

Comparison of Metadata Standards – A Proposal for Hungarian Core Metadata Standard

Enikő Kovács
Institute of Geodesy, Cartography and Remote Sensing
H-1149 Budapest, Bosnyák tér 5.
Eniko.kovacs@fomigate.fomi.hu

and

Tamás Prajcer
GeoX Kft.
1033 Budapest, Fő tér 1.
prajcer@geox.hu

1. Introduction

More and more data are developed by the geospatial community and these data often can serve many applications. The rapid increase of data products available in the public administration have resulted an increasing demand for metadata. Metadata – the data about data – describe the content, quality, conditions and other characteristics of data¹. Metadata help us to locate and understand data. Metadata have great importance in information exchange because it allows users to specify their requirements and suppliers to provide information about their products. It provides a means of selective data access.

¹ Several definitions are available for geospatial metadata.

"Metadata are »data about data«. They describe the content, quality, condition and other characteristics of data. Metadata help a person to locate and understand data"

"Metadata can be defined at its simplest as »data about«... This includes information about the content, representation, extent (both geographic and temporal), spatial reference, quality and administration of a dataset".

"Information describing a data set, including data user guide, description of the data set in directories, catalogues, and inventories, and any additional information required to define the relationships among these" .

"The data bases do not hold data sets themselves, but records that are known as »metadata«. Metadata are »data about data«. They provide such information as the characteristics of a data set, the history of a data set, and organisations to contact to obtain a data set".

"Data that characterise other data and are used to document the other data so that they may be properly and effectively used. Metadata describe the content, quality condition, and other characteristics of data".

"Metadata are information about data. Metadata contain information about the source(s), lineage, content, structure, and availability of a data set. Metadata also provide descriptions of the intent and potential uses of data".

The most important impulse to avoid parallel produce of data and to ensure sifting out them was given in the USA by the Federal Geographical Data Committee (FGDC). In 1992 the FGDC sponsored the "Information Exchange Forum on Spatial Data", which identified the need for a common definition of metadata. The "Co-ordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure (NSDI)" requires Federal agencies to use the standard to document their geospatial data. Recently more than 70 federal institutions are involved into the clearinghouse which supply metadata according to FGDC's metadata standard.

The standard specifies the elements needed to support three major uses of metadata:

- To maintain an organisation's *internal investment* in geospatial data
- To provide information to data *clearinghouses* and catalogues
- To provide information needed to process and interpret data *transferred* from another organisation.

The standard defines the information required by a prospective user

- To determine the *availability* of a set of geospatial data
- To determine the *fitness* of a set of geospatial data for an intended use
- To determine the means to *access* the set of geospatial data
- To *transfer* the set of geospatial data successfully.

The standard specifies the information content for a set of geospatial data. The standard establishes a common set of terminology and definitions for concepts related to metadata: the names of data elements and compound elements (groups of data elements), the definitions of data elements and information about the domains. The standard also specifies the elements that are mandatory, mandatory if applicable and optional.

Although a lot of standards follow the FGDC's standard in its spirit, it has got plenty of critics especially from two points of view: because of its size and its generality. The 'size-problem' has been solved with use of minimal version. The draft of ISO standard has already minimal version, called Level 1. The speciality problems appear first of all due to insufficiency of metadata for descriptive data. This problem has been solved by the adoption of the metadata standard for specific application areas (NBII, 1997, Cultural, 1997). The second version of the standard allows the application of specific extensions (FGDC, 1998).

In addition to the international standards a lot of national-wide and internal metadata descriptions exist. Internal 'standards' are used for remote sensed data and data-catalogues or product descriptions (ArcData).

Developments of metadata standards has started in Europe later, after a comprehensive surveying of the European Union in the frame of the Info2000 program. The European developments were sponsored by the European Union, were started through tender on base of CEN pre-standards. The Belgian, Holland and Portuguese metadata service are based on the CEN TC 287 metadata pre-standard.

After 1996 – as a consequence of the brushing up the activity of the ISO TC 211 and the Open GIS Consortium (OGC) the metadata standards have became richer with new elements.

Besides the European standardisation efforts North-American (FGDC) and the global (ISO) ones have been appeared the European applications (MEGRIN, ESMI). These metadata services have had a direct impact on the development of the Hungarian metadata content.

2. Ideas of the development of minimal metadata content

Because of the implementation of the Hungarian metadata service – METATÉR project – a minimal metadata content, called HunCore 1.0 has been developed.

At the development of this minimal „core” metadata content the existing international standardisation efforts were taken into account. Unfortunately they are not compatible with each other, therefore the Hungarian minimal version is larger than the used „core” versions.

At the development of the Hungarian core metadata content the following aspects were taken into consideration:

- Elements of the Hungarian core metadata have to contain the most important, generally accepted metadata elements (Dublin Core, GILS element, X.500 directories).
- The data description has to meet with the demand of users.
- The metadata content has to be usable or extendable for specific application areas (for instance geology, statistic, etc).
- The Hungarian minimal metadata content has to be compatible with the existing standards or initiatives (e.g. FGDC, ISO 211, CEN 287).
- The adoption procedure of the EU standards is planned.
- It has to have compatibility with the European metadata systems.

At the definition of the metadata content the following standards, drafts and proposals were taken into account:

- Content Standard for Digital Geospatial Metadata (Federal Geographic Data Committee, April, 1997)
- Geographic Information – 15046 Part 15: Metadata (ISO/TC 211/WG, December 22, 1997)
- Geographical Data Description Directory version 2.1 (MEGRIN Group)
- Geographic Information – Data Description – Metadata (CEN/TC 287/WG 2, November, 1996)
- GILS Element Definition
<http://www.usgs.gov/gils/elements.html>
- Dublin Core Elements
http://www.purl.org/metadata/dublin_core_elements
- ESMI Search Variables and Metadata
05/21/98, ver. 2.1

The idea of minimal geospatial metadata content is defined by the FGDC and ISO Pre-standard. The Dublin Core unambiguously defines the minimal elements, while the elements of the GILS-s are suitable for starting point because of their quantities. The minimal version is missing from the European proposals, the obligatory elements are not always the same as the above mentioned minimal element.

The definition of minimal metadata in the FGDC Standard is as follows:

„Mandatory sections in the standard have some elements that are always required for all types of geospatial data sets. For comparison with other metadata standards, these elements are referred to as “core” elements.“

The definition of minimal metadata in the ISO Standard is as follows:

„Compliance Level 1 is the minimum metadata required to uniquely identify a dataset ... This level of compliance shall be used to describe datasets only for the purposes of cataloging datasets and clearinghouse activities”.

The HunCore can not replace the Hungarian geospatial metadata standard or an adoption of the European one. But it is a base for creating descriptions and starting the services.

3. The HunCore 1.0 metadata content

Metatér ID	GEO Attribute ID	GEO Structure	Type	Domain	Name	Definition
1					Dataset overview	Information for overall description of the dataset.
1.1.1	4	105	text	free text	Title	The explicit name of the dataset, to sufficiently identify it by the users.
1.1.2	43	105	text	free text	Abbreviated title	Short name of the dataset.
1.1.3	5	105	text	free text	Series name	Name of the parent dataset to which this dataset belongs together with other ones.
1.3	3805	105	list element	cl	Product	The mode in which the data set is represented.
1.4	31	5	date	yyyymmdd	Publication Date	Date when the dataset was published or otherwise made available.
1.5	54	105	list	ISO 639	Language	Language and character set used within the dataset, identified by codes defined in ISO 639.
1.6	62	105	text	free text	Summary	Brief summary of the content of the dataset.
1.7	2003	105	text	free text	Purpose of Production	Summary of the intentions with which the dataset was developed
1.8	3108	105	text	free text	Progress	Status of the dataset
1.9	3109	105	text	free text	Update Frequency	Frequency with which changes and additions are made to the dataset.
1.10	2004	105	text	free text	Access Constraints	Constraints regulating the access of the dataset.
1.11	2005	105	text	free text	Use Constraints	Constraints regulating the usage of the dataset, other than copyright.
1.12	6911 (2001)	105	text	free text	Copyright owners	Name of the organisation(s) owning the copyright of the dataset
1.13	3137	204			Sample	One or more examples taken from the dataset being representative for the whole dataset.
1.13.1	3138	105	text	free text	Name	Names of the sample files (without file extensions).
1.13.2	3140	105	list	cl	Format	Formats in which the examples are available.
2					Administrative	Information about organisation(s) and person(s) associated with the data set.
2.1	3008	204			Organisation	Organisations and their role in relation to the dataset.
2.1.1	2024	105	text	free text	Organisation Name	Name of the organisation
2.1.2	6912 (2000)	105	list	cl	Organisation Role	Responsibility of the organisation in relation to the dataset.
2.2	3004	204			Contact person	Contact person
2.2.1	2023	105	text	free text	Contact person name	Whole name of the contact person for further information (title, first name, middle name, last name). Family name is indicated with capital letters.
2.2.2	6913	105	list	cl	Contact person role	Responsibility of the contact person in

	(3001)					relation to the dataset.
2.2.3	2032	105	text	free text	Telephone	Telephone number on which the contact person can be reached (without country code).
2.2.4	2033	105	text	free text	Facsimile	Facsimile phone number.
2.2.5	2030	105	text	free text	Contact electronic mail address	Address of the electronic mailbox of the contact person
2.3	3006	204			Point of Contact Address	Address of the contact person
2.5.1	2025	105	text	free text	Postal Address	Address line for the address (visiting address or P.O. Box)
2.5.2	2026	105	text	free text	City	City of the address
2.5.3	2028	105	text	free text	Postal Code	ZIP or other postal code of the address
2.5.4	2029	105	list element	ISO 3166	Country	Country of the address
2.5.7	3618	105	text	free text	Network Address	Network address from which more information or the dataset itself can be available.
3					Dataset Extent	The horizontal extent and time period covered by the dataset.
3.1	6923	204			Bounding Extents	The horizontal spatial coverage of the dataset
3.1.1	2060	201			Geographic Coordinates	WGS84 bounding coordinates
3.1.1.1	2038	200	float number	Earth longitude	West Bounding Coordinate	Western-most coordinate of the limit of coverage
3.1.1.2	2041	200	float number	Earth latitude	South Bounding Coordinate	Southern-most coordinate of the limit of coverage
3.1.1.3	2039	200	float number	Earth longitude	East Bounding Coordinate	Eastern-most coordinate of the limit of coverage
3.1.1.4	2040	200	float number	Earth latitude	North Bounding Coordinate	Northern-most coordinate of the limit of coverage
3.1.2	6918	201			EOV Coordinates	EOV bounding coordinates
3.1.2.1	6919	200	float number		Y min coordinate	Western-most coordinate of the limit of coverage
3.1.2.2	6920	200	float number		X min coordinate	Southern-most coordinate of the limit of coverage
3.1.2.3	6921	200	float number		Y max coordinate	Eastern-most coordinate of the limit of coverage
3.1.2.4	6922	200	float number		X max coordinate	Northern-most coordinate of the limit of coverage
3.2	2042	105	text	free text	Geographic Extent Name	Name of the areal units which are covered in the dataset
3.3	6914 (3906)	204			Temporal Extent	Time period covered by the content of the dataset.
3.3.1	2072	5	date	yyyymmdd	Begin Date	from date
3.3.2	2073	5	date	yyyymmdd	End Date	to date
4					Keywords	Words summarizing an subject of the dataset.
4.1	3122	204			Theme	Subjects covered by the data set.
4.1.1	2002	105	text	free text	Theme Keyword	Common-use words or technical terms used to describe the subject of the dataset.
4.1.2	2036	105	text	free text	Name of theme keyword	Name of the registered thesaurus or a similar authoritative source of

						keywords (e.g. standard, law)
4.2	3128	204			Stratum	Layered vertical locations characterised by the data set.
4.2.1	3130	105	text	free text	Stratum Keyword	Common-use words or technical terms used to describe the dataset.
4.2.2.	3129	105	text	free text	Name of stratum keyword thesaurus	Name of the registered thesaurus or a similar authoritative source of keywords (e.g. standard, law)
4.3	3131	204			Temporal	Temporal period(s) characterised by the data set.
4.3.1	3133	105	text	free text	Temporal Keyword	Common-use words used to describe the dataset.
4.3.2	3132	105	text	free text	Name of temporal keyword thesaurus	Name of the registered thesaurus or a similar authoritative source of keywords (e.g. standard, law)
4.3	6902	204			Other	Other keywords, not appropriate for the above mentioned three categories.
4.3.1	6903	105	text	cl	Other Keyword Type	Method used to group similar keywords (e.g. discipline). Place keywords are used at Geographic Extent Name.
4.3.2	6904	105	text	free text	Other Keyword	Common-use words or technical terms used to describe the subject of the dataset.
4.3.3	6905	105	text	free text	Name of other keyword thesaurus	Name of the registered thesaurus or a similar authoritative source of keywords (e.g. standard, law)
5					Quality	General assessment of the quality of the data set.
5.1	3212	205	text	free text	Overall positional accuracy	Average intervals in meters of the accuracy of the geographic position of the objects within the dataset compared to its nominal ground: horizontal accuracy or 3D accuracy.
5.2	3207	105	text	free text	Overall completeness	Average percentage of conformance of the dataset compared to its nominal ground with respect to the presence of objects, association instances, and property instances.
5.3	3503	105	list	cl	Spatial Object Type	Name of the point, vector and raster objects used to locate zero-, one-, two-, or three-dimensional spatial locations in the dataset (topology should be mentioned if exists).
5.4	3219	204			Linage	Information about the source data and processing steps
5.4.1	3221	105 (204)	text	free text	Source	List of sources used to create the dataset.
5.4.2	3228	105	text	free text	Processes	Explanation of the processing steps made during the construction of the dataset.
5.5	1012	5	date	yyyymmdd	Last dataset update date	Date when the dataset was last modified.
6					Spatial Reference System	Description of the spatial system used reference in the dataset
6.1	6915 (3301)	105	text	free text	Indirect Reference System	Indirect reference system in which the references to a position are given e.g.

						administrative units, street names.
6.2	6916 (3401)	204			Direct Reference System	The system used to represent space in the data set.
6.2.1	3462	105	list	cl	Name of Geodetic Datum	Name of the geodetic datum.
6.2.2	3408	105	list	cl	Name of Map Projection	Name of the projection or type of the coordinates.
6.2.3	3468	105	list	cl	Name of Vertical Datum	Name of the reference system used for elevation.
7					Distribution	General information about options for obtaining the data set.
7.1	6906	105	text	free text	Distribution Identifier	Identifier by which the distributor knows the dataset.
7.2	2055	105	text	free text	Price	Data about charges for the dataset, including price per nit and discount possibilities.
7.3	2017	105	text	free text	Order process	Data about how to order the dataset, including instructions and expected timescale for the delivery of the dataset.
7.4	6917 (2006)	105	text	free text	Distribution Unit	Data about the partitioning of the dataset (usually it means geographic partitioning e.g. mapsheets).
7.5	3632	105	list	cl	Distribution Media	Media on which the dataset can be received.
7.6	1031	105	list	cl	Distribution Format Name	Formats in which the dataset can be delivered.
8					Metadata Reference	General information about the metadata
8.1	3702	5	date	yyyymmdd	Last metadata update date	The date of latest review of metadata.
8.2	64	105	list element	ISO 639	Metadata Language	Language and character set used for the textual statements in the description of the dataset, identified y codes defined in ISO 639.

4. References

1. Content Standard for Digital Geospatial Metadata
(Federal Geographic Data Committee, April, 1997)
2. Dublin Core Elements
http://www.purl.org/metadata/dublin_core_elements
3. ESMI Search Variables and Metadata
05/21/98, ver. 2.1
4. European Committee for Standardisation, 1995-09-21
Geographic Information – Data description – Metadata
5. Geographic Information – 15046 Part 15: Metadata
(ISO/TC 211/WG, December 22, 1997)
6. Geographical Data Description Directory version 2.1
(MEGRIN Group)

7. Geographic Information – Data Description – Metadata
(CEN/TC 287/WG 2, November, 1996)
8. GILS Element Definition
<http://www.usgs.gov/gils/elements.html>

**Javaslat az egységes közigazgatási térinformatikai metaadat-szolgáltatás
konceptiójára és megvalósítására
MeH, 1998, p. 39.**

9. Kovács, Enikő, Prajczner, T., 1995
Evaluation of Metadata Standards
FÖMI, p.36+82.
10. Prajczner Tamás, Kecskés Zsolt, Székely Iván, 1998
A METATÉR projekt megvalósíthatósági tanulmánya
GeoX, p. 75