The Challenges of Semantic Interoperability in the Era of eScience on the Web

Maria Luiza de Almeida Campos*, Linair Maria Campos**
and Nilson Theobald Barbosa***

Programa de Pós-Graduação em Ciência da Informação, UFF,
Rua Tiradentes 148, Ingá, Niterói, RJ, 24210-510, Brazil,
*<maria.almeida@pq.cnpq.br>, **<lmcampos@id.uff.br>, ***<nilson@tbarbosa.org>

Maria Luiza de Almeida Campos holds a PhD in information science and is Researcher and Professor of the post-graduate programs at Universidade Federal Fluminense, Niterói, Rio de Janeiro, Brazil. Her areas of interest are knowledge organization, models and theories of knowledge representation, terminology, semantic interoperability and foundational ontologies.

Linair Maria Campos is Researcher and Professor in the Department of information science, graduate and post-graduate programs, at Universidade Federal Fluminense. She has a master's degree in informatics from Universidade Federal do Rio de Janeiro and a DSc in information science from Universidade Federal Fluminense. Research interests include knowledge organization, models and theories of knowledge representation, terminology and foundational ontologies.

Nilson Theobald Barbosa has a master's degree in information science from Universidade Federal Fluminense and presently is a doctoral student of information science and worked as a substitute professor at Universidade Federal Fluminense. He is also a systems support analyst at Universidade Federal do Rio de Janeiro. Research interests include knowledge organization, models and theories of knowledge representation, semantic web technologies and interoperability of vocabularies.

Campos, Maria Luiza de Almeida, Linair Maria Campos and Nilson Theobald Barbosa. 2020. "The Challenges of Semantic Interoperability in the Era of Escience on the Web." *Knowledge Organization* 47(8): 680-695. 75 references. DOI:10.5771/0943-7444-2020-8-680.

Abstract: This article is an exploratory study that intends to present semantic interoperability initiatives in the area of information science and computer science. It discusses these initiatives, focusing on open science and eScience, aiming at pointing out similarities and differences in the methodologies used for an intelligent retrieval of data in heterogeneous environments that do not benefit a universal language. We will illustrate seminal initiatives regarding the compatibility of languages in information science and the alignment and semantic mapping in the area of computer science, emphasizing what they have in common and their differences, showing that those seminal initiatives are still relevant nowadays. In this perspective, we want to bring up discussions about methodologies that supports a kind of "intermediate language" that allows the compatibility of research resources, semantically interconnecting their contents without forcing the use of a single language, but of a switching language, such as a metalanguage that guarantees to keep the differences of each community of speech.

Received: 24 August 2019; Revised: 13 December 2019, 28 January 2019, 22 May 2019; Accepted: 10 July 2020

Keywords: data, information, semantic interoperability, compatibility

1.0 Introduction

In this article we discuss the issues concerning semantic interoperability and the methodologies used to enable intelligent data retrieval. In this sense, we investigate methodologies that allow the development of an "intermediate language" that enables the matching of research resources, semantically interconnecting their contents in different contexts, in special open data and eScience. The question that arises is how to achieve semantic interoperability between the existing sources of information in these contexts, ensuring their compatibility without forcing the conversion of each one into a single vocabulary. Thus, in this article, we intend to point out similarities and differences in the methodologies investigated and used in the areas of information science and computer science aiming at intelligent retrieval approaches that do not rely on a universal language for data interoperability, showing that seminal methods and strategies proposed by information science should be allied to modern methods proposed by computer science to address the interoperability issue nowadays.

At the end of the nineteenth century, Paul Otlet and Henri La Fontaine conceived an environment where world scientific production could be found based on an expressive effort to collect and organize scientific information. This visionary achievement paved the way for other initiatives that would emerge in the area of knowledge and information organization, with a focus on the document. Otlet also conceived a future world where information in various media would be available to the user at his desk on a screen that would provide not only texts but images, sounds and videos from media such as television, microfilm and cinema (Rayward 1991). After a while, in 1945, Vannevar Bush imagined these information resources connected. So, he was considered the forerunner of hypertext (Robredo 2011). Nowadays, a few decades ahead, we are witnessing these future visions becoming true, and we have been going through an era where the access to information is more widespread and mobile, but, in addition, we live the possibility of access to an immense and diverse volume of data (structured, semi-structured and unstructured), in an open format and often interconnected with one another. We have access to another world of potential information, where a user is inserted whose focus is no longer to find and access documents but to obtain and make sense of this huge volume of data, asking questions for which answers are expected based on the available data. The borders for knowledge acquisition have been stretched out. In addition to it, we are challenged to provide the technological means and to think the requirements that the technologies must reach. We also need to gather differentiated kinds of expertise and efforts in order to ask the pertinent questions, make sense and interpret the answers obtained based on these data. There are many advances in this regard, and in this context DBpedia and Wikidata stand out as central data sources for the embryo of a semantic web being built in a collective and democratic way, which we can already perceive as a vast and varied worldwide database, whose content can be searched by the average user, albeit with some small degree of difficulty (Burgstaller-Muehlbacher 2016).

On the other hand, there are still many challenges to be faced, in particular when taking account the compatibility of vocabularies during the access to this increasingly open mass of data connected in the web. In this context of open data on the web, the relevance of the access to scientific research data stands out as a support for its reuse, to the generation of new knowledge, research visibility and other factors.

The current scenario that is related to the processes, which involve scientific research and the sharing and reuse of information among researchers, refers to the creation of mechanisms in an informational space that makes possible the scientific dissemination of methods, resources and products, the fruits of the activity of research. There is an increasing worldwide interest in an open science at all levels from scientific publications to scientific research data, through all stages and processes involved in scientific research, like public and private funding, resources used, accountability for the society, human resources utilized and similar services. On the other hand, as a result of research activities and the collective willingness to allow and give access to this information, the exponential increase in the amount of information creates a significant difficulty to find relevant information, also strangled by the enormous variety of areas of knowledge and institutions, a large number of data formats and various metadata used. Thus, we consider that in the field of information science it is necessary to investigate mechanisms, which make it possible to create a network of information that, despite the diversity and heterogeneity, can link information spread across several institutions and different areas of knowledge and make sense of them. In this context, ontologies are being used in online scientific activities, or eScience, mainly in papers related to the management and integration of data resources and workflows, with the view to allow a more explicit representation of scientific artifacts (Brodaric and Gahegan 2010). It is in this context, that we present the interoperability issue as a condition for heterogeneous information systems to interconnect semantically similar information.

For Payette et al. (1999), interoperability is defined as the ability of the components or services to be functionally and logically interchangeable, because they have been implemented according to a set of well-defined and known interfaces. However, the access to heterogeneous systems can involve documents (or datasets) from different research groups and institutions and even different countries, that have their information indexed quite differently, whether linguistically, structurally or in distinct conceptual schemas. In a diversified environment in some or all these aspects, it may not be possible to use the same interfaces with nomenclatures, and retrieval processes may lead to unsatisfactory results. In such cases, the users may be led to do multiple checking to get relevant results or sometimes even get some results. For the establishment of interoperability between different systems or indexing languages, some methods, at different levels, can be adopted.

The substantial increase in technologies that promote open interfaces and the creation of comprehensive metadata standards, which can represent academic research environments, are undoubtedly great steps towards advancing the process of interoperability. The problem is in the treatment of this research data at a level not only descriptive but also semantic. This includes addressing related issues and discussing the possibilities of sharing such data by relating their content. Within the semantic web environment, the study and the creation of ontologies have relevant aspect in the solution of the problem. Ontologies are a powerful way of interrelating systems. They are elaborated, mostly, aiming the structuring of knowledge bases or to be used as semantic tools in the support to the interoperability between information systems (Campos 2007).

In this case, it is very important for institutions from any field, to create their controlled and specialized ontologies or vocabularies, but the enormous semantic heterogeneity is not capable of allowing semantic interoperability by itself, since it would be unlikely to create a global ontology to which everyone would simply adopt. Therefore, in the real world it is necessary to create mechanisms that enable this semantic interoperability between heterogeneous systems. Interoperation between two systems seems to be possible only when we can overcome the linguistic, terminological, and verbal differences that are defined by the idiosyncratic conditions of each environment. So, we can compare and link the definitions and concepts associated with each object. In this sense, we are engaged to investigate methodologies, both in the scope of information science and computer science, which permit us to develop a kind of "intermediate language," that allows the compatibility of research resources, semantically interconnecting their contents.

Dahlberg (1981), in the area of information science, proposes the construction of a conceptual compatibility matrix based on her analytical-synthetic method. The conceptual compatibility matrix is a mapping of the semantic potentiality of the languages studied, providing the results of the compatibility analysis between languages from the semantic and structural point of view. Another work published in the area of information science that deals with the compatibility theme is the thesaurus reconciliation method proposed by Neville (1970), which is based on the principle that compatibilization must consider concepts (the conceptual contents of the descriptors, which are expressed in the definitions) and not the descriptors only. This method proposes an intermediate language approach, based on the numerical coding of concepts through which it is possible to establish the conceptual equivalence of descriptors of different languages, considering also the one to *n* correspondence treatment between the terms to be matched.

In the field of computer science, related to the studies of semantic representation languages, the literature has presented several forms of compatibilization. Here we emphasize those related to the concept of alignment (Bruijn et al. 2006), which allows the creation of links between ontologies, while preserving them without change.

This paper intends to contribute with research addressed to discussing and finding solutions to the problems of semantic opacity already pointed out by Pierre Lévy in his works, when he affirms that we are in the process of establishing a common participatory memory to all humankind. For Lévy (2014), the limitation we have today, at the beginning of the twenty-first century, to explore this immense memory of data are the problems of understanding its terminological meaning, incompatibility of classification systems and linguistic and cultural diversity. Under these circumstances, the lack of models that can be computationally treatable prevents the automation of most of the cognitive operations of analysis, selection, synthesis and interconnection of potential information, and, thus "we do not know yet how to systematically transform this ocean of data into knowledge, and still less how to transform the digital media into a reflective observatory of our collective intelligences" (Lévy 2014, 23).

The statement of Lévy, mentioned above, motivated us to take this path, that is, to raise initiatives that can provide some solutions for an intelligent retrieval for the immensity of data produced today. Then, we are starting these studies in our research group. Hence, this article has an exploratory objective, of raising questions that go through this issue. Thereafter, this article is then organized, besides this introduction, to present the questions that involve the great diversity of data currently produced, and the problem of intelligent access to this data. Section 2.0 "Research data in a new context: open science and eScience," presents specific questions that involve the research data in the context of open science and eScience, in addition to raising a current discussion, that is, restricted or unrestricted access to the retrieval of such data. It also raises the matter of the possibilities of relating this data through its contents in a significant way in an environment, which semantic technologies are already taking their first steps. Thus, in Section 3.0 "Semantic compatibility studies: contributions of information science and computer science," we present initiatives in information science and computer science of semantic compatibility and interoperability showing similarities and different methodologies presented in prospects of mapping possibilities for an intelligent retrieval to environments where the research presents itself. And finally, we present in our final considerations aspects that we consider fundamental for such initiatives to be undertaken.

2.0 Research data in a new context: open science and eScience

Research data as stated by the Organization for Economic Cooperation and Development (Pilat and Fukasaku 2007) are records of facts used as primary sources in scientific research and are generally accepted in the scientific community as necessary for the validation of the research results. These data, according to the nature of the knowledge in which they are generated, can be identified as numbers, images, software, algorithms, animations, texts, and other resembled items. The literature presents a range of typologies about the classification of research data (Borgman 2010; Harvey 2010; Lyon 2007; NSF 2007) where each classificatory form is linked to certain purposes. Then, as we can observe, the research data are collected to serve several purposes, using various methods, being possible either the purposes or the methods to influence on the data classification; moreover, these classifications also relate to how these researchers are willing to share their data with their peers (Sales 2014).

Research data are also called scientific data and have always been part of the process of doing scientific research, but they were not considered and are not always presented in the products/research results. We can discriminate such data as information objects produced, collected or used throughout scientific research. They are distinct and have a heterogeneous, diverse and complex nature. With the advent of the eScience and the open science movement, scientific data have gained prominence and recognition of its importance not only to the scientific community but also to society. These two concepts have a certain conceptual correlation, although they are not similar, as we can observe.

The term eScience was created by John Taylor in 2001 as a new way of doing science. That is, by establishing a global collaboration in key areas of science and every generation of infrastructure that will enable the sharing of information among researchers. From Taylor, other researchers have contributed to the ideals of eScience (Gray et al. 2005; Hey and Hey 2006; Hey, Tansley and Tolle 2009; Taylor 2001) in which ICTs and data repositories should play a prominent role. The creation of such digital repositories of scientific data distributed in various spaces where the research that is being done will require management services similar to the ones of the conventional digital libraries, as well as other data management services. In this case, it is important to point out that the availability of this data is expected for the relation between specialists. To make science (or e-science as it can be said today) requires manipulating and storing large quantities of these data, and still requires that such data may be available, to be shared and reused by other researchers and research groups, with similar objectives or not from the group that originally used or produced them. Though, disclosure in eScience, in a way, is more used to the availability of scientific data among peers.

In relation to the open science, its primary concern is to make research activity more transparent, more collaborative, and more efficient, in which scientific knowledge must be understood as world heritage. According to Akmon et al. (2011), scientists can no longer assume that the data generated during their research has value only to themselves or to their research group. Instead, they need to anticipate future uses of the data and, thus, preserve them and provide descriptions that facilitate their reuse. But these scientists do not have the knowledge, the techniques, the interest or the time to do it.

It is in this context that the proposals for the creation of digital curators of scientific research data, in accordance with Lee and Tibbo (2007), have as main tasks to boost the availability of data for the scientific community, to offer effective storage, to transform data, to preserve for a long term period, to make available authentic digital data to be reproduced and reused, to develop reliable digital repositories, to generate principles of metadata creation and capture and to use open standards for file formats and conversions. Besides, while data is central to the scientific process, metadata is central to the operation of digital files and to the curation of digital data, but the essence of the problem is to ensure that all the metadata and their associated documentation, that are required for the efficient use of data are correctly generated, understood and effectively accessible to its legitimate users (Davenhall 2011).

These repositories of scientific research data, despite being a major advance in the description, management and sharing of research data, confront a major problem. The extreme complexity of the data models involved and the enormous diversity of objects and domains to be represented form an almost insurmountable barrier to the efficient management of these environments, preventing its main purpose, which is, sharing, discovery and reuse of these data to be performed properly. It is a clear example that only the computer and computing infrastructure that forms the basis of the global information network are not enough to solve the existing problem.

The issue we want to address here is what contribution information science can make to open science and eScience. We discussed and used research data and eScience as a context for our argument, because despite the importance and all the effort dedicated to the preservation, dissemination and sharing of this data, there is still a problem to be solved nowadays, not only eScience data, but digital data in general; interoperability between heterogeneous databases, especially their semantic interoperability, so that researchers can reuse research data without the need to modify the vocabularies that describe them, that is, maintaining the local identities and idiosyncrasies that define the repositories available on the web.

Although the debate on restricted or unrestricted access is at stake when we discuss subjects related to scientific data from which we should not be absent. Here we are especially interested in discussing the *modus operandi* of how this access can be realized and investigating the possibilities of relating these data through its contents, in a significant way in an environment in which the semantic technologies are already giving their first steps. Therefore, it is our interest to join efforts and go through one of the aspects that will enable such access, whether restricted or unrestricted, that data can be gathered by their conceptual similarities. In this approach, this study aims to investigate issues involving semantic compatibility and semantic interoperability, which are fundamental for intelligent retrieval not only in research environments but in the whole web.

3.0 The studies of semantic interoperability: contributions of information science and computer science

For W3C, semantic interoperability means enabling different agents, services and applications to exchange information, data, in addition to knowledge, notably, both on the web and beyond it. For the purpose of making this type of interoperability possible, the agents involved need to share the same vocabulary comprehended by all or to create correspondences or mapping between different vocabularies (Ushold and Menzel 2005).

According to Degoulet, Fieschi and Attali (1997, vol. 9, 5), in order to achieve semantic interoperability, it is necessary the compatibilization of these vocabularies from the processes of creating and interpreting messages, which depends mainly on three key factors:

- Terminology or nomenclature
  It is the way in which concepts are expressed in a controlled vocabulary having the referential status of a field of knowledge or activity.
- Explicit semantic relationships related to terminology
  Relationships can be organized in a semantic network. This network aims to structure the connections between the concepts of the domain, specifying their nature; synonymy connections, heteronymy, derivation, etc. (for example, SBP and systolic blood pressure or angina and angina pectoris for synonymy connections).
- Representation of concepts in messages
  In general, two different institutions model and represent concepts in different ways. Simple examples are age (modeling created based on the age or from the date of birth ...) and the address of a certain person (separation of street number, ZIP code, number of lines ...). We could cite innumerable cases, as the modeling of the existence of a symptom by stage (or severity) or simply by yes or no.

The search for interoperability between data and records extends across all domains, along with areas of knowledge and at all levels. It is evident, as an example among many others, its need in healthcare, where patients, diseases and drug data need to be exchanged between different institutions, especially using as an example, one of the points where the problem is reflected: the patient's chart. Although the institutions and professionals of ICT and healthcare of these institutions do not know how to conceptualize correctly what are interoperability standards or even ignore this concept, according to the work of Miranda and Pinto (2015).

Regardless of global efforts, whether governmental or from the research institutes themselves, to promote open science with free access to publications and scientific research data, as shown above, this attempt runs up against the difficulty of interoperating these data between repositories and between different fields of research and human knowledge. The formation of transdisciplinary knowledge networks, that will allow a leap forward in the development of sciences, scientific research and innovation, depends on the ability of researchers, through their publications and research data, to interoperate and intelligently and semantically share their research and results.

As Lévy (2014, 231) states, "so much information is available in the digital world that the obstacle to accessing it is due to this abundance itself: how to find the relevant information needle in the giant haystack of digital data?" According to the author, in order for this information economy not to lose the value of its main good, it must become a semantic information economy, since the value of information depends on its meaning. That is, the information economy should be able to modify the meaningful contexts, the practical environments where the meaning is. In other words, the information economy should be able to model the meaningful contexts, the practical environments where the meaning is. Since information availability is no longer a problem to be overcome, informational goods gain and lose value as a function of their significance and relevance to the communities that use them. Then, it is necessary to imagine and build sociotechnical devices that are cable of answering a central question for a given user: how to find the information that has the most value.

Thereby, by proposing the bases of the semantic web, Tim Berners-Lee and others were saying that the original web operation, which was created and developed by Berners-Lee himself, could no longer meet the wishes of its users (Berners-Lee et al. 2001). The page-based web whether static, as it is at the beginning or dynamic, could not sort out the new information retrieval issues that are demanded by users. The current web was designed to be used by humans and we need to evolve into a web used by intelligent software agents and intelligent devices. These bases and semantic guidelines for this new web paradigm includes the capacity to express meaning, to represent knowledge, to use ontologies and to have software agents as users.

Dealing with interoperability with the intention to disseminate, share and reuse data, records and information is not a new problem to be faced by humanity and the scientists. On the contrary, the need to share information from different sources, with form and meaning, has already been in the information science for some time. On the way to creating a useful memory, the Belgian Paul Otlet, since before the Second World War, had proposed the creation of a universal library, its indexation and also discussed its theoretical problems. When Otlet created the Universal Decimal Classification (UDC), he transformed the classification developed by Dewey into a faceted language and, in parallel, popularized in Europe the use of microfiche, which had already been used in the United States of America. In 1930, Otlet showed that the documentary universe was growing strongly and predicted that electronic technologies would make information ubiquitous. His unfinished project of constructing a collective and universal memory for mankind can be seen in his 1934 work, called *The Treaty of Documentation*, where the author proposes an intellectual network universe, accessed by a universal classification system, which would be mandatorily updated constantly by the links created by the users between the various documents. In this case, even before the innovative and revolutionary constructions proposed by Vannevar Bush, Douglas Engelbart and Ted Nelson, Otlet presented the project to formulate and build the principle of hypertextual interconnection (Lévy 2014).

In this quest for interoperability, knowledge organization systems are essential to represent the different environments of organizations, institutions and research groups that seek to disseminate and share their information. These KOS can vary greatly in different aspects, i.e., knowledge domain, language, granularity or structure, making it difficult to recover information between different vocabularies. The definitions presented by ISO 25964—Thesauri and Interoperability with other Vocabularies (ISO 2011; ISO 2013)—can help solve these problems. It covers the interoperability of thesaurus, taxonomies, ontologies, classification schemes and other KOS. Its norms and definitions focus mainly on the implementation of mappings and address the characteristics of the different KOS vocabularies that make these mappings possible. However, despite the importance of the existence and formulation of these standards, it is not possible to guarantee that different vocabularies spread around the world today on the web can interoperate immediately, hence the importance of studying and improving the compatibility processes proposed by CI and by CC, which we will see in the next sections of this article.

Thereby, there are also the issues addressed by information science, in order to solve the problem of intelligent information retrieval adding semantics to these searches through methodological procedures that do not focus only on a universal language, but that allows a metalanguage in which the "languages" of the various contexts could be respected. This is a problem to be solved nowadays. In other words, how to make possible an intelligent retrieval of information with such diversity and volume of data generation. This diversity comes from the space composed of different areas of knowledge, ontologies, vocabularies, languages and cultures. Even in controlled environments, like a single company, for example, we have difficulties in establishing or even creating a single vocabulary, that caters to all producers and consumers of information. Thus, it is our understanding that intelligent information retrieval occurs when it is possible to retrieve information from different bases and indexed by heterogeneous vocabularies without the need to intervene or modify the original bases.

In this sense, we will discuss initiatives of language compatibility in information science and alignment and semantic mapping in computer science, which introduce the notion of a metalanguage. This metalanguage must be formed as an intermediary language between the different source vocabularies allowing different actors to navigate through this language and, in a contextual and semantic way, to retrieve the desired information, no longer based on character strings, nevertheless, rather in their meaning, as we will see later.

3.1 The studies of semantic compatibility in information science

When debating about interoperability it is important to declare the high importance, recognized for a long time, of the study of "vocabularies" or "languages" in the organization of knowledge and information retrieval, within the field of library and information science, as pointed out by Lancaster (1986) and Buckland (1999), among others. Recent and growing studies in the field of semantic web, ontology engineering, metadata and open linked data and markup ontology languages that spread across several disciplines and scientific fields, clearly show the critical role played by vocabularies to represent, access and retrieve information and knowledge (Park 2006; Hovy et al. 2001).

The studies of compatibility and convertibility between languages in the field of information science, more specifically in those concerning documentary languages, from the 1960s of the last century, intended to create tools that would allow the conversion of different languages, in order to enable the user access to different data sources. It is considered that in the 1960s these studies stood out, due to the "information explosion" and the subsequent loss of information control, because of the proliferation of data sources. At the same time, the possibility of different US agencies accessing each other's content and a possible speeding of service delivery encouraged the studies about compatibility (Lancaster 1986). Hence, the research in the area was contextualized at a political moment when information reached strategic and decisive status. In the 1970s, there was a decline in the studies about compatibility, due to the dissemination of research automation. Then, in the 1980s, the studies about compatibility arose again, now applied to the computational environment and to automated language issues (semantic and syntactic problems). So, in the 1990s, these studies dealt with the integration between languages of computational systems with emphasis on the elaboration of ontologies, that are used as semantic tools for the purpose of allowing interoperability between systems (Campos 2010; Souza and Campos 2007).

Compatibility can be understood in two aspects: semantic and structural compatibility. Semantic compatibility is the ability of two vocabularies to have similarity between the conceptual contents of their terms while structural compatibility can be understood as linguistic compatibility (Glushkov et al. 1978). In either of these two aspects, we emphasize that in information science the search for compatibility of information systems and knowledge organization systems is not a minor problem and is strongly related to one of its essential objectives, which is to allow and simplify the connections between those who need sources of information on a given subject and the potential relevant documents to meet this demand. In order to be able to advance in the solution of this problem, especially in the current times of massive production of information sources and advancement of information technologies, it is necessary to understand that the compatibility of languages and vocabularies is not an intrinsic quality of these systems, but rather, it is a goal to be achieved (Maniez 1997), as discussed below.

Although the authors debate about compatibility in thesauri, this discussion of semantic and linguistic aspects in the process of compatibility can be applied to ontologies. From the methods of compatibility and conversion of languages based on the integration of vocabularies, two stand out eminently. They are the "thesaurus reconciliation" method, suggested by Neville (1970, 1972) and the "conceptual compatibility matrix," proposed by Dahlberg (1981b, 1983). Neville's method is based on the principle that concepts must be made compatible, (the conceptual contents of the descriptors, which are expressed by the definitions) not the descriptors alone. This method offers an intermediate language approach founded on the numerical coding of concepts through which it becomes possible to establish the conceptual equivalence of descriptors of different languages. It also considers that, within a common thematic area, vocabularies should cover the same concepts even though there may be different terms for naming the same concept among these different vocabularies. Based on this principle, their strategy is established on identifying similar concepts and encoding them uniquely in each vocabulary. This coding would then allow the keywords of a vocabulary to be mapped to other vocabularies of the same subject that shared this coding scheme. For this, Neville (1972) recommended an intermediate language way, which implements this numerical coding of concepts making possible to establish the conceptual equivalence of descriptors of different languages, denominated by the author as reconciliation, that is, the possibility of integration and approximation of systems, which contemplate the same type of literature, but adopt different thesauri (Neville 1972, 622). In the author's plan, each term in each participating thesaurus receives a code with the inclusion of remissive to the keywords. Moreover, a key to the encoding is given, so that your application in third-party keywords will generate the corresponding keyword in the source thesaurus. This key is different for each thesaurus; it functions as a conversion mechanism.

The establishment of correspondences between concepts does not necessarily imply a one-to-one correspondence. There may be cases, for example, where a more specific concept in one of the vocabularies is covered by a broader concept in the other vocabulary, or there may be cases where there is no correspondence in the other vocabulary for a particular concept of the vocabulary of origin. Neville (1970) embraces a thesaurus as a basis and follows eleven levels of matching between the terms of the two vocabularies analyzing them from the exact match between terms, use of synonyms and homonyms, among other things. Like Neville, Dahlberg (1981b) also introduces a mapping of the semantic and verbal potentialities between languages, which she calls "compatibility matrix."

Dahlberg (1981b) states that the most sophisticated use of a compatibility matrix would be if it worked as a black box, in which one would enter a descriptor in an indexing language (IL), and it would be possible to obtain the corresponding descriptor in the other ILs involved. Dahlberg's conceptual compatibility matrix (1981b) is a mapping of the semantic potentiality of the languages studied, providing the results of the analysis of the compatibility between languages through the semantic and structural perspectives. The compatibility between languages, as said by the author, comprises three phases: 1) conceptual coincidence—when two concepts combine their characteristics—degree of equivalence; 2) conceptual correspondence—two concepts combine most of their characteristics—similarity; and, 3) conceptual correlation—two concepts are correlated through mathematical symbols establishing a correlation measure when they have different levels of detail, or when the relation between them is not of similarity (Dahlberg 1981b).

While Neville (1972) propounds the creation of an intermediate language in which the potentialities of reconciliation between two languages are presented, Dahlberg (1981b) goes a little further, since, besides presenting principles for the mapping, she also presents a proposal, which introduces the concept of "compatibility rate" in which one can quantify the degree of compatibility between the languages under analysis. Additionally, Dahlberg (1981b) uses the term "ordered systems" to name the various "terminological tools" that should be compatible and conceptualizes them as any instrument used in the organization, description and knowledge retrieval, composed of verbal or notational expressions for concepts and their relations, arranged in an ordered way. Dahlberg cites as examples classification schemes, thesauri, subject headings or another identical instruments.

Thus, when adopting "ordered systems," it extends not only the concept but also the universe of application of the compatibility methodology, since it makes the definition of the instrument to be compatibilized flexible and can include the questions asked by the users analyzed within a context that presents an internal organization offered by the retrieval system. Another issue presented in this study is the role of the definition in the mapping of the semantic potentialities of compatibility, when the established correlation between ordered systems should not be performed only at the level of the terms and their descriptors. The terms are only the bearers of the information and can only display the information about the content of the represented concept if the necessary relationships have been established through a definition.

A fundamental point presented by Dahlberg (1981b) in her methodology is the need that for each term there is a kind of annotation about how the term is structured in a given ordered system; she called the this information about the term "concept record." Therefore, for two languages to be compatible, it is necessary that each establishes a register of concepts. According to Dahlberg (1983, 6):

> In establishing an ordering system one must attempt to accumulate the necessary knowledge about the referents by a "concept record" ... If an ordering system has not been developed with the help of concept records it is necessary to establish such records at the time when the said comparisons with other OSs should be made of each single class and concept on every level of the hierarchies foreseen in order that the analyzed conceptual data of the concepts in question can be introduced into the comparisons.

The register of concepts, as told by to the author (Dahlberg 1983, 6), consists of a series of necessary fields:

00 – running number;
01 – name of concept or class of concepts;
02 – notation;
03 – definition with indication of source;
04 – next broader concept (generic and partitive relationship);
05 – highest concept in hierarchy;
06 – subject(s) field of a concept;
07 – other names of a concept or class of concepts (synonyms);
08 – source of concept; code for ordering system and
09 – remarks concerning corresponding concepts in other ordering systems.

Furthermore, the author suggest possible alternatives to these necessary fields: i) the name of the concept in other languages; ii) the category of the form of the concept (form category), which indicates if this is an object, a process or a quality; iii) additional information about the concept; and iv) related concepts.

The concept record supports the conceptual comparison between ordered systems, which is generally founded on the analysis of the hierarchical structure in which the term is situated, in addition to its definition and possible comments, to conclude whether the terms are or are not corresponding to the same concept.

For the preparation of a compatibility matrix, the first step is the verbal or linguistic joint of the terms, which can be automated and recorded in a preliminary matrix. From the analysis of the percentage of terms that were possible to match, we evaluate the feasibility of continuing with the compatibilization of vocabularies. The linguistic compatibility, however, does not ensure that the coincidences found are actually a conceptual correspondence, due, for example, to the possibility of homonymy. In addition, if different languages use different nomenclature for terms with the same meaning or present concepts at different levels of detail, the linguistic match is not enough to detect such occurrences.

In this way, the preliminary matrix obtained must be complemented, in a second stage, by means of the analysis of the concepts so that a semantic correspondence can be established. It is at this moment that the concept register is established to be able to infer the level of conceptual compatibility, that is, conceptual coincidence, conceptual correspondence and conceptual correlation, as presented above. As a result of this complementary conceptual analysis, we obtain the final compatibility matrix, which establishes, additionally to the correspondence of the concepts, a measure of compatibility and the correspondence type (“<,” “>,” etc.) as it was mentioned previously.

Both Neville’s (1970, 1972) and Dahlberg’s (1981a, 1981b, 1983) studies, although stemming from thesaurus activities, are considered to be appropriate for the investigation of principles related to the semantic aspects of terminological compatibility in heterogeneous environments of scientific data, in addition to the fact that they present elements to discuss the level of conceptual similarity independent of the types and data formats involved.

Nowadays, a significant number of conceptual and structural methodologies for vocabulary creation resemble those found in the history of pre-web knowledge organization systems. What makes them different are the approaches supported by semantic technologies, in consonance with LOD principles. With the growing number of KOSs being published in standardized and machine-understandable formats, institutions can refine and augment their data from external sources, accomplishing a major achievement of reusability of these vocabularies (Zeng and Mayr 2019). But while we have a steadily increasing advancement in vocabulary-building and sharing technologies, especially semantic web-related technologies for automated processes, on the other hand, we still need to make great progress in the methods and strategies for applying these technologies. Therefore, to advance and effectively reach the possibility of working in an open environment aiming at the interconnection of heterogeneous data with different vocabularies, without having to change or intervene in these vocabularies, we highlight two aspects of great value: to rescue the previous efforts mentioned above concerning the classical methods of information science, and to establish its similarities and differences from modern approaches, especially those directed to automatic processes.

Even with the incredible advance of information technologies, we have several mapping processes done at a syntactic level, that is, words, phrases and context, instead of being done at a semantic level. Advances obtained by the new processes of artificial intelligence and machine learning have great potential in reducing conflicts and promoting interoperability, in particular semantic interoperability (Zeng 2019), but despite these advances, the mapping of concepts in a semantic manner is still a challenge for those who depend on automatic mapping.

As we can see from methods and processes covered in this section, our view is that seminal information science studies for vocabulary matching, even if they were created in a pre-web era, are essential to moving forward in solving the problem of web interoperability and in creating automatic compatibility processes.

3.2 The semantic alignment studies in computer science

Leiva-Mederos et al. (2017) emphasize that there are several mechanisms to tackle semantic interoperability and that one of them is semantic alignment. The authors believe that semantic alignment fundamentally means finding the correspondence between distinct vocabularies. Li, Yang and Liu (2008) show that semantic alignment is the sum of a variety of methods and can be applied to many different types of thematic domains.

In the field of computer science, related to the studies of languages for semantic representation, by the end of the 1990s, ontologies began to appear. Ontologies, as a terminological artifact, can be compatibilized in different ways, depending on the needs that must be met and their availability. In this work, aiming the comparison with the initiatives of information science, for the purpose of intelligent data retrieval, we will treat ontologies as terminologies with the goal of representation and information retrieval.

In this field, the literature has presented several forms of compatibilization (Bruijn et al. 2006; Pinto and Martins 2001; Ziegler and Dittrich 2004, vol. 12). In this paper we will investigate the concepts of alignment (Bruijn et al. 2006; Choi, Song and Han 2006), which allow us to generate a set of links between ontologies. The alignment process, unlike others (e.g., join, integration) in relation to its result, instead of generating an additional ontology, keeps the reused ontologies unchanged and in their source locations but generates a set of (links) between these ontologies. These links contain a set of information about how to make reused ontologies compatible and they are expressed in a separate (physically existing) persistent model.

Alignments can be used for various purposes, such as the transformation of one ontology into another, or to allow queries or searches on features described with ontologies (Euzenat et al. 2004). Thus, a search can be made, for example, with terms of an ontology (O1) over features described with a another ontology (O2), because the alignment allows the automatic matching of the questions to the ontology model mentioned.

The set of links expressed in a persistent model produced by the alignment process is a mapping between the ontologies. There are different definitions in the literature for mapping (Ding and Foo 2002). In the context of this work, we adopt mapping as a formal expression, that establishes a correlation between two elements of distinct ontologies and that is stored in a persistent model, which can be separated or incorporated into an ontology. The elements related by the mapping can be the classes (used in the context of this work as a synonym of term), the properties or the attributes contained in the ontologies.

The information contained in the mapping will depend on the type of semantic link found between the elements and the type of formalism used in the ontology to represent its semantics. For example, two elements may be similar (to different degrees), or one may be part of the other or they may have some other type of relationship, which is identified with the help of a domain expert.

In relation to formalism, mappings can also be used to: a) make different formats compatible in the representation of ontology class attributes (for example, an attribute size measured in centimeters or millimeters); b) overcome problems of expressiveness derived from the use of different languages; c) make the shape of a class compatible, for instance, the wing of an airplane can be modeled as a part of the airplane (in this case the wing itself is a class); or, d) as an attribute of the airplane class. Other aspects can also be considered in regards to mapping such as the cardinality of relationships established when mapping, which can be one to one, one to many or many to many (Volz 2008).

Mappings of similarity may express different degrees of similarity. These can be represented as a numerical attribute of the mapping (Felicíssimo and Breitman 2004; Kalfoglou and Schorlemmer 2003) or also by means of a relation, that indicates the type of similarity, such as "equivalent," "narrower than," "broader than" (Aleksovski, Kate and Harmelen 2006; Su 2004). In order to determine the degree of similarity, several factors are generally considered: linguistic similarity between terms, compatibility of attributes, positioning of the term in the hierarchical structure of the ontology, among others.

As for the strategy of mapping, there are three ways to implement it: i) between local ontologies; ii) between a global ontology and other local ontologies; or, iii) in an ontology that is the result of the joining of others (Choi, Song and Han 2006). A local ontology reflects the conceptualizations of a domain that are relevant to a particular community, thus mirroring the perspective under which that community represents its world (or part of it) according to its objectives (Bouquet et al. 2002). A global ontology is a combination of several local ontologies in a unified ontology, which in this way contains all the concepts of local ontologies.

In the first way to implement a mapping (i), there are two local ontologies mapped through a separate persistent model (Choi, Song and Han 2006). In the second form (ii), there is a global ontology and one or more local ontologies and a separate persistent model that maps the terms of the global ontology to the terms of the local ontologies. In the third form (iii), there is an ontology, which is the result of a process of joining others and a set of mappings that is stored within the ontology itself, the result of the junction process, in the form of axioms and which maps the terms of the ontologies imported by this final ontology.

Several ontology matching systems have been designed recently and most of them use element-level techniques that aim at lexical information as essential elements, but with only this simple string comparison approach is impracticable to get useful results. On the other hand, there are structure-level techniques that rely on the analysis of the neighborhood of two entities to determine their similarity. But both techniques present weakness as they must extract semantics from the lexical information of entities. A possible solution to address this problem is to rely on external sources like WordNet or Wikipedia in order to obtain semantic similarities among elements. The limitation of this approach is that many vocabularies cannot find their correspondences in WordNet or Wikipedia. Another possible solution is the use the technique of word embeddings similarity, that can represent words as vectors in a semantic space. Even if different methods and strategies have been used, in almost all cases, the intervention of a domain expert in the ontology alignment process is necessary to avoid inconsistencies, showing the limitations of the processes (Zhang et al. 2014; Dhouib et al. 2019; Ardjani at al. 2015).

3.3 Similarities and differences between the two approaches

Based on the exposed approaches and techniques, it is possible to observe some common aspects in these diverse views. Primarily, in the proposals resulting from information science, we can observe a strong concern with the compatibility of the concepts involved and their conceptual contents, not only with the character strings that represent the terms, in other words, their descriptors, since this is a determinant characteristic of this approach. When Dahlberg proposed the semantic mapping, through semantic potentiality, she went beyond the coincidence between the concepts, analyzing its correspondences and correlations and suggesting the creation of a conceptual compatibility matrix, as exposed above. Neville, on the other hand, considers this semantic view through the integration of different thesaurus systems with the creation of an intermediary language, making a conceptual equivalence of the descriptors, not necessarily from one to one.

These proposals, in spite of privileging a semantic and conceptual view, have as an important aspect the fact that they require a great intellectual and manual effort to be implemented, requiring the work of professionals who have the knowledge of the processes to be carried out, and of professional experts in the vocabulary domains to be aligned. With the exponential growth of data—in the internet post-creation period and especially in the web post-creation period—allied to the growth of the diversity of vocabularies and indexing systems spread all over the world, in different domains, different cultures and different languages, this intellectual and manual work could not, by itself, be capable of meeting in its completeness the tasks of integration and compatibility of the heterogeneous data.

Concurrently, different techniques are developed, mainly by studies linked and supported by computer science so that, supported through computer programs they can automate and move towards integrating heterogeneous data sources. Therefore, we need to guarantee that in this case, as well as almost everything in this interconnected world, the practical application of interdisciplinarity, essential in the eScience era, be possible. The use of working tools for large volumes of data and large vocabularies should use compatibilization procedures which consider conceptual concepts that deal with the semantic similarity between terms.

DeRidder (2007) highlights that the semantic heterogeneity is inevitable, especially when information systems grow in a decentralized way. We can notice that this is what happens with the web. The author, in the context of ontology interoperability, explains that in order to achieve interoperability, there are three basic approaches that can be chosen: a global ontology for which local ontologies are mapped, local mapping between ontologies' pairs when necessary and a combination of both strategies. For DeRidder, although a single global and heavyweight ontology is a preferential option from the computational point of view, on the other hand, obtaining a global agreement for this ontology seems to be unfeasible. He proposes that an approach based on generality layers, with a group of ontologies for this purpose, instead of only one, would be more likely to succeed.

Park (2006, 30) shows in her study results that point out the need for mediation mechanisms, that it is necessary in order to provide contextual relations between metadata elements and their definitions to ease the mapping processes that can reduce semantic ambiguity. She proposes that the use of concept networks may serve this purpose; however, she suggests that the development of this kind of solution requires additional studies concerning not only the metadata but also the mapping practices. In this context, she highlights that "problems in metadata mapping result from the absence of the context in which a metadata element name and its usage (i.e., definition) occur."

On the other hand, although human mediation and intervention are still necessary, Greenberg, Spurgin and Crystal (2006) point out that it is not possible to rely exclusively on traditional manually-generated metadata approaches, especially considering the huge number and volume of digital resources in need of these metadata. In a study related to this discussion, focused on metadata generation and that contributes to the understanding of automated processes in this area, Greenberg (2004) positively concludes that there is a potential of these automatic processes for metadata creation.

For Martínez-Ávila et al. (2018), the semantic web, by proposing the creation of machine-readable data by different communities, has led to isolated information systems that deal with their own knowledge and are specific to their domain, limiting potential interoperability since publishing data as linked open data (LOD) is a big step towards making data available and interconnected but not enough to achieve full interoperability. The technology to link this data exists, but performing the linking process requires expert knowledge and does not happen completely automatically. Moving to a LOD environment requires extended interoperable vocabularies for better organization of large data, that is, we need producers and users who are able to provide and use richer semantics and structure in the vocabularies that are used to consume, describe and publish data. This situation could benefit from a greater application of knowledge organization principles (Martínez-Ávila et al. 2018).

According to Chan and Zeng (2002), as the advanced computerized processes were evolving, to achieve or improve this interoperability among vocabularies, computer technologies begun to be used to benefit from this interconnected environment. Thus, some new methods and some conventional ones have come to form a list of widely accepted methods.

In the method called "derivation" or "modeling," a simpler or specialized vocabulary is developed from a more comprehensive one chosen as a model. In "translation" or "adaptation," a vocabulary is controlled and developed from terms translated from another language. In the intellectual process of mapping a system, it is developed from the establishment of the equivalence of terms between different controlled vocabularies, requiring a great intellectual effort. This method can be migrated to use the computer technology in its implementation (Chan and Zeng 2002). In "alignment," for each entity (which can be a concept, an instance or even a relation) in the first vocabulary, we try to find a correspondent entity in a second vocabulary with the same intended meaning in the second vocabulary (Ehrig 2007).

In this environment where computer and information technologies are used, vocabulary alignment and mapping emerge as the main methods to be used in the vocabulary compatibilization process, because they can be applied to a variety of usages and scenarios such as vocabularies and ontologies importing, data schemes integration and links of different versions of ontologies. Thus, the ontologies alignment is a crucial condition nowadays to make possible the interoperability between semantic systems when it allows that the individual relations between elements of multiple ontologies to be identified (Ehrig 2007; Euzenat and Shvaiko 2007).

It is possible to observe in the scientific publishing of information science and computer science areas, different meanings and conceptualizations for the term's alignment and mapping (Ding and Foo 2002; Su and Gulla 2006; Ehrig 2007; Fielding et al. 2004). For some authors, such as Ehrig (2007), the alignment is a preliminary stage where it is possible to detect at which point the ontologies overlap each other and where they can be connected to each other. For this compatibility type, the terms to be aligned have the same meaning, but the process foresees that different relations be used and not only equivalence. In this alignment point of view, different relations such as identity, subsumption, instantiation and orthogonality can be used in the links between the terms (Ehrig 2007). For Euzenat and Shvaiko (2007), alignment is based on a term's matching process where it is sought to identify those terms that express similar concepts, being possible to associate values to express their degree of reliability and similarity (Ehrig 2007).

Regarding mapping, some authors, such as Pâslaru-Bontas (2007), consider it as a synonym to alignment, claiming that both terms are used in an interchangeable way in literature and define it as a process of relationship creation between the corresponding elements. Ehrig (2007) points out differentiations, claiming that mapping is a function between two ontologies represented by axioms that describe how to express concepts, relations or instances in terms of another ontology, focusing on the representation and establishment of the relations for certain tasks, whereas alignment only identifies the relation between ontologies.

Campos (2011) expresses a comprehensive definition when considering mapping as a formal expression that represents a function between two elements, equivalent or not, of distinct ontologies and allows the compatibility of such elements through actions for the execution of a determined task, being stored as a persistent model (which exists physically) and which can be separated or incorporated to one ontology. It is important to notice that both processes observed here have as a common characteristic to keep the original ontologies or vocabularies unaltered and in their places of origin but are able to generate links between these terms of these ontologies, expressing the type of relation that links the terms to be aligned.

Park (2006) and Boteram and Hubrich (2010) highlight the importance of the semantic issue in mapping initiatives. In this regard, we can notice here the relevance of Dahlberg's (1983) proposal of the concept registry. DeRidder (2007), in turn, highlights the usefulness of a central vocabulary connecting vocabularies to be mapped. At this point, we can observe the relevance nowadays of the studies for the construction of intermediary languages and how the information science efforts could be explored to support concrete vocabulary compatibility initiatives. We consider here the relevance of citing a bridge to be considered between the semantic web trends, using the computer science view and the traditional practices of vocabulary control and knowledge organization and representation from information science. This bridge is SKOS, which is one of the most important semantic web specifications for application in archives, libraries, and documentation centers. It emerged as a W3C recommendation and offers a data model to represent the structure and content of conceptual schemas such as thesauri, classification systems and taxonomies. In its official documentation, SKOS points to the use of semi-formal knowledge organization systems to differentiate them from ontologies that support complex automatic reasoning processes, i.e., its orientation is towards indexing and retrieving information rather than more complex processing, which can be achieved, for example, with OWL (González 2014).

4.0 Conclusions

The increasing digital data volume produced in all areas of human acting, notably in the production of scientific data, requires urgent production of scientific data action by researchers and professionals, especially in the information science and computer science areas so that the data can be used effectively in an intelligent way to generate knowledge.

Therefore, today we have an imperative need to walk towards the creation of methods and techniques to produce semantic compatibility without necessarily being a strong human intervention since manual processes have already proven to be incapable of solving this problem. That is, we need to create requirements and models that allow the development of software applications able to align the heterogeneous vocabulary terms based not only on their codification formats but also on their semantic content.

We already have a number of techniques and technologies that allow the creation of interconnected data sources using resources such as universal resource identifiers, open linked data, etc., but on the other hand we have a huge and overwhelming set of data indexed by equally huge numbers of different vocabularies that represent local and peculiar idiosyncrasies that cannot be linked, whether due to technical, financial, administrative or even political difficulties, negatively affecting the retrievability of this information. In this sense, the growth and diversity of the web with its heterogeneous data sources, indexed by different vocabularies and distributed throughout different countries, companies and organizations, raises the need to discuss paths, requirements, technologies, guidelines and solutions that point to the possibility of automation considering the compatibility between different environments, not with the creation of a new vocabulary to be used, because this is not always possible, but with the adoption of a metalanguage. This metalanguage must be formed as an intermediate language between the different source vocabularies enabling different actors to navigate this language and retrieve, contextually and semantically, the intended information no longer based on characters strings but on its meaning.

The techniques to implement the matching and compatibility of ontologies and vocabularies are complex processes that aim at reducing different representations, different perspectives and different modeling views. These techniques are a growing tendency since they are probably the best resources to encode the meaning of information, with the last decades being a period of extensive researches in this field. Nowadays these researches present an increase, and new publications where this problem is approached are continuously being published, reflecting the global interest in this matter (Otero-Cerdeira, Rodríguez-Martínez and Gómez-Rodríguez 2015).

Besides the proposals discussed here, we also verified the appearance, at the end of the last decade, of an investigation brought by the philosopher Pierre Lévy to elaborate a formal language called information economy metalanguage (IEML), which in a sense meets what we have been discussing so far. The IEML is proposal of a universal semantic addressing capable of indexing all the digital documents and has three basic procedures: 1) each distinct concept must have a single address; 2) the existence of a system of coordinated semantics must be open to any concept and relations between concepts (ontologies) regardless the cultural environment in which these concepts are created and transformed, without privileges and exclusions; and, 3) it must allow a group of mathematically defined operations (possible to be automated) on the semantic addresses, such as symmetries operations, logical inferences or comprehension, among others (Lévy 2007, 2009). The IEML semantic sphere presented by Lévy suggests a coordinated system where, first, the meanings are addressed and, then, represent movements and variations of the meanings through calculating functions. The IEML situates itself on the intersection between human languages and formalized languages and works as a categorizing system of all the culture, which results in development through digital means (Lévy 2014).

In this regard, we consider that in the investigation of such proposals and in the search for others it will be possible to identify methodological criteria that can allow an optimized appropriation considering the positive and negative points on the application of such methodologies in heterogeneous environment in which the scientific data present themselves.

References

Akmon, Dharma, Ann Zimmerman, Morgan Daniels, and Margaret Hedstrom. 2011. "The Application of Archival Concepts to a Data-Intensive Environment: Working with Scientists to Understand Data Management and Preservation Needs." *Archival Science* 11: 329-48.

Aleksovski, Zharko, Warner Ten Kate, and Frank Van Harmelen. 2006. "Exploiting the Structure of Background Knowledge Used in Ontology Matching." In *OM'06: Proceedings of the First International Conference on Ontology Matching – November 2006,* ed. Pavel Shvaiko, Jérôme Euzenat, Natalya Noy, Heiner Stuckenschmidt and Richard Benjamins. CEUR Workshop Proceedings 225, Aachen: CEUR-WS.org, 13-24.

Ardjani, Fatima, Djelloul Bouchiha, and Mimoun Malki. 2015. "Ontology-Alignment Techniques: Survey and Analysis." *International Journal of Modern Education and Computer Science* 7, no. 11: 67-78. doi:10.5815/ijmecs.2015.11.08

Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. "The Semantic Web." *Scientific American* 284, no. 5: 34-43.

Borgman, Christine L. 2010. "Research Data: Who Will Share What, With Whom, When and Why?". RatSWD Working Paper 161. doi:10.2139/ssrn.1714427

Boteram, Felix and Jessica Hubrich. 2010. "Specifying Intersystem Relations: Requirements, Strategies, and Issues." *Knowledge Organization* 37: 216-22. doi:10.5771/0943-7444-2010-3-216

Bouquet, Paolo, Antonia Donà, Luciano Serafini, and Stefano Zanobini. 2002. "Contextualized local Ontologies Specification via CTXML." In *Working Notes of the AAAI 2002 Workshop on Meaning Negotiation, Edmonton, Canada, July 28*, ed. Paolo Bouquet. Menlo Park: AAAI Press.

Brodaric, Boyan and Mark Gahegan. 2010. "Ontology use for Semantic e-Science." *Semantic Web* 1: 149-53.

Bruijn, Jos, Marc Ehrig, Cristina Feier, Francisco Martíns-Recuerda, Francois Scharffe, and Moritz Weiten. 2006. "Ontology Mediation, Merging, and Aligning." In *Semantic Web Technology: Trends and Research in Ontology-based Systems,* ed. D. Davies, R. Studer, and P. Waren. Chichester: John Wiley & Sons, 95-112. doi:10.1002/047003033X.ch6

Buckland, Michael. 1999. "Vocabulary as a Central Concept in Library and Information Science." In *Digital Libraries: Interdisciplinary Concepts, Challenges, and Opportunities: Proceedings of the Third International Conference on Conceptions of Library and Information Science (CoLIS3), 23-26 May 1999, Dubrovnik, Croatia*, ed. T. Arpanac et al. Zagreb: Lokve, 3-12.

Burgstaller-Muehlbacher, Sebastian, Andra Waagmeester, Elvira Mitraka, Julia Turner, Tim Putman, Justin Leong, Chinmay Naik, Paul Pavlidis, Lynn Schriml, Benjamin M. Good, and Andrew I. Su. 2016. "Wikidata as a

Semantic Framework for the Gene Wiki Initiative." *Database (Oxford)* 2016, 1-10. doi:10.1093/database/baw 015

Bush, Vannevar. 1945. "As We May Think." *Atlantic Monthly* 176, no. 1:112-24.

Campos, Linair Maria. 2011. "Diretrizes para definição de recorte de domínio no reúso de ontologias biomédicas: uma abordagem interdisciplinar baseada na análise do compromisso ontológico." PhD diss., Universidade Federal Fluminense.

Campos, Maria Luiza A. 2007. "Integração de ontologias: o domínio da bioinformática." *RECIIS* 1, no. 1: 117-21.

Campos, Maria Luiza A. 2010. "O papel das definições na pesquisa em ontologia." *Perspectivas em ciência da informação* 15: 220-38.

Chan, Lois Mai and Marcia Lei Zeng. 2002. "Ensuring Interoperability among Subject Vocabularies and Knowledge Organization Schemes: A Methodological Analysis." *IFLA Journal* 28: 323-7.

Choi, Namyoun, Il-Yeol Song, and Hyoil Han. 2006. "A Survey on Ontology Mapping." *ACM SIGMOD Record* 35, no. 3: 34-41. doi:10.1145/1168092.1168097

Dahlberg, Ingetraut. 1981a. "Conceptual Definitions for Interconcept." *International Classification* 8: 16-22.

Dahlberg, Ingetraut. 1981b. "Towards Establishment of Compatibility between Indexing Languages." *International Classification* 8, no. 2: 86-91. doi:10.5771/0943-7444-1981-2-86

Dahlberg, Ingetraut. 1983. "Conceptual Compatibility of Ordering Systems." *International Classification* 10: 5-8.

Davenhall, Clive. 2011. "Instalment on Scientific Metadata." In *DCC I Digital Curation Manual*, ed. Joy Davidson.

Degoulet, Patrice, Marius Fieschi, and Christophe Attali. 1997. "Les enjeux de l'interopérabilité sémantique dans les systèmes d'information de santé." In *Informatique et gestion médicalisée,* ed. François Kohler. Informatique et santé 9. Paris: Springer, 203-12.

DeRidder, Jody L. 2007. "The Immediate Prospects for the Application of Ontologies in Digital Libraries." *Knowledge Organization* 34: 227-46. doi:10.5771/0943-7444-2007-4-227

Dhouib, Molka Tounsi, Catherine Faron Zucker and Andrea G. B. Tettamanzi. 2019. "An Ontology Alignment Approach Combining Word Embedding and the Radius Measure." In *Semantic Systems: The Power of AI and Knowledge Graphs Semantic Systems; The Power of AI and Knowledge Graphs:15th International Conference, SEMANTiCS 2019, Karlsruhe, Germany, September 9–12, 2019, Proceedings*, ed. Maribel Acosta, Philippe Cudré-Mauroux, Maria Maleshkova, Tassilo Pellegrini, Harald Sack and York Sure-Vetter. Lecture Notes in Computer Science 11702. Berlin: Springer, 191-7. doi:10.1007/978-3-030-33220-4_14

Ding, Ying and Schubert Foo. 2002. "Ontology Research and Development: Part 2 – A Review of Ontology Mapping and Evolving." *Journal of Information Science* 28: 375-88. doi:10.1177/016555150202800503

Ehrig, Marc. 2007. *Ontology Alignment: Bridging the Semantic Gap. Semantic Web and Beyond*. Secaucus, NJ: Springer US. doi:10.1007/978-0-387-36501-5

Euzenat, Jérôme and Pavel Shvaiko. 2007. *Ontology Matching.* Berlin: Springer.

Euzenat, Jérôme, Thanh Le Bach, Jesús Barrasa, Paolo Bouquet, Jan De Bo, Rose Dieng, Marc Ehrig, Manfred Hauswirth, Mustafa Jarrar, Ruben Lara, Diana Maynard, Amedeo Napoli, Giorgos Stamou, Heiner Stuckenschmidt, Pavel Shvaiko, Sergio Tessaris, Sven Van Acker and Ilya Zaihrayeu. 2004. "D2.2.3: State of the Art on Ontology Alignment." http://knowledgeweb.semanticweb.org/semanticportal/deliverables/D2.2.3.pdf

Felicíssimo, Carolina H. and Karin K. Breitman. 2004. "Taxonomic Ontology Alignment: An Implementation." http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=E97414C34AFF1E7542B292ECFEF54DA1?doi=10.1.1.86.7321&rep=rep1&type=pdf

Fielding, James M., Jonathan Simon, Werner Ceusters, and Barry Smith. 2004. "Ontological Theory for Ontological Engineering: Biomedical Systems Information Integration." In *Principles of Knowledge Representation and Reasoning: Proceedings of the Ninth International Conference (KR2004), Whistler, Canada, June 2-5, 2004*, ed. Didier Dubois, Christopher A. Welty and Mary-Anne Williams. Palo Alto, CA: AAAI Press, 117-20. https://www.aaai.org/Papers/KR/2004/KR04-014.pdf

Glushkov, V. M., E. F. Skorokhod'ko, and A. A. Strongnii. 1978. "Evaluation of the Degree of Compatibility of Information Retrieval Languages of Document Retrieval Systems." *Automatic Documentation and Mathematical Linguistics Journal* 12: 18-26.

González, José Antônio M. 2014. *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: EDUFBA.

Gray, Jim, David T. Liu, Maria Nieto-Santisteban, Alexander Szalay, David J. DeWitt, and Gerd Heber. 2005. "Scientific Data Management in the Coming Decade." *ACM SIGMOD Record* 34, no. 4: 34-41. doi:10.1145/1107499.1107503

Greenberg, Jane, Kristina Spurgin, and Abe Crystal. 2006. "Functionalities for Automatic Metadata Generation Applications: A Survey of Experts' Opinions." *International Journal of Metadata Semantics and Ontologies* 1: 3-20. doi:10.1504/IJMSO.2006.008766

Greenberg, Jane. 2004. "Metadata extraction and harvesting." *Journal of Internet Cataloging* 6, no. 4: 59-82.

Harvey, D. Ross. 2010. *Digital Curation: A How-to-do-it Manual*. London: Facet.

Hey, Tony and Jessie Hey. 2006. "E-Science and its Implications for the Library Community." *Library Hi Tech* 24: 515-28. doi:10.1108/07378830610715383

Hey, Tony, Stewart Tansley, and Kristin Tolle. 2009. "Jim Gray on e-Science: A Transformed Scientific Method." In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, ed. Tony Hey, Stewart Tansley and Kristin Tolle. Redmond: Microsoft Research, xvii-xxxi.

Hovy, Eduard, Nancy Ide, Robert Frederking, Joseph Mariani, and Antonio Zampolli. 2001. *Multilingual Information Management: Current Levels and Future Abilities.* Linguistica Computazionale 14-15. Pisa: Instituti Editoriali e Poligrafici Internazionali.

ISO (International Organization for Standardization). 2011. *Thesauri for Information Retrieval.* Part 1 of *Thesauri and Interoperability with Other Vocabularies.* International Standards ISO 25964-1:2011(E). Geneve : ISO.

ISO (International Organization for Standardization). 2013. *Interoperability with Other Vocabularies*. Part 2 of *Thesauri and Interoperability with Other Vocabularies.* International Standards ISO 25964-2:2013(E). Geneve : ISO.

Kalfoglou, Yannis and Marco Schorlemmer. 2003. "Ontology Mapping: The State of the Art." *The Knowledge Engineering Review* 18, no. 1: 1-31. doi:10.1017/S0269888903000651

Lancaster, Frederick W. 1986. *Vocabulary Control for Information Retrieval*. Arlington, VA: Information Resources Press.

Lee, Christopher A. and Helen R. Tibbo. 2007. "Digital Curation and Trusted Repositories: Steps Toward Success." *Journal of Digital Information* 8, no. 2. https://journals.tdl.org/jodi/index.php/jodi/article/viewArticle/229/183

Leiva-Mederos, Amed, Jose A. Senso, Yusniel Hidalgo-Delgado and Pedro Hipola. 2017. "Working Framework of Semantic Interoperability for CRIS with Heterogeneous Data Sources." *Journal of Documentation* 73, no. 3: 481-99. doi:10.1108/JD-07-2016-0091

Lévy, Pierre. 2007. "Abrir o espaço semântico em prol da inteligência coletiva." *Revista eletrônica de comunicação informação & inovação em Saúde* 1: 129-40.

Lévy, Pierre. 2009. "Pierre Lévy e a linguagem IEML." Interview given to Carlos Nepomuceno. *Webinsider*, (blog). https://webinsider.com.br/entrevista-pierre-levy-e-a-linguagem-ieml/.

Lévy, Pierre. 2014. *A esfera semântica: tomo I: computação, cognição, economia da informação*. São Paulo: Annablume.

Li, Shuming, Zongkai Yang, and Qingtang Liu. 2008. "Research of Metadata Based Digital Educational Resource Sharing." In *Proceedings International Conference on Computer Science and Software Engineering CSSE 2008 12-14 December 2008, Wuhan, Hubei, China.* Los Alamitos, CA: IEEE Computer Society, 828-31. doi:10.1109/CSSE.2008.232

Lyon, Liz. 2007." Dealing with Data: Role, Rights, Responsibilities and Relationships. Consultancy Report." http://opus.bath.ac.uk/412/1/dealing_with_data_report-final.pdf.

Maniez, Jacques. 1997. "Database Merging and the Compatibility of Indexing Languages." *Knowledge Organization* 24: 213-24.

Martínez-Ávila, Daniel, Richard P. Smiraglia, Rick Szostak, Andrea Scharnhorst, Wouter Beek, Ronald Siebes and Laura Ridenour. 2018. "Classifying the LOD Cloud: Digging into the Knowledge Graph." *Brazilian Journal of Information Studies*: *Research Trends* 12, no. 4: 6-10.

Miranda, Nelson Julio de O. and Virginia B. Pinto. 2015. "Prontuário eletrônico do paciente: padronização e interoperabilidade." http://200.20.0.78/repositorios/handle/123456789/2966?show=full

Neville, Hugh H. 1970. "Feasibility Study of a Scheme for Reconciling Thesauri Covering a Common Subject." *Journal of Documentation* 4, no. 26: 313-36. doi:10.1108/eb026502

Neville, Hugh H. 1972. "Thesaurus Reconciliation." *Aslib Proceedings* 24: 620-6.

NFS (National Science Foundation). 2007. "Cyberinfrastructure Vision for 21st Century Discovery". https://www.nsf.gov/pubs/2007/nsf0728/nsf0728_1.pdf

Otero-Cerdeira, Lorena, Francisco J. Rodríguez-Martínez and Alma Gómez-Rodríguez. 2015. "Ontology Matching: a Literature Review." *Expert Systems with Applications Journal* 42: 949-71.

Park, Jung-ran. 2006. "Semantic Interoperability and Metadata Quality: An Analysis of Metadata Item Records of Digital Image Collections." *Knowledge Organization* 33: 20-34.

Pâslaru-Bontas, Elena. 2007. "A Contextual Approach to Ontology Reuse: Methodology, Methods, and Tools for the Semantic Web." PhD diss., Freie Universität, Berlin.

Payette, Sandra, Christophe Blanchi, Carl Lagoze, and Edward A. Overly. 1999. "Interoperability for Digital Objects and Repositories: the Cornell/CNRI Experiments." *D-Lib Magazine* 5, no. 5. http://www.dlib.org/dlib/may99/payette/05payette.html.

Pilat, Dirk and Yukiko Fukasaku. 2007. "OECD Principles and Guidelines for Access to Research Data from Public Funding." *Data Science Journal* 6. OD4–OD11. doi: http://doi.org/10.2481/dsj.6.OD4

Pinto, Helena Sofia and João P. Martins. 2001. "A Methodology for Ontology Integration." In *K-CAP '01: Proceedings of the 1st International Conference on Knowledge Capture, Victoria, British Columbia, Canada, October 2001,* ed. Yolanda Gil, Mark Musen and Jude Shavlik.

New York: Association for Computing Machinery, 131-8. doi:10.1145/500737.500759

Rayward, W. Boyd. 1991. "The Case of Paul Otlet, Pioneer of Information Science, Internationalist, Visionary: Reflections on Biography." *Journal of Librarianship and Information Science* 23, no. 3: 135-45. doi:10.1177/096100069102300303

Robredo, Jaime. 2011. "Do documento impresso à informação nas nuvens: reflexões." *Liinc em Revista* 7, no. 1: 19-42.

Sales, Luana F. 2014. "Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares." PhD diss., Universidade Federal do Rio de Janeiro. https://repositorio.ibict.br/bitstream/123456789/874/1/LUANA%20SALES%20D.pdf.

Souza, Joyce Cleide E. and Maria Luiza A. Campos. 2007. "Avaliação de linguagem de indexação aplicada à informação jornalística: um estudo de caso." http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/161/GT2--105.pdf?sequence=1

Su, Xiaomeng and Jon Atle Gulla. 2006. "An Information Retrieval Approach to Ontology Mapping." *Data and Knowledge Engineering* 58: 47-69.

Su, Xiaomeng. 2004. "Semantic Enrichment for Ontology Mapping." PhD diss., Norwegian University of Science and Technology. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.215.1832&rep=rep1&type=pdf

Taylor, J. M. 2001. "The UK e-Science Programme." Paper presented at the e-Science London Meeting. https://web.archive.org/web/20090720233315/http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/research/esci/jtaylor.pdf

Ushold, Mike and Christopher Menzel, eds. 2005. "Semantic Integration & Interoperability Using RDF and OWL." https://www.w3.org/2001/sw/BestPractices/OEP/SemInt/

Volz, Raphael. 2008. *Semantics @ Work: Ontology Management – Tools and Techniques.* Morrisville: Lulu.com.

Zeng, Marcia Lei and Mayr, Philipp. 2019. "Knowledge Organization Systems (KOS) in the Semantic Web: a Multidimensional Review." *International Journal on Digital Libraries* 20: 209-30.

Zeng, Marcia Lei. 2019. "Interoperability." *Knowledge Organization* 46: 122-46. doi:10.5771/0943-7444-2019-2-122

Zhang,Yuanzhe, Xuepeng Wang, Siwei Lai, Shizhu He, Kang Liu, Jun Zhao and Xueqiang Lv. 2014. "Ontology Matching with Word Embeddings." In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*: *13th China National Conference, CCL 2014, and Second International Symposium, NLP-NABD 2014, Wuhan, China, October 18-19, 2014. Proceedings,* ed. Maosong Sun, Yang Liu and Jun Zhao. Cham: Springer, 34-45. doi:10.1007/978-3-319-12277-9_4

Ziegler, Patrick and Klaus R. Dittrich. 2004. "Three Decades of Data Integration: All Problems Solved?" In *Building the Information Society: IFIP 18th World Computer Congress Topical Sessions 22–27 August 2004 Toulouse, France*, ed. Renè Jacquart. IFIP International Federation for Information Processing book series 156. Boston, MA: Springer, 3-12. doi:10.1007/978-1-4020-8157-6_17-6_1