

Definition and Standardization of Data Elements' Attributes in Land and Resources Management

Lei Cong, Yongxia Yang, Dehai Zhu, Min Yin, Jianlin Li

► **To cite this version:**

Lei Cong, Yongxia Yang, Dehai Zhu, Min Yin, Jianlin Li. Definition and Standardization of Data Elements' Attributes in Land and Resources Management. Daoliang Li; Yingyi Chen. 6th Computer and Computing Technologies in Agriculture (CCTA), Oct 2012, Zhangjiajie, China. Springer, IFIP Advances in Information and Communication Technology, AICT-393 (Part II), pp.309-320, 2013, Computer and Computing Technologies in Agriculture VI. <10.1007/978-3-642-36137-1_37>. <hal-01348246>

HAL Id: hal-01348246

<https://hal.inria.fr/hal-01348246>

Submitted on 22 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Definition and Standardization of Data Elements' Attributes in Land and Resources Management

Lei Cong¹, Yongxia Yang^{1*}, Dehai Zhu¹, Min Yin², Jianlin Li²

¹College of Information and Electrical Engineering, China Agricultural University,
Beijing 100083, China

² Beijing Municipal Bureau of Land and Resources, Beijing 100013, China

Abstract. The standardization of data element in land and resources management is an important component of the standardization of information technology in the field. It's an essential method to guarantee the interconnection, intercommunication and interoperability, and achieve the information sharing. This study is a case study on Beijing Municipal Land and Resources Management, especially cadastral management and land use management. Considering of the particularity of data in land and resources management from the aspects of language, expression, area, data and application, we study and develop many rules to describe the attributes of data element based on ISO/IEC 11179 and GB/T 19488. This paper shows the specific definition and standardization of data elements' attributes in our study. The whole study can direct the other related working in land and resources management, and give the ideas to carry out the study on the standardization of data element in the other fields.

Keywords: Standardization, Data Element, land and resources, attribute

1 Introduction

The standardization of data element is an important component of the standardization of information technology. It guarantees the development and running of information system. The success of the system applications is on the basis of it.

*Corresponding author; E-mail: yongxia.yang@qq.com

The standardization of data element is defined as the process of developing and implementing uniform standards, which includes the data elements' general rules, definition, description, classification, expression and registration.

Its importance can be broadly divided into five areas:

- Standardizing and unifying the collection and application of data;
- Building the basis of data sharing for different systems with multiple modes of data;
- Achieving data management in the level of metadata and making the separation of software and data, so the development of software is more flexible.
- Providing unified standard of data exchanging in business internal areas or between internal and external areas;
- Being the basis of building a unified, integrated and efficient data model in business.

Land and resources management is a multi-disciplinary integrated system relating to land, geography, minerals, etc. Business management needs cooperation and mutual penetration in the area, while it actually lacks of common data models and standards. The present situation causes many problems, such as one word with several equivalents of different meanings and one word with multiple equivalents of the same meaning. The former can make users understand the word with ambiguity like Digital Gap; the later can cause data redundancy and update data inconsistently, and further, it goes ill with the integration of information system, which is called Information Island. These problems add the confusion of the system, and limit data exchanging and sharing. In summary, the standardization of data element in land and resources management is a vital foundation in this field, especially for the goals of "A Map" and "Golden Land Project".

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) has released "ISO/IEC 11179-Information technology-Specification and standardization of data elements" in 1999, and updated it to "Information technology--Metadata registries (MDR)" in 2004. It's an international standard for defining, describing, standardizing and registering data element or metadata for an organization. For the standard's importance, China National Institute of Standardization translated it into "GB/T 18391-Information technology-Specification and standardization of data elements" in 2002 and "GB/T 18391- Information technology--Metadata registries (MDR)" in 2009 for Chinese

organizations' use. In addition, CNIS also released "GB/T 19488 Data Elements for e-government" for better understanding and more practical, which refines and embodies the contents of principle and theories in GB/T 18391.

This study is on the standardization of data element in Beijing Municipal Land and Resources Management, especially cadastral management and land use management. The above standards are the basis of the study, whereas, because of language, expression, area, data and application, there are still some problems that need further research. In this paper, we record and summarize the issues and our solutions that focus on the standardization of data element's attributes.

2 Definition of Data Element' Attributes

Data element is described by a set of attributes. The basic attributes of data element can be organized into five categories, including a total of 22 attributes as following^[3].

- **Identifying:** Name (mandatory), Identifier (conditional), Version (conditional), Registration Authority(conditional), Synonymous Name (optional), Context (conditional)
- **Definitional:** Definition (mandatory)
- **Relational:** Classification Scheme (optional), Keywords (optional), Related Data Reference (optional), Type of Relationship (conditional)
- **Representational:** Representation Category (mandatory), Form of Representation (mandatory), Data Type of Data Element Values (mandatory), Maximum Size of Data Element Values(mandatory), Minimum Size of Data Element Values (mandatory), Layout of Representation(conditional), Permissible Data Element Values (mandatory)
- **Administrative:** Responsible Organization (optional), Registration Status (conditional), Submitting Organization (optional), Comments (optional)

The above attributes are just recommended in the standards. In the actual study, they can be added, deleted and modified with many factors. First, we can identify priorities on attributes considering of the application of data elements. For example, if the standardization of data element is for the internal needs of an organization, the attributes of Administrative should be completed and detailed for future use in business management; while the standardization of data element is for data sharing

among different organizations, the attributes of Administrative may have no much sense. Second, because of the difference of languages and areas, we can add some specialized attributes adapt to the current environment, such as Chinese name, English name, Chinese spelling. In addition, we can also merge and resolve some attributes in order to make them clearer and more concise.

Considering of the actual need in business areas of Beijing Municipal Land and Resources Management, we determine the following factors to choose the attributes with the aim of less attributes but also a clear expression.

- 1) Requires attributes, such as name, identifier, definition, data type, etc.
- 2) Attributes for management, such as version, registration authority, submitting organization.
- 3) Attributes for inquiring and reading by user, such as keywords, synonymous name.

By the above factors, we define the attributes which are used to describe data elements in land and resources management (see table 1).

Table1. Attributes of Data Element in Land and Resources Management

Category	Attribute	Obligation
Identifying	Chinese Name	M
	Chinese Spelling	M
	Identifier	M
	Version	M
	Registration Authority	M
	Synonymous Name	O
	Context	C
Definitional	Definition	M
	Object Class Term	M
	Property Term	M
Relational	Keywords	O
	Relationship	O
	Representation Term	M
Representational	Data Type	M
	Data Format	M
	Value Domain	M
	Unit of Measurement	O

	Status	M
Administrative	Submitting Organization	O
	Date of Approval	O
	Comments	O

3 Standardization of Data Element's Attributes

This section shows the key points of the standardization of data element's attributes in our study fields, which are cadastral Management and land use management. We study and develop many proprietary and practical rules to solve the problems caused by language, expression, area, data and application. This study complements and improves the processes and methods for the standardization of data element's attributes in specific field, which are abstract and conceptual in the standards.

3.1 Chinese Name

Chinese name is the name of data element with single or multiple Chinese words. It's generally composed of object class term, property term, representation term and sometimes also qualifier term, which are related to the data model of data element. In order to make data element's concept, classification and representation more clearly, we add the first three terms in the attributes of data element, and the methods to identify them are showed in the ensuring chapters. In this part, the development of naming rules is the key, even in our whole research. A standardized name is the basis to solve "Digital Gap" and "Information Island" in the system.

Besides the basic naming rules in the standards as following ^{[5] [7]}:

Semantic rules:

- *One and only one object class term shall be present.*
- *One and only one property term shall be present.*
- *One and only one representation term shall be present.*
- *Qualifiers will be added as needed to make the name unique within a specified context. The order of the qualifier terms is not significant. Qualifier terms are optional.*

EXAMPLE

宗地法人代表姓名	Name of Parcel Legal Representative	
法人代表	Legal Representative	(object class term)
姓名	Name	(property term)
名称	Name	(representation term)
宗地	Parcel	(Qualifier term)

NOTE “姓名”and “名称” are used in different situations in Chinese, while they are a same word in English. We explain it later in the article.

Syntactic rules:

- *The object class term shall occupy the first (leftmost) position in the name.*
- *The property term shall occupy the next position.*
- *The representation term shall occupy the last position. If any word in the representation term is redundant with any word in the property term, one occurrence will be deleted.*

EXAMPLE

宗地法人代表姓名 Name of Parcel Legal Representative

The property term is “姓名”, and the representation term is “名称”. They make the meaning repeated, so the representation term is deleted in the name.

- *Qualifier terms shall be prefixed to the part qualified. The order of qualifiers shall not be used to differentiate names.*

EXAMPLE

“变更前宗地地籍号” and “宗地变更前地籍号” both mean parcel number before changing. The position of Qualifier term “变更前 (Before Changing)” have no effect.

Uniqueness rule:

All names in each language shall be unique within one context.

We also develop the additional lexical rules to adapt to the particularity of this field, considering of the use of Chinese words.

Lexical rules:

- *If the object class of data element is human, its name should use “姓名” to represent. While the object class of data element is thing or organization, its name should use “名称” to represent.*

EXAMPLE

宗地法人代表姓名 Name of Parcel Legal Representative

宗地权属单位名称 Name of Parcel Ownership Unit

Comparing to a word of name in English, it seems complex in Chinese. More

than one word can correspond to one meaning but also have a little difference. So rules are much needed to prevent the confusion.

- *International abbreviations and acronyms are allowed to use in the name.*

EXAMPLE

DEM Digital Elevation Model
DTM Digital Terrain Model

Although they are the abbreviations and acronyms of English words, almost everyone knows and consistently uses them in this field. On the contrary, if they are translated to Chinese, someone may have doubts about them and the name of data element would be too long.

- *Arabic numerals can be used in the name to distinguish data elements.*

EXAMPLE

线状地物权属单位 1 Name of Linear Feature Ownership Unit
线状地物权属单位 2 Name of Linear Feature other Ownership Unit

This is a means to make the name of data element simple and brief, but it's on the premise of a clear definition.

These are all of rules we have studied and developed for naming data elements. As there aren't unified English name for data elements in this field, we don't consider English name as one attribute of data element.

3.2 Chinese Spelling

Chinese spelling is a specific attribute for Chinese data element, which is composed of the spellings of Chinese characters in the name of data element. All the letters should be lowercase, and connect with a hyphen”-”.

There are two points needed to pay attention to:

NOTE 1:

“u” and “v” should be distinguished properly and used in the right place, such as “吕(lv)” and “区(qu)”.

NOTE 2:

If the Chinese name of data element with Arabic numerals, the Chinese spelling should retain the Arabic numerals.

For example, 线 状 地 物 权 属 单 位
1(xian-zhuang-di-wu-quan-shu-dan-wei-ming-cheng-1).

3.3 Identifier

Each data element should have an identifier, which is defined by the registration authority. This process includes the study of classification of data elements. There we suggest that the authority just provides the idea of classification instead of changeless classification number, because the standardization of data elements is an increasing long-term work.

Since there are so many horizontal and vertical business areas in land and resources management, we use the method of classification combining surface and line. First, we divide the field of land and resources management to three areas of land, resource and geological environment by the means of surface classification. Second, we use the combinative method to get business class and business sub-class. After that, the classification can be refined by the object class that data elements belong to. Fig.1 shows the classification result of land and resources management.

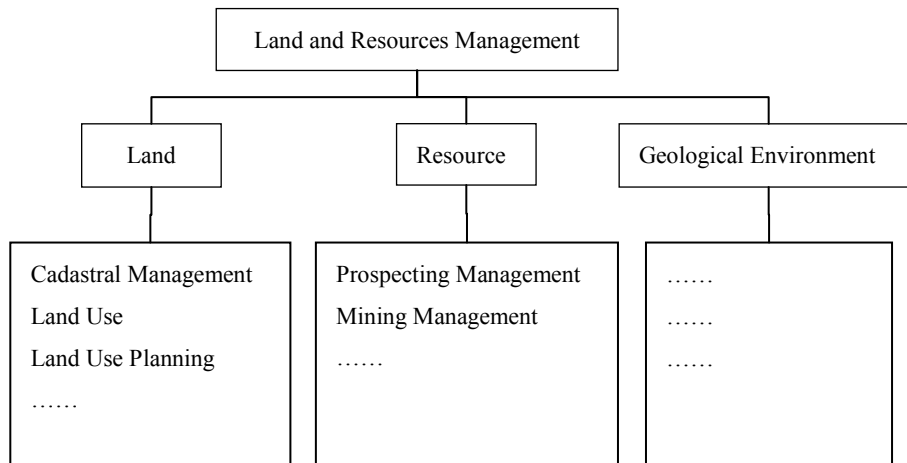


Fig.1. The classification result of land and resources management

Then we can define identifier of data element based on the above result. Fig.2 shows the frame of identifier.

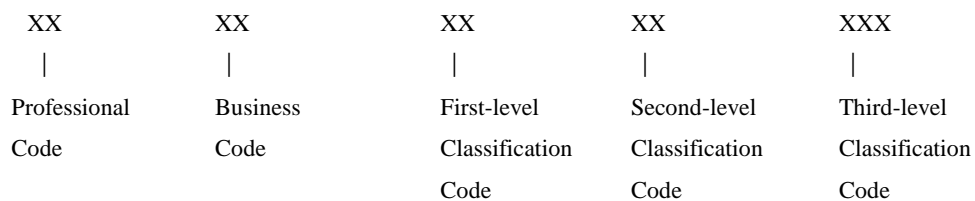


Fig.2. The frame of identifier in land and resources management

1) Professional Code is the code of three professional areas in land and resources management, land, resource and geological environment. It is represented by two letters.

2) Business Code is the code of business types. It is represented by two numbers.

3) First-level Classification Code is about the object class that data elements belong to.

4) Second-level Classification Code can refine the object class in first-level classification.

5) Third-level Classification Code is the sequential number.

EXAMPLE

宗地地籍号 cadastral number of parcel

Identifier TD020101001

Professional Code: TD

Business Code: 02

First-level Classification Code: 01

Second-level Classification Code: 01

Third-level Classification Code: 001

3.4 Synonymous Name and Context

Synonymous name refers to the different names in the different application environment for one data element. A data element may have none, one or more synonymous name, which are separated by semicolon. The collection of synonymous name can further solve “Digital Gap” and “Information Island” in the system, considering of the traditional habits in the business management.

As synonymous name’s corresponding, context records the application environment and process generating and using the data element. A data element may have none, one or more context, which are separated by semicolon, that’s decided by the quantity of synonymous names.

3.5 Definition

Definition is the statements used to express the essential characteristics of data

element. It can distinguish data element from all other ones. We should make them brief and accurate. Besides, definitions should be unique in a field and use singular form. Reading various terminology standards for reference is the best way to make definition.

3.6 Object Class Term

Object class term can show the objects and concepts that data elements belong to, and represent an activity or object in a context. The extraction of object class term is a problem in our study, because it usually confuses with qualifier term. To solve this problem, we introduce the thought of modeling, such as Entity Relationship Diagram (ERD).

an object class is a set of ideas, abstractions or things in the real world that are identified with explicit boundaries and meaning, and whose properties and behavior follow the same rules. It corresponds to the entity or relation in the data conceptual model. Through rebuilding the ERD, We can sort out all the entities and relations in the whole business areas. The results are the object class and their attributes are data elements.

In order to prevent object class term confused with qualifier term, the entities and relations must be minimal. The object class that data element directly belongs to, which isn't able to be subdivided, generates the object class term of data element, while the object class at the higher level may be the qualifier term. Sometimes the object class is too small to present in the ERD, so it uses the object class at the higher level instead of the direct one.

3.7 Property Term

A property is a characteristic common to all members of an object class. Each property has a name. Property term represents the characteristics to distinguish the object classes of data elements significantly. It generally corresponds to the attributes in the ERD, such as address, name, price, etc.

3.8 Representation Term

Representation term describes the form of the set of valid values of an administered item which includes representation. Often, the representation term may be redundant with part of the property term. When this occurs, one term or part of one term may be eliminated in a structured name. This can be established as a naming rule.

Representation term is helpful to achieve the functions of data element naming, analysis, analog and inquiry. The international used representation terms include date, code, name, sum and so on. They can also be added by the demand in the study field.

3.9 Relationship

Relationship describes the connection between the present data element and other ones. We can know the deep-rooted network of relationship in this business area by the attribute. It includes the following basic relationships: derive-from, compose-of, replace-of, link-with^[7].

EXAMPLE

Parcel compose-of Boundary lines

3.10 Data Type and Date Format

Data type is the type of data element value. The basic data type includes character, number, Date, Date Time, Boolean and binary-type. Because of the particularity of data in land and resources management, we should define the additional data types for spatial data. Based on the standards of OPENGIS, we add seven data types of spatial data, which are point, linestring, polygon, multipoint, multilinestring, multipolygon and geometrycollection.

Data format orders the format requirement for data elements in a business perspective. It combines three basic attributes which are data type of data element values, maximum size of data element values and minimum size of data element values. There are the exact basic requirements in the existing standards. Besides them, we rule that the data format of spatial data is represented by “O”, and the data format of binary data is represented by “B”.

3.11 Value Domain and Unit of Measurement

Valued domain shows the allowable value set of data element. It's given by four ways as following ^[7]:

- 1) By name, pointing out the name of valued domain directly.
- 2) By reference.
- 3) By list, representing every possible value with its instance or meaning.
- 4) By rules indirectly.

Unit of measurement should refer to the related international standards. It needs to be represented in Chinese to avoid users' not understanding the abbreviations.

3.12 Status

Status shows the current state of data element for registration during its whole lifetime. We define four types of status for data element in land and resources management.

- 1) Draft: the relevant departments and units can give widely comments and suggestions during this stage.
- 2) Trial: the result can be used for a try after a certain technical procedures. The feedbacks in this stage are some practical advices.
- 3) Implementation: the content of data element has become the standards to be complied with.
- 4) Abolished: the content of data element would be deleted from the standards.

4 Discussion and Conclusion

This study is based on the theories and concepts provided by the international standards about the standardization of data element; in addition, we consider of the particularity of data in land and resources management from the aspects of language, expression, area, data and application. The proper attributes of data element in this field are defined and the specific rules to describe them are developed in our study. They can direct the other related working in land and resources management, and give the ideas to carry out the study on the standardization of data element in the other fields.

Although the study has been basically completed, there are some uncertain issues needed for further study.

- 1) Chinese name is too long. That's because of the semantic rules. The rules make the name standardized, but also different from the previous used name. Can we add an attribute called "short title" to record the previous used name?
- 2) Synonymous name and context are two corresponding attributes. Context shows the synonymous name's application environment. How to show the application environment of Chinese name?
- 3) We define the data type and data format of spatial data. But they can be only used for small range. There should be an international standard to define them.

By this study, we have achieved more than 400 data elements' standardization in the areas of cadastral management and land use management. We complete and improve the whole process and rules during the case study. Besides, our research results also include a work guideline to direct the related work in the future. Further, we are making an effort to develop it up to a local standard in Beijing. For the unified management of the standardized data element in the future, a management information system is needed. It ensures the interoperability of data elements in the different departments and business, and achieves data sharing successfully. So the design, development and maintenance of the MIS will be the important work.

This study is only about two areas of cadastral management and land use management. There may be still many problems in the case study of other fields. So the results of this study needs to be improved constantly.

Acknowledgements. This research was funded by China Agricultural University Young Teacher Innovative Project "study on the standardization of data element in Beijing Municipal Land and Resources Management". Young Teacher directed and contributed to the whole study. The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

References

1. ISO/IEC 11179-1 Information technology—Metadata registries (MDR)—Part 1: Framework[S](2004)
2. ISO/IEC 11179-2 Information technology—Metadata registries (MDR)—Part 2: Classification[S](2005)
3. ISO/IEC 11179-2 Information technology—Metadata registries (MDR)—Part 3: Registry metamodel and basic attributes[S](2003)
4. ISO/IEC 11179-2 Information technology—Metadata registries (MDR)—Part 4: Formulation of data definitions[S](2004)
5. ISO/IEC 11179-2 Information technology—Metadata registries (MDR)—Part 5: Naming and identification principles[S](2005)
6. ISO/IEC 11179-2 Information technology—Metadata registries (MDR)—Part 6: Registration [S](2004)
7. GB/T 19488 Data Elements for e-government[S](2004)
8. Wenfeng Liu, Feng Bao: Representation Method of Intelligent Transport Systems(ITS) Data Elements. Transport Standardization[J](2011)
9. J. Nogueras-Iso, F.J. Zarazaga-Soria, J. Lacasta, R. B éjar, P.R. Muro-Medrano: Metadata standard interoperability: application in the geographic information domain. Computers, Environment and Urban Systems[J] (2004)
10. Yongtao Liu: Study on the Standardized model of data element. Science & Technology Information[J] (2010)
11. Kai M. Hüner, Boris Otto, Hubert Österle: Collaborative management of business metadata. International Journal of Information Management[J] (2011)
12. Yao Lin: Traffic Scientific Data Design Based on Data Elements Technology. Science Technology and Engineering[J] (2011)
13. Bin Wu: Study on ecological metadata and its standard. Journal of Central South University of Forestry & Technology[J] (2010)
14. Tschangho John Kim: Metadata for geo-spatial data sharing: A comparative analysis[J] ,1999
15. Yi Ouyang, Li Yao, Changhua Dai: Data Element Distill Method on Demand-oriented of Information System. Computer Engineering[J] (2006)
16. Hong Wei, Jianfang Zhang: The Definition and Application of Data Element in E-government Affair Standards System. Information Technology & Standardization[J] (2004)