IMEKO TC6 International Conference on Metrology and Digital Transformation
September 19 – September 21, 2022, Berlin, Germany

SEMANTIC INTEROPERABILITY IN METROLOGY THROUGH CONTROLLED VOCABULARY

Julia Neumann [a] [*]

[a] Physikalisch-Technische Bundesanstalt, Braunschweig/Berlin, Germany
[*] Corresponding author. E-mail address: julia.neumann@ptb.de

Abstract – Information always transmits a meaning. Human and machine communication relies on both the technological and the semantical exchange of information. The application of controlled vocabulary can support the semantical information exchange if certain interoperability criteria are fulfilled. The commentary paper will give an overview about semantics and semantical interoperability with a focus on controlled vocabulary. As a result, two exemplary semantic representation forms will be analysed. Further possible steps to enhance the semantics in metrological data and how to achieve semantic interoperability will be concluded from the commentary overview.

Keywords: information science, semantics, controlled vocabulary, semantic interoperability, metrological data, commentary paper

1. INTRODUCTION TO SEMANTICS IN METROLOGY

1.1. Contextualization

The metrological domain is very diverse. It describes many topics and is influenced by other disciplines. Following the lifecycle of science metrology as a domain also delivers input in the form of data to other involved parties. Those parties themselves can be very diverse. They are not restricted to science only. Aside from researchers there are, for instance, controlling bodies, industry partners and administrative bodies involved in the data driven processes. In this paper the data workflow will be considered from three core aspects:

- the production of data
- the sharing of data
- the archiving of data

When the paper refers to the combination of these three aspects it is described as data management. Data management includes more details but for this paper these three aspects have been chosen for further observation [1]. The type of data within the management processes is very heterogenous due to the broad metrological domain and its interaction with other disciplines. This paper will mostly refer to metrological data. It does not limit itself to any specific type of data such as images, simulations, modelling, coding, or text-based documents. Metrological data will be understood as any kind of information that might be of interest to the metrological domain. Respectively, information is the knowledge that a receiver transmits to a recipient through any information channel with the goal to achieve processing. For instance, the increase of knowledge or the categorization as a relevant/irrelevant type of information are such goals [2] [3]. This is a very broad scope which will allow to transmit the complexity of interoperability within the semantical context. However, it also sketches a limit to not everything in the data management workflow being of the same relevancy to metrologists. The assignment of semantics, e.g., the definition of terms or concepts, can be a highly subjective discussion. In theory, there is no limitation to how semantics is deployed and interpreted because knowledge is constantly accumulated, becomes outdated or changes. This is called the open world approach. One major side effect of this approach is that it is impossible to bring every knowledge, and thus every perspective, together [4] [5]. The goal of the paper is to show that a semantical approach with a focus on controlled vocabulary is an important, inevitable part of the solution to achieving interoperability in the data workflows. Domains like metrology are representing a challenge in the sense of the open world approach. Due to its high complexity, the metrology domain also offers a substantive basis for research about semantic interoperability. The results of this paper are an exemplary overview and initial analysis of existing approaches in metrology that are either already using partly semantics, or that use methods which would possibly benefit from adding semantics. Derived from the analytical overview, it is possible to sketch further steps for the implementation of semantics in metrology. The paper does not intend to provide a complete overview or analysis. It will restrict itself to specific examples that either can be understood as representatives for other sources with similar characteristics or/and with potential to adapt semantics. In consideration of this focus, this paper will be a commentary of the status of semantic interoperability in metrology.

1.2. Semantics and Controlled Vocabulary

Semantics is a discipline of information science that describes the meaning of signs. It applies to various research fields such as programming, philosophy and linguistics. In this paper semantics refers to its linguistical understanding. The purpose of semantics has an overall wide scope which can be condensed to three core aspects:

- the communication between humans
- the communication between humans and machines
- the communication between machines

Communication is a crucial part for the insurance of interoperability. It describes the exchange and transmission of information. To the discipline of metrology communication is important since it already produces and shares data. This might be through well-established channels or media like e-mails, directories, or library catalogues but also other, newer, multimedia approaches like data repositories and cloud systems. Even though the channels for communication exist, the data management is often not yet fully developed. In some areas, like in the field of research data management, metrologists have come to the realization that it is not enough to produce and receive information. Data requires pre-processing and archiving if the goal is to be able to understand the transmitted information and reuse it [6] [7] [8]. This is shown, for example, through the increased awareness of the FAIR data principles [9] and developments of data models such as the Digital System of Units (D-SI) [10], digital certificates like the Digital Calibration Certificate (DCC) [11] or Digital Certificate of Conformance (D-CoC) [12]. Through such developments data management is being positioned into workflows that enable metrologists to communicate with the understanding of the information that circulates in those workflows. The understanding can be either taken over by a machine or humans themselves, depending on the use cases. Metrologists that write together a proposal through different metrological disciplines might need a human-based understanding whereas conformity assessment processes need a more technical approach to analyse documents just because of the big amount of existing information. The purpose of semantics is to add a meaning to the data. Therefore, it can improve the interoperability of metrological data in the broader sense of communication [7]. The levels of interoperability will be described in chapter 2.1.

The complexity of the transmitted meaning of information is described through semantical representation forms. All of them have two characteristics in common. Firstly, they use linkage to express relation between linguistic concepts. Secondly, they apply vocabulary which is controlled either weakly or strongly. Controlled vocabulary in this context will be understood as a set of terms which has been previously elaborated and agreed upon by a target group. The elaboration process can be executed manually by humans or (semi)automatically by calculating the relevance through natural language processing methods. Lower ranking complexity semantics is expressed through e.g., glossaries and taxonomies. A glossary depicts descriptors in an alphabetical order. A descriptor can be, for instance, a keyword or a specific concept. Every descriptor relates to a human-readable explanation of the given term. Typically, the explanation is a definition of the term or concept. The definition itself can be structured in different ways, e.g., by using connotations and denotations. Therefore, the types of relations (such as synonyms, topics etc.) found in a glossary can vary depending on the structure of the definition. It usually needs to be coordinated within the community, context and data managers what approach is eventually used to make a definition. A glossary gives the opportunity to harmonize processes on a human-readable scale while the enhancement of a meaning proceeds through contextual relations [6] [13] [14] [15] [16].

Another low-level representation form for semantics is a taxonomy. The taxonomies, as well as glossaries, are based on controlled vocabulary. The focus of taxonomies, however, is usually the hierarchical depiction of the vocabulary. Effectively, the prominent relation types that connect the terms and concepts are broader and narrower term relations. Because of this, taxonomies are often applied when there is a need of a systematic classification rather than a human-readable connotative explanation of a term or concept [6] [13] [14] [15] [16].

Thesauri are an example of higher-ranking representation forms for semantics. They combine characteristics from both glossaries and taxonomies. On the one hand, they offer human-readable definitions. On the other hand, they simultaneously structure the contents fit for computational usage through different relations, among them synonyms and related terms. This machine-interpretability is achieved, for example, through formats like RDF (Resource Description Framework) which were specifically developed for this purpose. Thesauri usually require a significant amount of human intervention. Controlled vocabulary needs continuous maintenance. If something is not described, it will be difficult for both humans and machines to find the information [6] [13] [14] [15] [16].

Higher ranking representation forms offer more automatization. Ontologies are an example for it. Instead of relying as heavily on the structuring of terms, ontologies allow to describe and define concepts. This adds more flexibility to the overarching goal of harmonization of information. The machines will be able to make simple assumptions and to fill in gaps. For example, if two concepts are described in an ontology, however the linkage between those two concepts has not been fully described, it still will be possible for the machine to fill the gap to some extent solely based on what kind of information is given. It does not achieve the same capacities as an AI would. An ontology will not allow a machine to make predictions. The flexibility of ontologies can also have disadvantages. Human feedback and maintenance are still necessary. However, ontologies have a high complexity level [6] [17]. Linguistical and computational training as well as an understanding of semantics are a barrier that must be considered. The different characteristics, including the advantages and disadvantages, of the representation forms can be balanced well if the use cases are clear and, more importantly, if the interoperability is provided. The following chapter will describe the main interoperability levels that need to be considered when a data workflow is based on semantics. Their importance will be described, followed by a discussion of two examples within the metrological scope.

2. APPLICATION OF CONTROLLED VOCABULARY IN METROLOGY

2.1. Semantic Interoperability Levels

Semantic interoperability describes the ability to exchange information between different systems. In contrast to the general definition of interoperability, semantic interoperability specifically focuses on the transmission of meaning between involved parties. These involved parties are generally computational systems or human interaction with machines. In a broader sense, the human-human interaction is also a part of an information workflow. As such it can represent an information exchange which requires a common understanding of everyone involved in the workflow without the obligatory use of computational system. This paper therefore will consider it as an equally important aspect of semantic interoperability. Semantic interoperability, in its simplified form, deals with the disambiguation of information. Disambiguation in this paper is used in the context of word-sense identification. Words mainly get a meaning through their contexts, e.g. in sentences and the domain. It is possible to exchange information without disambiguation, however it will not be possible to always ensure an understanding of the meaning. Computational approaches like natural language processing methods and the application of controlled vocabulary are providing disambiguation. However, the usage of controlled vocabulary does not automatically provide semantic interoperability. Similar to the FAIR principles and the requirements for metadata quality, controlled vocabulary (and therefore the respective representation forms) itself can adapt certain characteristics that will provide interoperability. The paper refers to them as semantic interoperability levels [2] [8] [18][19] [20] [21] [22] [23] [24] [25].

The first interoperability level of importance to controlled vocabulary is the technological level. It mainly refers to the capability of mapping data. Some mapping problems have already been provided with solutions. As mentioned previously, there is a widely accepted standard for the description of semantic representation forms and (meta)data models. It also provides a formalization for the description of relations between terms or/and concepts within the representation forms. As an effect, it is possible to upgrade or downgrade representation forms. The application is shown in Fig.1, Fig 2. and Fig 3. This is important for the adaption of vocabulary with flexibility and with a low threshold for heterogeneous applications.

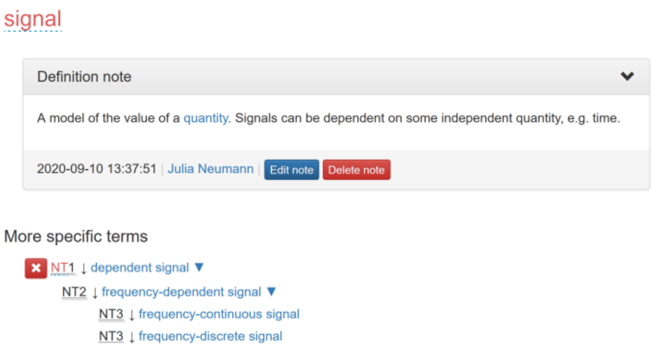

Fig. 1. Excerpt of a thesaurus. The information from the thesaurus will be represented as a glossary.

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT DISTINCT ?label ?definition WHERE {
 GRAPH ?g { ?s skos:prefLabel ?label .
          ?s skos:definition ?definition}
}
LIMIT 200
```

Fig. 2. Example of a SPARQL query. Specific information from the thesaurus will be extracted.

Table 1. Excerpt of the glossary extracted from the thesaurus.

input signal	The input signal is the signal that enters the system. For example, the acceleration signal measured by an accelerometer is its input signal.
non-dynamic system	A degenerated differential equation, whose derivatives equal zero momentarily and locally, represents the preliminary description of a dynamic system in a steady (static) state, i.e. of a non-dynamic system. Some multivariate algebraic equations represent the system behavior of non-dynamic systems.
non-parametric equation	An equation that does not contain equation parameter.

However, the existing standards cannot fully solve content-related challenges of controlled vocabulary. It does not provide guidance for the curation of data. It does not provide relevant use cases for the controlled vocabulary. Lastly, it does not substitute linguistic proficiency. For a heterogeneous and highly interconnected domain like metrology it means that there need to exist representation forms that provide standardized frameworks for mapping but at the same time they must provide flexibility for the creation of its content. A good balance is required.

The linguistic proficiency represents a challenge for semantic interoperability. There are several aspects that to the date are relevant research topics. One of them is multilinguism. There are roughly 7000 languages in the world. Every one of them has a different mindset. The linguist Ferdinand de Saussure has described in a semiotic model that it depends on the speakers of the languages which meaning they link to a sign [26]. Cultural diversity, therefore, is one important reason why so many linguistic concepts exist. It occurs that different cultures can use the same words for different concepts. For example, the German word Handy means mobile phone, however handy in English expresses that something is convenient or useful. Metrology is an international discipline. It is expressed through many different languages. Therefore, it is confronted with these linguistic challenges. A common example is the word quantity. Without a context it can apply to many semantic concepts such as being an amount, number, size, parameter etc. A direct translation is further complicated if the intrinsic cultural agreement of conceptual understandings is considered. For instance, the direct translation of quantity to French grandeur implies the preferred concept “size” over the other possible concepts of the English term. The introduced representation forms can overcome this challenge because they allow to link the different concepts with each other. Supported by natural language processing and information retrieval methods it is possible to (semi)automatize the processes and to give multilingual outputs that are use case relevant. In metrology there have been already made efforts to find solutions to these challenges. For example, there has been developed an official International Vocabulary of Metrology (VIM) [27] which is multilingual. Furthermore, there are different ISO norms (ISO 25964) [28] and national vocabulary developments such like from Deutsches Institut für Normung (DIN-Terminologieportal) [29]. Besides this, there are various internal national metrology institutes and industrial projects describing initial efforts in the development of a controlled vocabulary or/and representation forms for sub/crossdomains of metrology like the Semantic Sensor Network Ontology [30], NFDI (Nationale Forschungsdateninfrastruktur) ontology collections and developments [31], e-class [32] and the Common Data Dictionary [33].

There are six other semantic interoperability levels. They will be mentioned in less detail because they either exist due to similar reasons as the other levels or because they are applied outside of semantics and it is expected that they need less explanations to the target group of metrologists. However, this does not mean that in their deployment they do not have special semantic patterns that require innovative treatment. The deployment will not be the scope of this paper. In the first chapter the open world approach has been described. Its philosophy requires for the semantic interoperability to develop concepts for the accessibility, completeness, conciseness, legibility, and provenance of the vocabulary. The idea behind semantics as highlighted previously is to improve communication. This can only be achieved if the controlled vocabulary can be used by others. The accessibility can refer to concepts like open access principles or the existence of a representation form for the controlled vocabulary where the information can be found and retrieved, e.g. through a common standard like RDF. It is often not enough to just have access if the overall goal is the usage and retrieval of relevant information. It is necessary to know the provenance of the vocabulary, e.g. for it to be permanently accessible and also to be able to reuse it for own projects. Persistent identifiers are a common method to achieve this. Furthermore, the representation forms need to reflect the designated use cases without missing any linkage between concepts. Only then every information can be found. Introducing high-level automatization processes like natural language processing and machine learning components can help to achieve this goal. Classic methods involve collaborations with target groups such as interviews and the analysis of corpora from domain specific controlled vocabularies. To have a complete controlled vocabulary is just one side of the coin. Completeness does not always equal coherence or relevancy. Because of this, an important semantic interoperability level is the conciseness. It behaves similarly to the idea of effectiveness and efficiency. A system can be effective, however this does not mean it is automatically efficient. Of course, a controlled vocabulary can never be complete, but it can become mostly complete. If a proper context for its usage is selected, effective concepts and measurements can be chosen to deploy the vocabulary in the most meaningfully complete and concise manner. The interoperability level of legibility interacts with the other levels. It describes how well the controlled vocabulary can be understood by the machines or humans who are using it. On the one hand it relies heavily on the existence of technical standardization and the capability to describe the intercultural meaning completely. On the other hand, it also expresses the need for a semantic proficiency. It means that the vocabulary is either designed along the needs of the users, so it can be understood by them without prior knowledge of semantics, or there needs to be provided training that makes the users proficient enough to understand the vocabulary. The latter case is especially important, if the users are involved in the development of the controlled vocabulary. For the application of vocabulary, it is not as necessary because the information behaviour is best satisfied if the existence of a controlled vocabulary is not noticed by the end-users. Legibility is not exclusively a challenge to humans, however there are solutions like SPARQL queries [34] which support the legibility for machines.

The last semantic interoperability level which will be mentioned is the maintenance of the controlled vocabulary. Like for any other operational systems it is necessary for the controlled vocabulary to have concepts which ensure that the representation forms are running smoothly. For the semantical approach it is especially important to have concepts for the management of the content. Concepts can change their meaning over time. Additionally, there will be empty spaces in the controlled vocabulary due to the nature of the open world approach. If various controlled vocabularies are linked to each other, there need to be mapping solutions for when one of them changes its content. This means that the controlled vocabulary will require attendance in the short and long run. A suitable maintenance workflow needs to be sketched [8] [18][19] [20] [21] [22] [23] [24] [25].

2.2. Conclusion - Controlled Vocabulary in Metrology

Controlled vocabulary in metrology has been existing for a long time. However, it has not yet been widely used for human-machine or/and machine-machine communication. The following short analytical overview of two representation forms in metrology will highlight the current state of the art and draw a conclusion for the application of controlled vocabulary.

The scope of VIM [27] is very broad. Its target group are scientists and engineers of any discipline. The vocabulary is being updated and curated periodically. There is a printed publication and a digital representation of the latest version. Both can be accessed in English and French. Other translations of VIM like in German exist, however only in a fee-paying publication format. Due to the scope of VIM, the content reflects general concepts of metrology and does not specify subdomains. The overall structure of VIM includes characteristics from taxonomies and glossaries. Both the printed and the digital representation clarify that the defined terms are most often concepts. However, the concepts itself are only loosely connected. Most of the contents rely on a human-readable structure or definition instead of machine-readability. There is a low-level hierarchy which is expressed through broader and narrower term relations. There are also synonyms and a small set of related terms. Especially the latter are usually highlighted by being printed bold. This is a visual pattern which is suitable for the human comprehension and can be found often in glossaries. Besides this, there are a lot of commentary notes that have the function of annotations for further in-depth explanations. However, there is no machine-interpretable processing. The digital representation introduces an information retrieval component. Whereas the printed version relies on an alphabetical order of the entry terms for browsing, the digital VIM allows to search for specific terms through a search box. In the context of semantic interoperability, VIM is having some major gaps. Nevertheless, it does a good job at introducing the basics of structuring a controlled vocabulary on a content-based level. For instance, it acknowledges the need of concepts over a strictly term related approach. It also introduces relation types between the terms. These might be too flat, considering the completeness level of semantic interoperability and the possible use cases. On the one hand, the scope is to give to none-metrologists access to metrology, too. It might make little sense to overcomplicate the vocabulary by controlling it too much through a high-level structure. Additionally, the users are encouraged to use VIM as a basis for their own projects. On the other hand, the lack of well described relations between terms might hinder the overall machine-interpretability and the usefulness within the metrological workflows because the completeness has not been achieved. A solution to the machine-interpretability might be an upgrade of VIM to a higher-ranking semantic representation form such as a thesaurus. This is relatively easy to achieve since VIM already uses a good, similar basis.

The Common Data Dictionary [33] is a free to access domain specific semantic representation form that is using controlled vocabulary. It is based on the vocabularies of the International Electronic Commission (IEC). Although it is maintained by the IEC, users outside of the commission can hand in their feedback as well. The CDD currently covers four domains in five languages. Rather than terms or concepts it refers to classes. The classes serve in this case a similar purpose like topics or broader terms. The classes always introduce one or more subclasses. This is a hierarchical structure. The lowest class in the hierarchy is often a specific term that describes characteristics of the overall general classes. The information about the classes is structured into metadata schemes which consist of metadata fields and the metadata contents. The metadata contains, for example, information about the definitions. They also contain technical data like provenance information. The content description is often not easy to understand for a human without prior training. This can be a side effect of a vocabulary that is controlled too much. The advantage of this approach is that it can be rather easily made available in a machine-readable format. The chosen format in the CDD is based on Excel. Excel formats cannot represent complicate semantic structures because of their highly interconnected nature. It is likely that the CDD will not be fully machine-interpretable. Many content fields in the metadata schemes are yet empty, so there might be no need for an enhanced export of the semantic representation. It might also indicate that there is no concept yet for how to achieve completeness of a controlled vocabulary and how to maintain it although users are already encouraged to give feedback. The CDD is a good example for how to initiate a controlled vocabulary. It draws its content from existing sources, and it pays attention to a machine-interpretable structure. It depends on the scope and the use cases whether this approach is suitable. It does not support annotations or as much textual explanations as a glossary would. One should be aware of it and consider it prior to using the CDD.

Semantics is important for a successful data workflow on a human and machine-based level. The paper has shown that in metrology the awareness for this has increased. However, for semantics to be implemented successfully, it is necessary to consider the various interoperability levels. The analytical overview has shown that the existing representation forms relevant to the metrological domain are offering a good basis, but they do not implement semantics thoroughly. It is recommended to analyse what needs and relevancies should be considered during the development of a controlled vocabulary. These processes can be either supported by qualitative approaches like interviews or by applying relevancies through information retrieval methods. For such a heterogenous and interactive discipline like metrology it most likely means that there is a vast amount of relevant information. Representation forms like thesauri or ontologies could be a solution to structuring the information. Their primary advantage is the cross-linkage of information. This works for terms and concepts but also for different subdomains. At the PTB (Physikalisch-Technische Bundesanstalt) a metrology thesaurus [35] is in development. The primary goal of the thesaurus development is to research concepts that will offer a balance between flexibility and control of the vocabulary in metrology. Patterns and characteristics of the different metrological subdomains will be analysed and structured. The structuring will happen at least on two levels – within the metrology domain as a whole and within the network of the metrological (sub)domains. Automatization processes will be considered as well as human interaction. It is certainly a difficult task, however the paper has shown that there are already a lot of sources that can be used because data and communication management has already been a part of metrology. Effectively, creating an interoperable semantic information network is the most important goal.

ACKNOWLEDGMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[February revision cycle] [1] Thomer, A. K., Akmon, D., York, J., Tyler, A. R., Polasek, F., Lafia, S., ... & Yakel, E. (2022). The craft and coordination of data curation: complicating" workflow" views of data science. arXiv preprint arXiv:2202.04560.

[2] Breznik, Kristjan (2018). Procedia - Social and Behavioral Sciences, ISSN: 1877-0428, Vol: 238, pp. 141-148.

[3] Virkus, S. and Garoufallou, E. (2019), "Data science from a library and information science perspective", Data Technologies and Applications, Vol. 53 No. 4, pp. 422-441. https://doi.org/10.1108/DTA-05-2019-0076.

[4] Maree, M., Kmail, A. B., & Belkhatir, M. (2019). Analysis and shortcomings of e-recruitment systems: Towards a semantics-based approach addressing knowledge incompleteness and limited domain coverage. *Journal of Information Science*, *45*(6), 713-735. https://doi.org/10.1177/0165551518811449.

[5] Keet C.M. Open World Assumption. In: Dubitzky W., Wolkenhauer O., Cho KH., Yokota H. (eds) Encyclopedia of Systems Biology. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9863-7_734

[6] Burgin, Mark (2016): Theory Of Knowledge: Structures and Processes. Singapur: World Scientific.

[7] Borgman, C.L., Scharnhorst, A. and Golshan, M.S. (2019), Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. Journal of the Association for Information Science and Technology, 70: 888-904. https://doi.org/10.1002/asi.24172.

[8] Tognoli, Natalia, and José Augusto Chaves Guimarães. "Provenance as a knowledge organization principle." KO KNOWLEDGE ORGANIZATION 46.7 (2020): 558-568.

[9] Annika Jacobsen, Rajaram Kaliyaperumal, Luiz Olavo Bonino da Silva Santos, Barend Mons, Erik Schultes, Marco Roos, Mark Thompson; A Generic Workflow for the Data FAIRification Process. Data Intelligence 2020; 2 (1-2): 56-65. doi: https://doi.org/10.1162/dint_a_00028.

[10] Hutzschenreuter, Daniel, Härtig, Frank, Heeren, Wiebke, Wiedenhöfer, Thomas, Forbes, Alistair, Brown, Clifford, Smith, Ian, Rhodes, Susan, Linkeová, Ivana, Sýkora, Jakub, Zelený, Vít, Ačko, Bojan, Klobučar, Rok, Nikander, Pekka, Elo, Tommi, Mustapää, Tuukka, Kuosmanen, Petri, Maennel, Olaf, Hovhannisyan, Kristine, … Paciello, Vincenzo. (2019). SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data (D-SI 1.3). Zenodo. https://doi.org/10.5281/zenodo.3522631.

[11] https://dccwiki.ptb.de/en/DCC_basic_ideas (19.04.2022)

[12] https://www.ptb.de/cms/ptb/fachabteilungen/abt9/fb-94/metrologische-dienstleistungen.html (19.04.2022)

[13] Knorz, Gerhard (2004): Informationsaufbereitung II: Indexieren. In: Kuhlen, Rainer u.a. (Hg.): Grundlagen der praktischen Information und Dokumentation. 5. Auflage, München: Saur. p. 179-188.

[14] Burkart, Margarete (2004): Thesaurus. In: Kuhlen, Rainer u.a. (Hg.): Grundlagen der praktischen Information und Dokumentation. 5. Auflage, München: Saur. p. 141-154.

[15] Kasprzik, Anna. (2018). Community-getriebene Entwicklung von Sacherschließungswerkzeugen rund um die GND // Von der Normdatei zum Thesaurus, und vom Thesaurus zur Ontologie -- strukturelle Stolpersteine. 10.13140/RG.2.2.14154.06087.

[16] Pearly Jade N. Embajador 2019 J. Phys.: Conf. Ser. 1254 012037.

[17] Qin, Jian. "Knowledge Organization and Representation under the AI Lens" Journal of Data and Information Science, vol.5, no.1, 2020, pp.3-17. https://doi.org/10.2478/jdis-2020-0002.

[18] Rahman, H, Hussain, MI. A comprehensive survey on semantic interoperability for Internet of Things: State-of-the-art and research challenges. Trans Emerging Tel Tech. 2020; 31:e3902. https://doi.org/10.1002/ett.3902.

[19] Venceslau, A.; Andrade, Rossana Maria; Vidal, V.; Nogueira, T. and Pequeno, V. (2019). IoT Semantic Interoperability: A Systematic Mapping Study. In Proceedings of the 21st International Conference on Enterprise Information Systems - Volume 1: ICEIS, ISBN 978-989-758-372-8, pages 535-544. DOI: 10.5220/0007732605350544.

[20] J. Nilsson and F. Sandin, "Semantic Interoperability in Industry 4.0: Survey of Recent Developments and Outlook," 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), 2018, pp. 127-132, doi: 10.1109/INDIN.2018.8471971.

[21] Marko Marković & Stevan Gostojić (2020) A knowledge-based document assembly method to support semantic interoperability of enterprise information systems, Enterprise Information Systems, DOI: 10.1080/17517575.2020.1793389.

[22] Elouataoui, W., El Alaoui, I., Gahi, Y. (2022). Metadata Quality in the Era of Big Data and Unstructured Content. In: Maleh, Y., Alazab, M., Gherabi, N., Tawalbeh, L., Abd El-Latif, A.A. (eds) Advances in Information, Communication and Cybersecurity. ICI2C 2021. Lecture Notes in Networks and Systems, vol 357. Springer, Cham. https://doi.org/10.1007/978-3-030-91738-8_11.

[23] Debattista, J., McKenna, L., Brennan, R. (2018). Understanding Information Professionals: A Survey on the Quality of Linked Data Sources for Digital Libraries. In: Panetto, H., Debruyne, C., Proper, H., Ardagna, C., Roman, D., Meersman, R. (eds) On the Move to Meaningful Internet Systems. OTM 2018 Conferences. OTM 2018. Lecture Notes in Computer Science(), vol 11230. Springer, Cham. https://doi.org/10.1007/978-3-030-02671-4_32.

[24] Király, P., Stiller, J., Charles, V., Bailer, W., Freire, N. (2019). Evaluating Data Quality in Europeana: Metrics for Multilinguality. In: Garoufallou, E., Sartori, F., Siatri, R., Zervas, M. (eds) Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science, vol 846. Springer, Cham. https://doi.org/10.1007/978-3-030-14401-2_19.

[25] Eisenstein, J. (2019). Introduction to natural language processing. MIT press.

[26] De Saussure, F. (2011). Course in general linguistics. Columbia University Press.

[27] https://jcgm.bipm.org/vim/en/index.html https://www.oiml.org/en/files/pdf_v/v002-200-e07.pdf (19.04.2022).

[28] http://www.niso.org/schemas/iso25964 (19.04.2022).

[29] https://www.din.de/de/service-fuer-anwender/din-term (19.04.22).

[30] https://www.w3.org/TR/vocab-ssn/ (19.04.2022).

[31] https://nfdi4cat.org/services/ontologie-sammlung/ (19.04.2022).

[32] https://www.eclass.eu/index.html (19.04.2022).

[33] https://cdd.iec.ch/cdd/iec61360/iec61360.nsf/ TreeFrameset?OpenFrameSet (19.04.2022).

[34] https://www.w3.org/TR/rdf-sparql-query/ (19.04.2022).

[35] https://digilab.ptb.de/vocab/index.php (19.04.2022).