Research on Standard Digitization Technology based on Knowledge Graph and Semantic Data Dictionary

Chengcheng Wang and Chunxi Wang
Instrument Technology and Economy Research Institution, Guanganmenwai 397A, Beijing, China

Keywords: Standard Digitization, Semantics, Knowledge Graph, Data Dictionary, Interoperation.

Abstract: Standard digitization is the inevitable trend of standard development. IEC, ISO and other international standardization organizations have laid out their layout in the field of standard digitization and built a hierarchical model of machine-readable standards. National standards in the field of industrial process automation can be divided into basic standards (terms and definitions, symbols, classification, etc.), the method of standard (test, procedures, guidelines and interface, etc.) and product standard (product, function, service, data, etc.), digital requirements also exist in application scenarios such as guidance, testing, and certification. Based on the analysis of key technologies such as machine-readable, knowledge graph, semantic information model, semantic interoperation and semantic data dictionary, a standard correlation model with pressure meter as the core is preliminarily constructed in this paper, which provides technical support for the subsequent standard digital transformation.

1 INTRODUCTION

As a new round of technological revolution and industrial revolution worldwide, digital transformation drives production mode, life style and the way of governance profound changes, and a comprehensive, profound and revolutionary impact on economic growth, standardization as a guide and standardize the key elements of the national economy development in our country, plays an important role in the wave of the digital transformation. Although the object and scope of standardization are different, there is a demand for digital transformation.

At present, the three major international standardization organizations (ISO, IEC, ITU), European Standardization Organization (CEN/CENELEC), as well as the United States, Germany, Russia and so on have started the process of standard digital transformation. IEC's MSB (Marketing Strategy Committee) white paper "semantic Interoperability: Challenges in the Digital Transformation" has been published; SMB/SG12 (Digital Transformation Strategy Group) was set up to work on digital work, machine-readable standards, semantic interoperability, systems approach, etc. The database type standard platform has been established, which can formulate, publish, maintain and download IEC international standards online. In industrial process automation, electric power and other fields, the establishment of machine-readable standards has been carried out (Wang, 2021, Wang, 2021, Cao, 2016, Li, 2020, Lu, 2020).

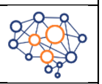
In 2018, ISO established the SAG/MRS (Machine Readable Standards Strategy Advisory Group), published the machine-readable Standards implementation Roadmap, which was incorporated into the ISO 2030 Strategy. It has also established an ISO international standard online browsing platform, which can retrieve symbols, codes, terms, definitions, etc (Chen, 2021, Huang, 2021, Chen, 2020, Kou, 2019).

2 THE PROGRESS AND SIGNIFICANCE OF STANDARD DIGITIZATION

Based on the significance of digital transformation of standards, ISO, IEC and regional and national standards organizations have carried out active research on this topic. IEC and ISO have agreed on a hierarchical model for machine-readable standards (see Table 1).

Wang, C. and Wang, C.
Research on Standard Digitization Technology based on Knowledge Graph and Semantic Data Dictionary.
DOI: 10.5220/0011180000003440
In *Proceedings of the International Conference on Big Data Economy and Digital Management (BDEDM 2022)*, pages 394-399
ISBN: 978-989-758-593-7

Table 1: Machine readable standard classification model,

Classification	***Level 0***	***Level 1***	***Level 2***	***Level 3***	***Level 4***
Name	Paper	Open digital Format	Machine readable document	Machine readable and executable content	Machine interpretable content
Icon					
Format	Paper, TXT	PDF	XML	-	-
Concept	Traditional text format	Display and search related content.	Contains standard text-structured content, with software that recognizes file structure and performs basic operations.	Standard content with semantics can be selectively accessed depending on the application scenario, and application program interfaces (API) can be used to perform more complex operations on standard content.	Machines can perform or parse standard content in more complex ways. The standard contains an information model representing the standard content and the relationship between the elements, which can achieve breakpoint-free and unambiguous data flow.

ISO's SAG-MRS also recommends the unified definition of level 3 and Level 4 Standards as ISO SMART (Standards Machine Applicable, Readable and Transferable), that is, the machine can be available, readable and resolvable without the participation of personnel (Sabbir, 2020, James, 2020). Therefore, based on the classification of IEC and ISO for standard digitization, we can regard standard digitization as the primary stage of "human readable" and the advanced stage of "machine readable".

Primary stage (0 and 1) is mainly the realization of digital reading and retrieval of standard, namely through the analysis of demand for standard in digital environment, help users to timely, accurately and quickly find relevant standards, solve the problem which hard to find the standard. At present, several standard resource retrieval platforms have been developed in China, which can basically meet the query and retrieval of keywords. However, most of the technical content of the standard is read in the way of documents, without a complete analysis. The main reasons include:

- Data protection and copyright ownership of source documents;
- The technical level of standard content analysis is insufficient;
- The mutual application of standards and the timeliness of version update iteration.

Therefore, the digital transformation of standards must be jointly participated by standardization engineers and data engineers, otherwise the digital standard intelligence stays in the stage of circulation and storage, unable to play the role of application on the level of "human readable", more difficult to achieve the function of "machine readable".

The advanced stages (levels 2 to 4) focus on implementing the digital application of the standard. We analyse the digital application scenarios of standards from the types of standards. Generally, standards in China can be divided into basic standards (terms, definitions, symbols, classification, etc.), method standards (tests, procedures, guidelines, interfaces, etc.) and product standards (products, functions, services, data, etc.). Compared with basic standards, it is more meaningful to transform method standards and product standards into machine-readable standards. The main application scenarios of the two standards vary with different application industries and can be generally summarized as guidance, testing and certification.

2.1 Guidance

Which mainly refers to the process in which digital standards guide and standardize activities within the scope of application. Take the manufacturing industry as an example, digital standards can provide standardized design guidance, manufacturing requirements and development methods in the process of product design, manufacturing and development, and cooperate with the development of intelligent manufacturing to realize the intelligent design, manufacturing and development process, such as the self-organization of equipment functions.

2.2 Test

Which include the test from enterprise, as well as the third party, digital standard can support the rapid test plan formulation, the rapid generation of test data, test results of quick judgment and rapid formation of test report, most of all, eliminate the inconformity of understanding of standard terms in the process of test, ensure the consistency of different main body to carry out the test results. Under the cooperation of information tools, the openness and transparency of the testing process will be improved to the maximum extent to enhance the effect of the implementation of standards.

2.3 Certification

As a transmission process of trust, in this paper not only refers to the certification activities of products, systems and capabilities carried out by third-party certification bodies, but also includes enterprises' self-evaluation of their standards compliance. On the basis of the standard digital, can pass, the technical requirement of the standard quantitative into different indicators, enterprise control system related indexes, the submission and upload data and supporting materials, the system through the data model and standard algorithm, the input of product data, such as enterprise information necessary parameters for the item scores and comprehensive scores. Only to the enterprise system to upload the material authenticity review unit manual audit, again according to the system automatically score, form the comprehensive score of jitc or standard products conform to the degree of hierarchy, rather than simply to meet and do not conform to the evaluation conclusion, on the other hand, also can be realized based on the standard digital traceability of the products or services.

3 REQUIREMENTS ANALYSIS OF STANDARD DIGITAL

Among the standard types with digital requirements, product standard, protocol/interface standard, test and evaluation method standard, operation and installation method and other standards account for more than 80%.

The digitization of product standard can support the retrieval and judgment of standard applicability, reference relationship between standards and correlation analysis between technical content.

The digitalization of protocol/interface standards can greatly reduce the workload of communication protocol development and support the rapid testing of protocol consistency and interoperability.

The digitalization of test standards can clarify the applicability of standards, and realize the guidance of test evaluation and validity verification.

Although the requirement for digitalization of different types of standards are different, the significance is clear. From the perspective of technology path, XML language is one of the optional technology paths from level 2 to level 4 to realize standard digital transformation. However, XML language is essentially a standard for data storage, exchange and expression, which is used to mark data and define data types. The semantics of data still need to be defined in the process of digitalization. Based on the above analysis, the semantics of the data can be used "standard model" or "product model", the so-called "standard model" in standard elements as the core, such as product standards for technical requirements, test methods, inspection rules, in order to establish the semantic model structure, the type test evaluation standard to test process, test requirements, test methods and results analysis in order to establish the semantic model structure, The definition and update of the "standard Model" are quick and simple, and the relationship between them is clear. However, the "standard model" developed for one standard is difficult to be applicable to all standards. "Product model" is a semantic model with product elements as the core and product structure, function and requirements as the structure. The definition and correlation of this model are relatively complex, but it can be applied to various types of standards at the same time.

Different standardization areas can adopt the "standard model" or develop compatible "middleware" models, but from the perspective of the standard digital transformation roadmap published by IEC and ISO, the "product model" is also chosen as the semantic modeling approach. For example, IEC 61360 series standards put forward the concept of common data dictionary (CDD), hoping to establish cross-domain knowledge base for all equipment and services in the field of electrical technology. The knowledge base/database is maintained by the IEC SC3D (Product Attributes, Categories and Identification) and the TC (Technical Committee) or SC (sub-technical Committee) of each related technical area. And IEC/TC65 is developing IEC

61987 series standards, according to different product types into a number of parts.

4 KEY TECHNOLOGY OF STANDARD DIGITIZATION

According to the above analysis, standard digitization in a broad sense is to realize standard representation digitization, content digitization and application digitization through semantic technology and information technology. And machine readable is the core form of standard digitization, refers to the machine without manual operation available, readable, transferable process. To realize machine readable, we must first ensure the interaction between the machine and the real world. The basis of the interaction is the reception of language (the process of converting human language into program, including speech recognition, image recognition, natural language processing, expert system, etc.) and semantic recognition (which can be understood as program dictionary and rule base).

According to the definition given by IEC, semantics are concepts and are represented as data structures through classes and their attributes. Data structures have rules or models, and the model is an information model, that is, a declarative model that accurately describes the machine ontology and its interactions and can be recognized by other machines. In real life, semantics are difficult to cover or fully define, but information models can. Through the standardization of semantic information model, the mapping or fusion between information models can be realized, and the knowledge base built on this basis can be shared through the information model.

Semantic interoperability refers to multiple assets (such as facilities, machines and systems, etc.) between right exchange and the ability to understand each other data, which means "to understand each other data can be implemented by the transformation of the information model and the meaning of" right "refers to the assets without using artificial intervention or additional programming, information model conversion can be realized. Therefore, semantic interoperability is based on language reception and automatic transformation of semantic information models, corresponding to level 3 or level 4 of standard digitization.

Common Data dictionary is a evolving database containing all the necessary information for describing objects (equipment, products, services, etc.) in the form of categories and lists of Properties (LOP), these include administrative property list (ALOP), operational property list (OLOP), device property list (DLOP), and Business property list (CLOP). Therefore, a common data dictionary is a collection of semantics and a resource base for building semantic information models, so it is also called semantic data dictionary. The information system based on data dictionary can realize the interconnection among equipment layer, enterprise layer and industry layer, thus realizing the function of "machine readable".

The relationship of the above core concepts is shown in Figure 1. It can be seen that semantic data dictionary is the core technology to realize standard digitization, and the key to establishing semantic data dictionary lies in the construction of data dictionary architecture and knowledge association modelling. Especially in the immature stage of standard digitization research, knowledge association lacks training sets and it is difficult to realize automatic extraction and association of knowledge.

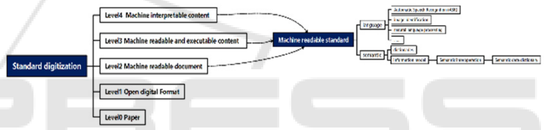


Figure 1: Standard digitization core concept relationship.

5 CASE ANALYSIS OF STANDARD DIGITIZATION

5.1 The Establishment of Knowledge Graph Model

This paper takes the national standard of industrial process automation as the analysis object, and takes pressure instrument as the core to build a partial knowledge graph model, which provides guidance for the construction of semantic data dictionary. The knowledge graph model includes the following models:

5.1.1 Class

A collection of individual objects, for grouping individual objects that have something in common, for example:

Class(ID(Person)),

Classassertion(:Person :Mary).

5.1.2 Object Attributes

Object attributes are used to represent an association between two entities, for example:

ObjectProperty(IS(hasWife)),

ObjectPropertyAssertion(:hasWife :John :Mary).

5.1.3 Data Attributes

Data attributes are used to associate entities with data values such as Integer and String, for example:

DatatypeProperty(ID(hasAge)),

DataPropertyAssertion(:has Age :John "51"^^xsd:integer).

5.1.4 Definition Domain and Range Domain

The knowledge graph model supports the declaration of domain and range for attributes and implicit additional information for attributes. The range of a data attribute is a data type, and the range of an object attribute is an entity.

According to the above research, the model of "product model" is adopted to realize modeling. According to the technical standards related to pressure instruments in the field of industrial process automation, pressure instruments are divided into pressure gauge, pressure gauge, pressure instrument, pressure transmitter and pressure controller, as shown in Figure 2.

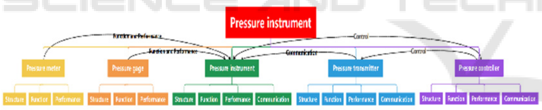


Figure 2: Product model of pressure instrument.

5.2 The Establishment of Knowledge Graph Model Transformation between Standard CDD and Knowledge Graph Model

A standard knowledge graph model in the field of product and industrial process automation can be established. MySQL is used as the modelling tool of the standard data dictionary. The transformation relationship between the pattern information and the ontology elements of the knowledge graph is as follows:

- The entity table (ET) in the data dictionary (CDD) is mapped to OWL class, which is named after the data table, namely:

$$\forall ET \in CDD \rightarrow Class(ID(ET)) \quad (1)$$

- For the column (C) in the data table, the non-foreign key column is mapped to an OWL data attribute, which is named after the column. The domain of the attribute is the class mapped from the current table, and the value field is the data type of the column, namely:

$$\forall \in attr(ET) \cap IsNotForeignKey(C, ET) \rightarrow DatatypeProperty(ID(c), domain(ID(ET)), range(datatype(C)) \quad (2)$$

- For the two data tables T and R, T is associated with R conventionally through its foreign key column FK, which is mapped to an OWL object attribute named after this column. namely:

$$\forall FK \in attr(T) \cap IsNotForeignKey(FK, T) \cap Relation(FK, T, R) \rightarrow ObjectProperty(ID(FK), domain(ID(T)), range(ID(R)) \quad (3)$$

Based on the above rules, automatic mapping of data dictionary schema information to OWL ontology classes, data attributes, object attributes and other elements can be encoded.

5.3 The Establishment of Standard Semantic Data Dictionary

According to the analysis of the standard content, the semantic data dictionary of the standard includes general semantics and special semantics, among which the general semantics is the general technical content of the standard, and the special semantics of the standard is related to the type and content of the standard, as shown in Table 2:

Table 2: List of standard semantic relationships

Generic standard semantics	*Abb.*	*Special standard semantics*	*Abb.*	*semantic relation*	*Abb.*
Foreword	FOR	Summary	SUM	Coordination	COO
Introduction	INT	Overview	OVE	Complementary	COM
Scope	SCO	Gen	GEN	Subordinate	SUB
Normative reference	NOR	Met	MET	Contains	CON
Terms and Definitions	TER	Performance	PRO	Exclusion	EXC
Abbreviations	ABB	Function	FUN	Relevance	REL
Symbol	SYM	Requirements	REU		

Generic standard semantics	***Abb.***	***Special standard semantics***	***Abb.***	***semantic relation***	***Abb.***
Bibliography	BIB				

This paper establishes mutual relations from the aspects of classification, structure, function and performance, and test methods, including reference, inclusion, mutual exclusion, and supplement, as shown in FIG. 3.

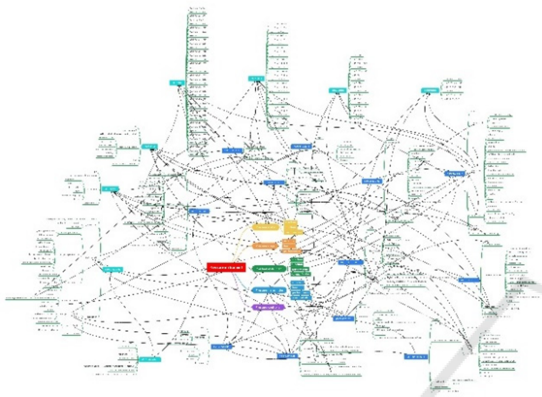


Figure 3: Standard knowledge graph model for pressure instrument

6 CONCLUSIONS

Through the understanding and analysis of the concept of standard digitization, this paper establishes the standard digitization model of typical products of industrial process automation, which provides the method of thinking for the establishment of semantic data dictionary. The main viewpoints include:

- Standard digitalization can be divided into "primary stage" and "advanced stage". Not all standards need to be "machine-readable", but "machine-readable" standards are necessarily related to application scenarios. In addition to industry differences, the main application scenarios are guidance, testing and certification.
- Semantic interoperability is the key in standard digitization. In order to achieve semantic interoperability, a standardized semantic data dictionary is needed.
- The construction of semantic data dictionary includes two approaches with "product model" and "standard Model" as the core, which is recommended in this paper. Different standard scopes mean different modelling paths, but the business path at the core is recommended.

The following deficiencies still exist in the research process of this paper, which need to be further strengthened and improved in the subsequent research process:

- The analysis of standard digitization is not in-depth enough, the analysis level only stays in the name, scope and structure, the analysis of function, performance, test methods and other content is insufficient, the most critical is that the general path and method of standard digitization has not been established.
- The analysis of standard digitization is not comprehensive enough, mainly confined to the field of industrial process measurement control and automation, and there is not enough research on relevant supporting technical standards in other fields. As a modelling method, there is also a lack of guidance for the digitalization of standards in other fields.

This paper is supported by the national key research and development program "NQI" machine readable standard generic technology and international standards in key fields (2021YFF0601400).

REFERENCES

Cao yongsheng. (2016). Construction and implementation of standard digital system[J]. Standardization in China.
Chen jin. (2020). The application of standard digitization and quantization[J]. Management informatization in China.
Chen xiaorong, Huang jianmin. (2021). Exploration of standard index retrieval method based on semantic technology[J]. Standard Science, 2021(08).
Kou jie. (2019). Exploration and research on standard application mode in digital environment[J]. Aviation standardization and quality, 2019(06).
Li jun, Lu Hongwei. (2020). Research on enterprise standard digitization thinking based on artificial intelligence[J]. value engineering.
Sabbir M. Rashid1, James P. McCusker1. (2020). The Semantic Data Dictionary – An Approach for Describing and Annotating Data[J]. Data Intelligence.
Wang chunxi, wang shuo. (2021) Research on machine readable standards in industrial automation[J]. Standardization in China.