoDesc appears, but not both. They also refer to items directly in this level (table B3), so the forward referencing difficulty is avoided.

2. On page 27, it states that (item B.1.3) "Implementation using SGML and XML is not mandatory". Life would be VERY much simpler is this statement was changed to "Implementation using XML is mandatory". XML is, by definition (see http://www.w3.org/TR/1998/REC-xml-19980210.pdf) a subset of SGML, and all XML documents are conforming SGML documents. Forcing people to use XML would very much standardize the way in which geographical metadata is communicated around the world, with proper beginning and ending tags for each construct. It would also make the implementor's job much easier, as well as conform to much larger efforts in other domains (e.g. medical information, manufacturing EDI).

3. Context free grammar specification implies an order to the appearance of tokens. I assume that the order implied in tables B2 and B3 are required.

Ying and I are drafting a small technical report that shows our current interpretation of tables B2 and B3 in CFG. I will forward a copy on to you when it is ready (later this week).

Regards, Brad Nickerson

Dr. Brad Nickerson
Professor and Director of the Information Technology Centre
University of New Brunswick
Faculty of Computer Science
P.O. Box 4400
Room GE-119, Gillin Hall, 540 Windsor Street
Fredericton, N.B. E3B 5A3
Canada
phone: (506) 458-7278 fax: (506) 453-3566 E-mail: bgn@unb.ca

On Thu, 3 Feb 2000, Yves Hudon wrote:

> Brad,
> Thank you for your last email.
> I should have mentionned what are the target dates I have on my shoulders to make
> possible that Canada sends officially comments to ISO/TC 211. The ISO TC 211 target
> date is 2000-02-27, but the Standard Council of Canada due date I have to meet was
> 2000-01-27, for which I asked a delay until tomorrow 2000-02-04.
>
> Also, I am leaving on Tuesday to Toronto to attend a Geoconnections meeting on
> standards on Febr 9th and a CGSB meeting on geomatics on Feb 10th. The best for me
> would be to get your example not later than Monday 10 AM. If not possible, as soon as
> possible later this target. I feel after nxt Friday, it will be hard to manage this
> comment at ISO/Tc 211 level.
>
> Do you best. Appreciate very much your input.
>
> Amitiés
>
> Yvcs
>
> Brad Nickerson a écrit :
>
> > Yves:
> >
> > Thanks for your prompt response and to Jake and yourself for
> > suggesting a way ahead. I don't have the time this week to
> > complete such a task as outlined below this week, but I will try
> > to get a look at completing something by next week.
> >
> > Regards, Brad Nickerson
> >
> > On Wed, 2 Feb 2000, Yves Hudon wrote:
> >
> > > Brad,
> > >
> > > Thank you very much for your input. You raised an important an interesting
> > > issue.
> > >
> > > Canada could send your comment to ISO/TC 211. For acceptance by ISO/TC 211, a
> > > brief description of what are the problems will help to convince people. As
> > > suggested by Jake,
> > > «Consequently and in conclusion, I urge you to take existing CD 19115 and to
> > > cast it into a formal representation CORBA IDL equivalent and submit the
> > > same as an Normative/Informative Annex. In doing this, you will undoubtedly

> > > have questions for clarification, possible illogical aspects, etc. which
> > > need to be addressed/resolved in the plain text clauses of CD 19115.»
> > >
> > > Could you do a short example (let's say tables B.2 and B.3)? Doing your example,
> > > you should then list all questions for clarification that needs to be solved
> > > because the UML model and the Data dictionary is not enough or are sometimes in
> > > contradiction. Can you send to me this example ASAP, let's say tomorrow Thursday
> > > 16h00, or Friday 13h? Please confirm.
> > >
> > > Amitiés
> > >
> > > Yves
> > >
> > > "Infoman Inc." a écrit :
> > >
> > > > Yves,
> > > >
> > > > Thanks for passing on this e-mail to me. I am replying to the two parties
> > > > concerned and have copies in Ray Gates as well.
> > > >
> > > > To Brad and Teng, my name is Jake Knoppers. I used to be Head of Delegation
> > > > for Canada for JTC1/SC30 Open-edi. After the JTC1 re-engineering this SC was
> > > > incorporated with several others into a new SC32 "Data Management and
> > > > Interchange. ISO/IEC JTC1 is the joint committee of ISO and IEC for all
> > > > standardization activities pertaining to information technology, i.e. those
> > > > which are not application area specific such as those of specific ISO TCs.
> > > > For all practical intends and purposes an ISO/IEC JTC1 SC functions as an
> > > > ISO TC.
> > > >
> > > > 1. I fully agree that any standard whose requirements and specifications are
> > > > intended to be implemented through the use of information technologies
> > > > should have those requirements and specifications captured using a Formal
> > > > Description Techniques(FDT) defined in ISO/IEC JTC1 Directives as "a
> > > > specification method based on a description language using rigorous and
> > > > unambiguous rules both with respect to developing expressions in the
> > > > language (formal syntax) and interpreting the meaning of these
> > > > expressions(formal semantics).
> > > >
> > > > 2. Experience has demonstrated that as information technologies
> > > > improve/progress so thus the capabilities and scope of FDT tools. Initially
> > > > we had LOTUS and ESTELLE as ISO standard FDTs, then later EXPRESS(-G) and
> > > > IDL (and others), and now UML is being progressed as an ISO/IEC standard via
> > > > JTC1/SC7. There is also MOF of OMG.
> > > > Consequently, the capabilities of FDTs will change and improve and whatever

> > > > the "flavour of the month" it will be superseded by new and improved "X"
> > > > every 3-5 years.
> > > >
> > > > 3.In our standardization work in the area of electronic data
> > > > interchange(EDI), now marketed as e-commerce, e-business, etc., we
> > > > encountered quasi-religious/theological debates on the "best" FDT. The
> > > > issues was settled through the introduction of the following term +
> > > > definition,
> > > >
> > > > "Open-edi Description Technique(OeDT): a specification method such as a
> > > > Formal Description Technique, another methodology having the characteristics
> > > > of a Formal Description technique or combination of such techniques to
> > > > formally specify BOV concepts, in a computer processible form." [See further
> > > > ISO/IEC 14662:1997 "Information technology - Open-edi reference model".
> > > >
> > > > For your information "BOV" refers to the term/definition "Business
> > > > Operational View(BOV): a perspective of business transactions limited to
> > > > those aspects regarding the making of business decisions and commitments
> > > > among organisations which are need for the description of a business
> > > > transaction" [ISO/IEC 14662:1997(3.1.3).
> > > >
> > > > 4. To relate and bring back the above to your comments on TC211's CD 19115
> > > > Geographic information - Metadata my comments are as follows:
> > > >
> > > > 4.1 It is vital that CD 19115 contains CLEAR, PLAIN TEXT logical statements
> > > > of agreed upon user/business requirements and specifications for geo
> > > > metadata. Personally, I would like to see these cast in the form of a
> > > > rule-base with concise terms and definitions.
> > > >
> > > > 4.2 The standard, in this case CD 19115 would then include
> > > > normative/informative Annexes (to be decided by TC 211) which capture such
> > > > clear, plain text user/business requirements using one or more FDTs.
> > > > Personally, I would like to see more than one FDT being used in the Annexes.
> > > > For example, it would be of benefit to all, if CD 19115 would contain as
> > > > Annexes both the use of UML and CORBA IDL equivalent.(MOF accepts the
> > > > existence of both).
> > > >
> > > > 4.3 Further the use of one or more FDTs serves as a quality control check on
> > > > any standard which is to be implemented via IT. The key purpose here is to
> > > > ensure that user/business requirements have been stated clearly and
> > > > unambiguously. This is/should be an interactive standards development
> > > > process.
> > > >
> > > > 4.4 However, the main body of the standard must focus specifying the

> > > > "WHAT's" in terms of agreed upon user/business requirements (includes both
> > > > for-profit and not-for-profit)in clear, plain text. After all the languages
> > > > of ISO standards are English (as well as French and Russian). Mapping such
> > > > requirements into FDTs serves as the crucial bridge to specifying the
> > > > "HOW's". The two are complimentary and NOT competitive.
> > > >
> > > > 5.0 To conclude, your comments on CD 19115 are valid since in its current
> > > > state it is neither "fish nor fowl". The contents of CD 19115 in terms of
> > > > user/business requirements need to be stated in PLAIN TEXT, preferably
> > > > rule-based. In addition, this standards development work needs to be
> > > > accompanied by Annexes where these user/business requirements are captured
> > > > using FDTs,i.e. in your words "formal grammar defining the metadata".
> > > >
> > > > Consequently and in conclusion, I urge you to take existing CD 19115 and to
> > > > cast it into a formal representation CORBA IDL equivalent and submit the
> > > > same as an Normative/Informative Annex. In doing this, you will undoubtedly
> > > > have questions for clarification, possible illogical aspects, etc. which
> > > > need to be addressed/resolved in the plain text clauses of CD 19115.
> > > >
> > > > I trust that these comments are of some help.
> > > >
> > > > Best regards - Jake Knoppers (613) 234-3244[fax:613-234-3935]
> > > >
> > > > > -----Original Message-----
> > > > > From: Brad Nickerson [mailto:bgn@unb.ca]
> > > > > Sent: February 1, 2000 4:42 AM
> > > > > To: Yves Hudon
> > > > > Cc: TENG YING
> > > > > Subject: Re: ISO TC/211 N 831 CD 19115, Geo.inf. - Metadata
> > > > >
> > > > >
> > > > > Yves:
> > > > >
> > > > > We have been looking at the above document, and I would like
> > > > > to record our comment that the above document requires a
> > > > > formal grammar defining the metadata. Such formal grammars
> > > > > defining the production rules are an essential part of any
> > > > > modern computer-based language no matter what it represents.
> > > > > For example, the CORBA Interface Definition Language uses a set
> > > > > of 98 production rules to clearly define the IDL grammar. See
> > > > > http://www.omg.org/corba/cichpter.html
> > > > > and click on "IDL Synax and Semantics" for a good discussion
> > > > > of this representation. Chapter 3.4 gives a precise definition
> > > > > of the IDL grammar. Note that it is imperative to use a formal

> > > > > notation such as the Extended Backus-Naur Form (EBNF) used for
> > > > > IDL. This form is defined well in the above reference, but
> > > > > essentially requires rules to be written using the notation
> > > > >
> > > > > Symbol Meaning
> > > > > ------ -------
> > > > > ::= is defined to be
> > > > > | alternatively
> > > > > <text> nonterminal
> > > > > "text" literal
> > > > > * preceding syntactic unit can be repeated zero or more times
> > > > > + preceding syntactic unit can be repeated one or more times
> > > > > {} enclosed syntactic units are grouped as a single syntactic unit
> > > > > [] enclosed syntactic unit is optional - may occur zero or
> > > > > one time
> > > > >
> > > > > Please note that the CORBA IDL is a completely object oriented
> > > > > language.
> > > > >
> > > > > Such a formal representation is essential to enable the
> > > > > development of computer based tools for handling, translating
> > > > > and interpreting ISO geographic information metadata, as well
> > > > > as any other ISO representation. Please don't take this comment
> > > > > the wrong way. The current document is valuable, and does lay
> > > > > out the design well using UML diagrams to formally describe the
> > > > > object relationships. A formal grammar is, in my opinion, a
> > > > > necessary addition to correctly define what the ISO TC/211
> > > > > geographic information metadata is.
> > > > >
> > > > > A form of such production rules is included with the FGDC
> > > > > Content Standard for Digital Geospatial Metadata (although it
> > > > > does not use a commonly accepted standard for defining it), as
> > > > > well as the SAIF (Spatial Archive and Interchange Format)
> > > > > description. The ISO TC/211 geographic information metadata
> > > > > definition would be improved substantially with the addition
> > > > > of a formal grammar defining it.
> > > > >
> > > > > Regards, Brad Nickerson
> > > > > Dr. Brad Nickerson
> > > > > Professor and Director of the Information Technology Centre
> > > > > University of New Brunswick
> > > > > Faculty of Computer Science
> > > > > P.O. Box 4400
> > > > > Room GE-119, Gillin Hall, 540 Windsor Street

> > > > > Fredericton, N.B. E3B 5A3
> > > > > Canada
> > > > > phone: (506) 458-7278 fax: (506) 453-3566 E-mail: bgn@unb.ca
> > > > >
> > > > >
> > > > >
> > > > >
> > >
> > > --
> > > YVES HUDON (Yves.Hudon@sct.gouv.qc.ca)
> > > Secrétariat du Conseil du trésor du Québec / SSIRI
> > > 875, Grande Allée Est, Section 3-C
> > > Québec (Québec) G1R 5R8 Canada
> > > Tél.: +1 418-528-6125; Téléc. (fax): +1 418-646-3571
> > >
> > > ? Métadonnées principales en géomatique? Voir la proposition pour
> > > le gouvernement du Québec qui s'appuie sur les travaux de l'ISO/CT 211
> > > (http://www.tresor.gouv.qc.ca/autorout/geo-meta.htm)
> > > --
> > >
> > >
>
> --
> YVES HUDON (Yves.Hudon@sct.gouv.qc.ca)
> Secrétariat du Conseil du trésor du Québec / SSIRI
> 875, Grande Allée Est, Section 3-C
> Québec (Québec) G1R 5R8 Canada
> Tél.: +1 418-528-6125; Téléc. (fax): +1 418-646-3571
>
> ? Métadonnées principales en géomatique? Voir la proposition pour
> le gouvernement du Québec qui s'appuie sur les travaux de l'ISO/CT 211
> (http://www.tresor.gouv.qc.ca/autorout/geo-meta.htm)
> --
>
>
>

Appendix IV - Suggested comments to ISO/TC 211 Secretariat on document ISO/TC 211 N831 2. CD 19115, Geographic information - Metadata

Country	Clause/ subclause	Paragraph/ Figure/ Table	Type of comment	Comment	Proposed change
BGN-1	B.1.3	1	T	The note that implementation using SGML and XML is not mandatory should be changed	Change to read "NOTE: Implementation using XML is mandatory" to bring into line with other modern metadata standards. XML is a subset of SGML. A suggested XML encoding is in Nickerson and Teng, [2000][1]
BGN-2	Annex A and Annex B	class diagram for MD_Metadata, data dictionary in Annex B	T	Missing relationship with and definition of FT_FeatureCollection, GF_PropertyType, GF_FeatureType, FT_FeatureAttribute, DS_Aggregate	Add missing relationships and missing class diagrams for FT_FeatureCollection, GF_PropertyType, GF_FeatureType, FT_FeatureAttribute, DS_Aggregate in Annex A, define these in Annex B, or eliminate them from table B.2
BGN-3	B.2	row 16 of table	E	4th column, last word "applices"	change spelling to "applies"
BGN-4	B.2	row 22 of table	T	missing obligation/condition and maximum occurrence for applicationSchemaInfo	add missing obligation and maximum occurrence
BGN-5	entire document	additional Annex	T	the document is missing a formal	add an Annex defining a formal context free

				context free grammar definition of geographic information metadata. Such a formal definition is required to define the geographic information metadata properly, and to enable the development of useful computer software for interpretation, translation and verification of ISO/TC 211 geographic information metadata	grammar for geographic information metadata. An initial incomplete attempt at defining such a grammar is in Nickerson and Teng, [2000][1]
BGN-6	Annex B	all tables	T	Conditional (C) obligations are problematic for use in a formal grammar definition. It is unclear what is required and what isn't required in each case.	Change conditional (C) obligations to Mandatory (M) or Optional (O), and use the formal grammar to clearly define what is required for each entity. An initial attempt at defining such a grammar is in Nickerson and Teng, [2000][1]

[1] Nickerson, B.G. and Teng, Y. "A Formal Grammar for Geographic Information Metadata", University of New Brunswick Faculty of Computer Science Technical Report TR00-132, Fredericton, N.B., Canada, February, 2000, 25 pages.