THE STANDARDISATION OF OBJECT DEFINITION:

INTERNATIONALLY AND THE DUTCH CASE

Henri Aalders*

Preface

During Theo Bogaerts' carrier, in the 80'ties, he was very much involved in the definition of LIS objects, creating the BOCO classification, finally resulted in a SSVI report [SSVI, May 1982]. Since, standardisation of data definition has been a hot topic in the Netherlands, finally resulting in the Standard NEN 3610 Geo-Information Terrain Model. This standard describes terms, definitions and general rules for the classification and coding of objects related to the earth's surface. The necessity of this standard became apparent after the introduction of the NEN 1878 standard for the transfer of GIS data. NPR 3611, containing application rules using both NEN 3611 and 1878 follows NEN 3610. Due to the development of Internet - resulting in a remote use of database at the user's terminal - standard ways of data definition is internationally a topic on the standardisation agendas of many organisations. These appear both in the field of application - e.g. European projects as MEGRIN/La Clef/GDDD (Topography), GEIXS (Geology/Soils), AVID (Hydrography), CLEAR and MADAME (property) - as well in research applying ontology.
In this contribution for Theo Bogaerts' farewell a review of work in this field is described both from theoretical and practical point of view.

* About the author

Henri Aalders finished his study in Geodesy in 1969 at the Delft University of Technology and worked since in photogrammetry and remote sensing, cartography, land surveying and GIS at the Survey Department of the Ministry of Public Works, the International Institute for Aerial Survey and Earth Sciences (ITC) and was project-leader for the cartographic part of the Dutch Cadastre. Since 1998 he is Associated Professor in GIS-technology at the Department of Geodesy of the Delft University of Technology

working with Theo Bogaerts and since 1995 he Professor at the Katholieke Universiteit Leuven in Belgium in Topography. He was also Dutch representative for the CEN/TC 287 commission on standardisation of 'Geographic Information' and chairperson of its working group on 'Fundamentals'. He is also Netherlands representative for the ICA commission on Standardisation. Also, he is an expert for the Word Bank and the European Union, designing cadastral systems in many central European countries after the political and economic change in 1989.

INTRODUCTION

In the analogue era, concern for spatial data and its distribution were tasks for mapping organisations earning a reputation for producing high quality products based on their spatial characteristics, including their visualisation in printed map form to be used for many different types of problems.
In contrast, from the second half of the twentieth century onwards, with the developments in computer technology, digital cartography and GIS were also developed to capture, store and analyse spatial data, replacing the tedious analogue map making process.
Nowadays, there is a rapid growth of the availability of digital spatial data and a growing need to use it for all kinds of applications in the field of spatial activities. In addition, with the development of today's communication technology, it becomes possible for every user, surfing the Internet, to collect datasets from a variety of sources and different types of application. This requires translation from the original source data into the user's system, to make the received data user understandable and usable.
One can consider the set of available databases as the bases of the web that makes access to the datasets possible for all kind of users.

Users of such datasets encounter specific problems in understanding the information, each dataset contains by at least two aspects:

1. each data set has its own structure defined by the Geo-DBMS (Data Base Management System) that is used for storing the data as well as the application specific and user-defined organisation of the data in the Geo-DBMS. The receiving geo-system should be enabled to understand the organisation of the data in the original dataset. This is done through standards, defined by official national and international standardisation bodies (such as ISO and CEN, where many experts in the field of concern co-operate to design standards) as well as by the OpenGIS Consortium, where many private organisation try to enable the possibility to understand the organisation of the data in the different Geo-DBMSs for the receiving Geo-DBMS. This aspect of data transfer can be referred to as the *geo-structural database definition*. The organisation of geo-objects in an

object-relational or object-oriented database serves the swift access to the data for query by all kind of applications. Especially for large databases such access methods use specific geometric access keys - as Morton, Peano, Hilbert, etc., together with a well-defined spatial clustering method as B-tree, R-tree, etc.;

2. once the dataset is received by the user, the understanding of the data to make it worthwhile information to use together with the user's own information or with information of other datasets is the next problem the user has to overcome. This aspect of data transfer can be referred to as the *semantic object-definition*.

This contribution will deal with the semantic object definition as developed in The Netherlands geometric applications as was used in earlier times in large and medium scale maps.

HISTORY OF SEMANTIC OBJECT DEFINITION IN THE NETHERLANDS

The standardisation of object definition has been a matter study in the Netherlands since the introduction of digital topographic data in the mid-eighties. Before that period, object-definition was done by legends accompanying each map type. The Topographic Service had its own map legend for different scales of maps, the Cadastral Office used a definition for information displayed on property maps and the Survey department of the Ministry of Transport and Traffic control issued its own legend for large scale topographic maps. On top of that many municipality and provincial maps, utility maps and maps for water management and water quality were provided with there own specific legend.

In the digital era however, the basic principle is to make computer programmes understanding the object definition allowing them to understand each other datasets. First attempts were initiated in November 1978 when the BOCO (Bestuurlijke Overlegcommissie Overheidsautomatisering) contracted the working group Uniformering of the Rijkscommissie voor Geodesie to perform a study for the possibilities to transfer digital topographic and other basic data between users. This study was separated into four sub-studies named: classification, quality, data structures and access and conversions. The department of Geodesy of the TU Delft - resulting in a study report [SSVI1982/6] - executed the classification study, chaired by Prof. Dr. Ir. M.J.M. Bogaerts. A result a subsequent study was performed by the foundation SSVI (Stichting Studiecentrum voor Vastgoedinformatie), creating a set of definitions for topographic elements that were used in different datasets [SSVI 1982].

This object definition resulted in a scheme of 400 objects being defined and in time more and more terms were added for objects necessary for different users resulting finally in 700 definition.

With the introduction of the cadastral system LKI (Landmeetkundig en Kartografisch Informatiesysteem) a huge database was created (up to about 50 Gbyte nowadays) containing property and so-called large scale topography[1]. The large-scale topography, known in the Netherlands as the GBKN (Grootschalige Basiskaart van Nederland - large-scale topographic base map of the Netherlands) has been stored in the cadastral database also and used its own data definition. In the beginning of the creation of this database, the data was collected provincially where different data definitions were used. With the introduction of the CKR (Centrale Karteringsraad, central mapping board, later changed in CTB, Commissie Topografische Bestanden – commission for topographic databases - of the RAVI, voorlopige Raad voor de Vastgoedinformatie – pre-council for land-information), this problem was solved but different nomenclature was used from the BOCO definitions. Prof. Dr. Ir. M.J.M. Bogaerts chaired both commissions. The cadastral office published [LKI extern, 1989], containing tables to convert such differences.
The co-ordination Minster for Landinformation, the minister for Housing, Physical Planning and Environment has requested the RAVI to propose a official standard (norm) for the classification of spatial objects in July 1992. In order to speed up the process making a norm a co-operation with the Netherlands Normalisation Institute was created. Prior to this attempt to create an official standard several researches have been performed in the field of cables and pipelines [RAVI 20 and RAVI 24] that formed the criteria for the research to create a standard to select only those objects that were to be transferred. All other objects were considered to be 'company-own', not to be available for transfer.
The standard [NEN 3610], called Geo-information Terrain Model [RAVI 29], a Netherlands standard for: terms and generic rules for the classification and coding of objects related to the earth's surface, was published in July 1995.

[1] In fact the data in this database contain real co-ordinates resulting in a DLM (digital landscape model), so one cannot speak of large scale topography but the term is still often used. It means that the database contains data with the amount of detail related to a large scale topographic map

THEORETICAL ASPECT OF SEMANTIC OBJECT DEFINITION

The main reason for the standardisation of semantic object definition is the use of the same data in different applications by transferring those data between datasets. As such, multiple data use is efficient and necessary because data collection is very expensive and. For the use of the semantics when transferring data between different applications, one has to distinguish between descriptions of objects as they appear in the terrain and descriptions of objects in map representations. In literature, the DLM (Digital Landscape Model) and DCM (Digital Cartographic Model) often indicate these. In the DLM the terrain is described as a set of objects defined by their spatial and non-spatial characteristics. In the DCM objects as obtained in the terrain according to the DLM are defined as they are represented in a graphics (e.g. on maps or drawings) or digital picture in a image or on-screen.
In this study, we will only consider the DLM description of the terrain.

Preparing definitions for topographic objects can be seen a preparing a nomenclature of the objects in order to provide terrain objects with the correct name.
With the inventory of topographic objects appearing in different datasets, similarities were found referring to two aspects: the concept and the object's naming. Objects appearing in two different datasets that fulfil the relations [SSVI 1982]:

$$(\text{name}_1 = \text{name}_2) \quad \text{and} \quad (\text{concept}_1 = \text{concept}_2)$$

are equally transferable, or in other words for these object in both datasets the definition is used. In other case where one of these conditions is not fulfilled standardisation of object definition is required to make data transferable in an understanding way for both the provider and the recipient.

Secondly, after naming the topographic objects, they can be classified into classes with common characteristics. Usually classes are hierarchical organised. Often the object definition and the classification of objects into classes are called classification though they are different aspects.

Thirdly, classes are usually assigned codes, in order to avoid lengthy names to be stored and make it easier for computer programmes to handle the semantic definitions. Coding also serves efficient selection in the dataset. These codes can be seen as a second name for the object. However, users are not obliged to use these codes. Still, users need to use the same names and definitions to allow the transfer and mutual understanding of the definitions.

There are several aspects for the definition of objects appearing in the terrain. Definition may be very generic in terminology e.g.: rural area versus urban area, or very specific as is used in most nomenclatures using a strict taxonomy. Mostly used are definitions using the appearance of the object in the terrain as well as its usage. Classification then can be done according to the appearance or usage of the objects but also to its value, juridical or management status.
Classification also serves the possible aggregation for other applications and if necessary, representations at smaller scales. Map generalisations operations as 'deletion' of unwanted objects for the new application as well as the 'combination' of objects into new ones, also influence the hierarchical structure of a classification system. For instance, a highway may contain the lanes for driving but also bridges and viaduct, direction signs, lampposts, kilometre indications etc., all belonging to the object 'highway'. It is questionable whether all these object should be defined separately and/or also all together as one object. It depends on the application for the user how detailed or combined objects have to be defined.
User may select from the list of defined objects only that are necessary for his/her application. Assuming the hierarchical nature of the classification, by this choice, he determines also the amount of detail in his dataset as well as the potential level of transfer.

The semantic definition of objects involves also the definition of their characteristics. The main characteristic determines the name – and the code - of the objects. Other characteristics may also be used for classification or for selection in other applications. All characteristics are either of qualitative of nature, i.e. a discontinue qualification of the object's characteristic or quantitative of nature, i.e. a measurable qualification of the object's characteristic defined in a continuous scale.

Classification systems maybe open or closed, i.e. closed classification systems contain object names assigned to instances of terrain objects and only those are a are allowed. In open classification systems, there will be always a class containing objects that do not fall under other classes. E.g. a building can be characterised into public buildings (open for all citizens), private buildings (open for only specific citizens) and other built-up constructions (as a bridge, a viaduct, a factory, etc.). Such 'other' classes may contain all kind of instances that cannot be controlled by the classification system and make unified understanding of the dataset difficult.
Other problems occurring in the definition of classes is the fuzziness in usage for different applications and the amount of details that is necessary for different types of applications. Different applications may also use application

specific dominance at least resulting in different classification systems and may be also in different definitions.

ISO REQUIRED ATTRIBUTES FOR DATA DEFINITION

In July 2001 the ISO (International Organisation for Standardisation, Geneva) technical committee ISO/TC211 has issued the final text for CD 19110 (Committee Draft) describing a methodology for data.

Type / Object catalogue element	Object type	Object operation	Attributes	Attribute value	Feature relationship
Name	M	M	M		M
Definition	C		C	O	C
Formal definition		O			
Code	O		O	O	O
Aliases	O				
Object type names		C			
Operation names	O				
Attribute names	O	M			
Value data type			M		
Value measurement unit			O		
Value domain type			O		
Value domain			C		
Label				M	
Relationship names	O				
Inverse relationship					O
Object types included					M
Order indicator					M
Cardinality					O
Affected attributes					O
Relationship constraints					O
Subtype of	O				

M = Mandatory: to be included in the catalogue
O = Optional: may be included in the catalogue
C = Conditional: the condition is stated as a question; in case of positive answer the section should be included in the catalogue

Table 1. Required object catalogue contents according to ISO/TC211 CD 19110.

Definition [ISO/TC 211, 2000]. The ISO/CD 19110 makes distinction between the definition for object types, object attributes and object relationships. Additionally descriptions of object operations may be defined. It is stated that all definitions shall be given in natural language. For all types, a list is given indicating the mandatory, optional and conditional parts of the definitions as indicated in table 1. The domain of the content elements is free text or integer type. In addition, the content is given for the object catalogue itself in the CD 19110, as well as UML formatted model.

THE GEO-INFORMATION TERRAIN MODEL FOR A STRUCTURED SEMANTIC DATA DEFINITION

Preparing the Geo-Information Terrain Model (GITM), finally leading towards the Netherlands Standard NEN 3610, object definitions were used that existed already within the following organisations in the Netherlands: Cadastral Office, Municipalities, Utility companies, Water Control Boards, Netherlands Railway, Provinces, Topographic Service, Department of Physical Planning of the Ministry of Housing, Physical Planning and Environment, Directorate of Environment of the Ministry of Housing, Physical Planning and Environment, Ministry of Defence, Ministry of Transport and Traffic, Ministry of Agriculture, Nature management and Fishery, Society of companies in the geo-information (VNBG), Union of City Planners (BNS) and NOB transportation.

The main aim - to create an application independent, generic, object-based terrain object definition for which the semantics could be used in transferring data - should fulfil the following conditions:

1. only objects and characteristics, as appearing in the terrain, are described, becoming entities and attributes in the GITM;
2. physical characteristics of objects may be abstracted into attributes to describe their appearance;
3. object definition is independent from the applications and technology;
4. object definition must be flexible; i.e. extendible for later developments;
5. classifications within certain sector applications should stay valid;
6. relations between objects are single and mutually level co-ordinated. Users may compose complex objects form these.

Also, objects in such generic, application-independent object definition should have an earth surface related position, to be defined by location, shape and topology, have a recognisable abstraction, resulting in a closed classification system (further details may be created by respective application sectors).
The characteristics of such objects should have a recognisable abstraction level, resulting in a closed classification system. The usage characteristics should be avoided as much as possible because they are linked too much to applications and only terrain appearance should be stressed, while minimal

attributes for linkage to other (more detailed) classification systems should be allowed.

In order to disconnect specific applications and to determine the common datasets, a methodology for object definition was used where terms and aspects play an important role. The term is the building brick of the definition system, while an aspect is a point of concern of a concept. Therefor the definitions contain definitions of entities – being the abstraction of the object's main characteristic (name), attributes – being the definition of the other object's characteristics and other terms.

Object definitions contribute to the solution of identification of common objects in different datasets where the following aspects are of concern:

1. time differences appearing between objects in the different datasets where chances may have occurred to the objects and are registered differently in the respective datasets;
2. using a common dataset in several separate applications each user may add applications dependant information to its dataset. Delivery of updates of the common dataset may cause identification problems for each user possibly loosing the applications specific added value;
3. using general identifications controlled by third parties may cause identification problems for users. E.g.: municipalities control addresses. When a municipality decide to change the house numbering system in a street users may misunderstand the changes when identifying house according to the street name and house numbers.

Basic concept to the definition of the semantics of terrain objects is the object-oriented approach with a recognisable delineation of objects in the terrain, making entities with attributes classifiable. The level of detail of those objects was determined by their geometric discernibility in the terrain. Elements in the reality are called objects, while their descriptions in the model are called entities (see table 2, [Aalders, 1989]).

<table>
<tr><th colspan="2">REALITY</th></tr>
<tr><td>Object is a phenomenon in reality which exists independently of others and is separately recognisable</td><td>Feature is a descriptive characteristic of an object</td></tr>
<tr><th colspan="2">MODEL OF THE REALITY / CONCEPTUAL SCHEMA</th></tr>
<tr><td>Entity is an abstraction of an object instances with the same characteristics</td><td>Attribute is a descriptive characteristic of an entity with a domain for the respective values</td></tr>
</table>

Table 2. Relations between definition concepts.

In order to enable as many users as possible to apply a structured semantic data definition, The level of abstraction is very generic for the basic semantic object definition. Each sector may extend the generic semantic data definition with its specific more detailed application oriented definitions. In turn, users within a specific sector can detail the sector oriented semantic definitions for its own application (see fig. 1), resulting in a extendible structured semantic data definition according to the user requirements.

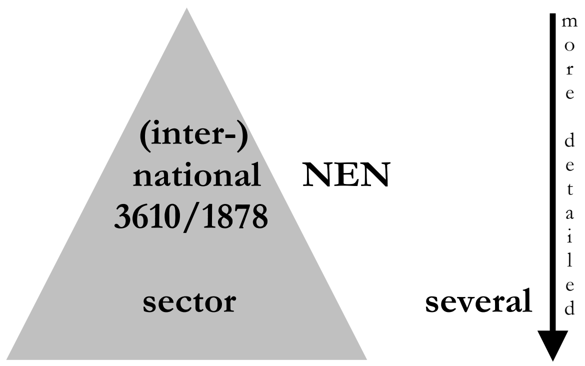

Figure 1. Hierarchy of semantic data definition.

According to the GITM the set of spatial objects covering the earth's surface is subdivided into real and virtual objects that may overlap (see fig. 2).

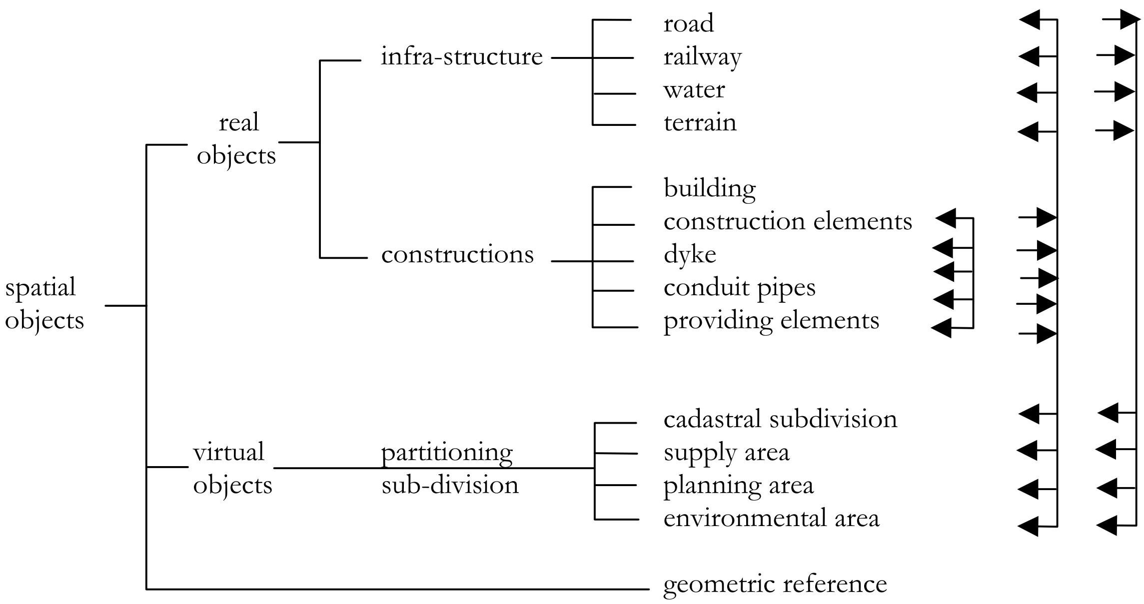

Figure 2. Hierarchical subdivisions and adjacent relations.

The real objects are subdivided into infra-structural objects being roads, railways, water and terrain and constructions as buildings, construction elements, dykes, conduit pipes and providing elements. The virtual objects are subdivided into partitioning sub-divisions as cadastral lots, service areas (e.g. for utility companies, water control boards, municipalities, etc.), planning areas (for physical planning) and environmental control areas. As such these create a hierarchical subdivision.

On top, entities show mutual level co-ordinated relations, e.g. building or constructions may appear in ranging areas or constructions may overlap sub-divisions. Features are of the following types:

1. identification features in order to identify the object instance from any other instance; both unique codes or location can be used for identification ;descriptive features to add more detail to the object. One can describe the usage, nature, purpose, planning permission, right, access, paving, material, product, voltage pressure class, profile or status. No one of the descriptive features are required, but all may be used. The function of an object is not defined. An example may show this. A church (nature) may be used as dwelling (usage) but is intended as a car park (purpose). Planning permission may alter the function in time. Right has a twofold function: it may be attached to the object itself or to the data about the object (making it metadata about the object);

Entity \ attributes	identification	location	infra-structure	building	construction	reference	virtual area	usage	nature	purpose	planning permission	legal	access	paving	material	product	voltage / pressure	profile	status	geometric	data management
Road	x	x	x						x			x	x						x	x	x
Railroad	x	x	x						x			x							x	x	x
Water	x	x	x						x			x	x						x	x	x
Terrain	x	x							x			x	x	x	x				x	x	x
Building	x	x		x				x	x	x	x	x	x		x				x	x	x
Construction element	x	x		x					x			x			x				x	x	x
Dyke	x	x	x						x			x	x	x	x				x	x	x
Conduit pipe	x	x	x						x			x			x				x	x	x
Providing element	x	x			x				x			x		x	x	x	x	x	x	x	x
Geometric reference	x	x				x						x			x					x	x
Cadastral subdivision	x	x					x					x							x	x	x
Supply area	x	x					x					x							x	x	x
Planning area	x	x					x			x									x	x	x
Environmental area	x	x					x			x									x	x	x

Table 3. Relationships between entities and attributes.

Entity \ Types of depiction	Point	Line	Areal
+ROAD	x	x	x
RAILROAD	x	x	x
WATER	x	x	x
TERRAIN	x	x	x
BUILDING	x	x	x
CONSTRUCTION ELEMENT	x	x	x
DYKE	x	x	x
CONDUIT PIPE	x	x	
PROVIDING ELEMENT	x	x	x
GEOMETRIC REFERENCE	x		
CASDASTRAL SUBDIVISION	x		x
SUPPLY AREA	x		x
PLANNING AREA	x		x
ENVIRONMENTAL AREA	x		x

Table 4. Mathematical types of depiction by entity.

2. geometric describing the object's location (defined by co-ordinates in a national reference system or by addresses), shape (describing the interpolation methods to represent the object) and topology (for describing neighbourhood relations between the objects);
3. meta descriptions containing information about other features such as data management, quality descriptions and conceptual schemas about the objects, containing entity and attribute descriptions with their domains and relationships as well the mathematical depictions of the entities (see table 3 and 4).

Metadata model

In order to ensure definitions are application independent a methodological definition of organisational data is used according to the meta model in fig. 3.

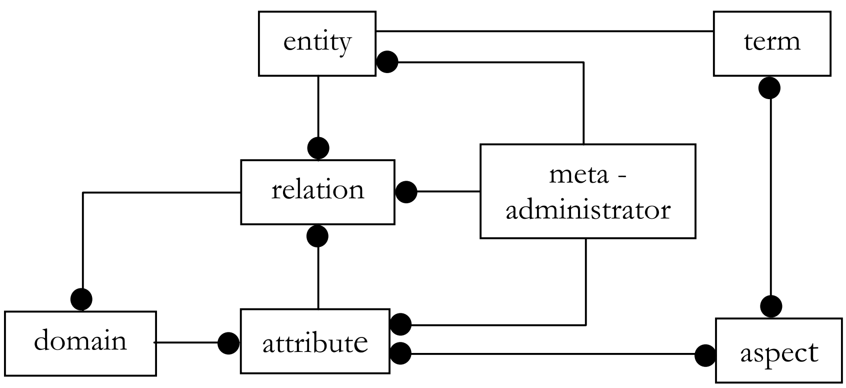

Figure 3. Metadata model for defining data.

Essential in this model is the use of terms and concepts to:
1. establish a framework for data definition;
2. fostering an unambiguous understanding of the meaning;
3. aid for establishing definitions;
4. support data management activities.

CONCLUSIONS

In general, there is a common part in the different applications of geographic data that requires the multiple use of these data and so the common data definition to make the data understandable and meaningful to all users. This is a difficult process since most data definitions are application-dependant. In the Netherlands, an effort is made to design an system for data definition that is application-independent resulting in a generic data definition that is extendable for different sectors and applications. This hierarchic approach proved to be very successful since utility companies, water control boards and city planners have designed their more detailed sector objects definitions following the structure laid down in the generic data definition.

References

Aalders, 1989, Gegevenstheorie voor ruimtelijk informatie, H.J.G.L. Aalders, Dictaat behorende bij het college Vastgoedinformatie II-ge61. Versie 1991, TU Delft, Faculteit der Geodesie.

ISO/TC 211, 2000.

CD 19110, Geographic Information - Feature cataloguing methodology, N938.

ISO/TC 211/WG 3/Editing committee CD 19110, July 2000.

LKI - extern, 1989, Handboek LKI - extern, technische aspecten. Dienst van het Kadaster en de Openbare Registers. Uitgave: directie geodesie, november 1989.

NNI 1995, NEN 3610 Terreinmodel vastgoed, Delft, NNI. UDC 001.4.69.001.4:681.3:003.62.

RAVI 20, Kabel- en Leidinggegevens voor elkaar, deel 1 Classificatie voor uitwisseling, RAVI rapport 20, Apeldoorn, The Netherlands.

RAVI 24, Kabel- en Leidinggegevens voor elkaar, deel 2 Uitwisselingformaat gedefinieerd, RAVI rapport 24, Apeldoorn, The Netherlands.

RAVI 29, Terreinmodel vastgoed. Een algemene classificatie voor de gegevens over de aan het aardoppervlak gerelateerde gegevens, RAVI rapport 29, Apeldoorn1992, The Netherlands.

SSVI 1982/6, Classificatie topografische elementen; verantwoording gebruikersrapport 1e uitgave, Publicatiereeks van de Stichting Studiecentrum voor Vastgoedinformatie, Delft, October 1982.

SSVI 1982, Classificatie topografische elementen; gebruikersrapport 1e uitgave, Publicatie-reeks van de Stichting Studiecentrum voor Vastgoedinformatie, Delft, May 1982.