ABSTRACT

According to the FAIR guiding principles, one of the central attributes for maximizing the added value of information artifacts is interoperability. In this paper, I discuss the importance, and propose a characterization of the notion of Semantic Interoperability. Moreover, I show that a direct consequence of this view is that Semantic Interoperability cannot be achieved without the support of, on one hand, (i) ontologies, as meaning contracts capturing the conceptualizations represented in information artifacts and, on the other hand, of (ii) Ontology, as a discipline proposing formal methods and theories for clarifying these conceptualizations and articulating their representations. In particular, I discuss the fundamental role of formal ontological theories (in the latter sense) to properly ground the construction of representation languages, as well as methodological and computational tools for supporting the engineering of ontologies (in the former sense) in the context of FAIR.

1. INTRODUCTION

In their seminal paper, Wilkinson et al. [1] propose the so-called FAIR guiding principles as the cornerstone for maximizing the added value of information artifacts. The principles are organized around the four
general notions of Findability, Accessibility, Interoperability and Reusability (hence the acronym). Here, I focus on Interoperability. Firstly, we should be reminded that interoperability is not about finding ways to connect data artifacts but ultimately about affording the interoperation of humans mediated by these artifacts\(^2\). Information artifacts are instruments used by humans to harmonize their conceptualizations and, hence, interoperability approaches succeed to the extent that they can safely connect these conceptualizations.

Secondly, in the description of the FAIR principles, interoperability is described in a recursive manner stipulating that, in order for an information artifact to be interoperable, it should be described using semantic resources that follow FAIR principles. In other words, these artifacts can only be interoperable if they are grounded on artifacts that are themselves interoperable. In this article, I discuss the role of Formal Ontology, as a discipline, and of representation languages based on formal ontological principles, for grounding this entire enterprise.

The remainder of this paper is organized as follows. In Section 2, I discuss the importance of information integration and information systems interoperation, and the challenges of our current scenario in which the information is needed for addressing fundamental questions exists but is dispersed in multiple silos. In Section 3, I propose a theoretical characterization of the notion of semantic interoperability, as a relation between worldviews, or more technically, a relation between Conceptualizations. Finally, in Section 4, I defend a view of ontologies as kinds of “meaning contracts”, i.e., as artifacts that precisely characterize a given domain conceptualization. Moreover, I discuss the essential role of Formal Ontology and of ontology-driven representation languages for addressing this semantic interoperability challenge.

2. THE DATAVERSE IS A WORLD OF SILOS

Information is the foundation of all rational decision-making. Without the proper information, individuals, organizations, communities and governments can neither systematically take optimal decisions nor understand the full effect of their actions. In the past decades, information technology has played a fundamental role in automating an increasing number of information spaces. Simultaneously, there has been a substantial improvement in information access, motivated not only by advances in communication technology, but also by more recent demands on transparency and public access to information.

Despite these advances, most of these automated spaces remained as independent components in large and increasingly complex silo-based architectures. The problem is that, nowadays, several of the critical questions in large corporations, governments and scientific communities can only be answered by precisely connecting pieces of information distributed over these silos.

\(^2\) In the FAIR literature, authors commonly speak of semantic interoperability involving interoperation between humans, between humans and machines, and between machines. In the spirit of the semiotic engineering literature [2], I defend here that, with the possible exception of a scenario in which by “machine” we mean strong artificial intelligence (AI), semantic interoperability is always about interoperation with meaning preservation between humans, even in the cases in which these are mediated by machines and information structures. In Section 3, I elaborate on the notion of semantic interoperability defended in this paper.
An example illustrating this point is put forth by Wilkinson et al. [1]:

“Suppose a researcher has generated a dataset of differentially-selected polyadenylation sites in a non-model pathogenic organism grown under a variety of environmental conditions that stimulate its pathogenic state. The researcher is interested in comparing the alternatively—45 polyadenylated genes in this local dataset, to other examples of alternative-polyadenylation, and the expression levels of these genes—both in this organism and related model organisms—during the infection process. Given that there is no special-purpose archive for differential polyadenylation data, and no model organism database for this pathogen, where does the researcher begin?”

Now, suppose that the information required to answer this question exists “in the ether” but also that (as it is usually the case) it only exists in dispersed forms in a number of autonomous information silos. As a consequence, despite the increasing amount of information produced, as well as the improvements in information access, answering such critical questions is still extremely hard. In practice, they are still answered in a case-by-case fashion and still require a significant amount of human effort, which is slow, costly and error-prone. The problem of combining independently conceived information spaces and providing unified analytics over them is termed the problem of Interoperability [3].

Wilkinson et al. approach this thought experiment by considering (in the case the appropriate data sets exists) the multiple aspects of data set discovery, of operational and authorship rights over them, of formatting, and of integration. In this paper, I focus on the latter. Having access to data sets as well formatting issues are indeed aspects connected to interoperability. The former is connected to physical or communication interoperability, i.e., to how we can connect networked systems to allow for distributed access to information in heterogeneous computational platform. The latter is connected to syntactical interoperability, i.e., to how we can agree on standard syntactical structures for symbol processing that can be shared among parties. We have managed to make substantial advances in both aspects in the past decades and having heterogeneous networked systems that exchange information in standardized formats (e.g., XML) is state of the practice. A much more difficult problem that we are far from solving in a larger scale is that of Semantic Interoperability [3].

3. INFORMATION STRUCTURES AND SEMANTIC INTEROPERABILITY

I subscribe here to the so-called representation view of information systems [4]. Following this view, an information system is a representation of a certain conceptualization of reality. To be more precise, an information system contains information structures that represent abstractions over certain portions of reality, capturing aspects that are relevant for a class of problems at hand. There are three direct consequences of this view.

Firstly, it is that all information systems make ontological commitments. This has been discussed by several authors [3, 5], but it was even clear in the first paper to mention the term “ontology” in computer science, namely, the classical Another Look at Data by George Mealy [6]. In that paper, Mealy defends that “data are fragments of a theory of the real world (my emphasis), and data processing juggles
representations of these fragments of theory [...] The issue is ontology, or the question of what exists”. For an information structure to represent a conceptualization, it must commit to the existence of the entities constituting that conceptualization. For example, an information system that records information about organ transplants commits to a theory about the existence of certain entities such as persons, surgeons, transplants, donors, donees (organ recipients), etc. This is illustrated in Figure 1. Let us postpone questions of notation for now. The important point here is that this commitment to a particular ontology of transplants is what defines the real-world semantics [3, 7] of that information structure and this is inevitable, even if the designers of those information systems are not aware of this commitment. Paraphrasing Collier [8], I defend that the opposite of ontology is not non-ontology, but just bad ontology.

![Figure 1. An information system and its inevitable Ontological Commitment.](image)

Secondly, a direct consequence of this view is that the quality of an information system directly depends on how truthful its information structures are to the aspects of reality it purports to represent. These structures must represent all the relevant aspects of the underlying conceptualization in an unambiguous way and constrain the possible states of that information system to the states that represent intended state of affairs according to that conceptualization [9, 10]. For example, they should proscribe the existence of data populations in that transplant system that reflect state of affairs in which the donor, the donee and the surgeon involved in a transplant are the same person! This is a fundamental issue for semantic interoperability (as we will see soon) because, if these information structures are under-constrained [3, 10], we can have two information system agreeing on their possible data populations without having them agreeing on their

---

*Assuming that this is indeed an unintended state of affairs in this domain.*
intended populations. This is the so-called False Agreement problem [5]. To put it loosely, it is not a problem if two systems disagree but it can be a significant problem if they falsely believe they agree.

Thirdly, in order to connect two information systems A and B, we first need to understand the precise relation between the abstractions of entities in reality represented in A and B. Take for example the situation depicted in Figure 2. These two systems commit to different theories (ontologies) of transplants. Can we assume that just because the same term (e.g., Person of Transplant) is used in both structures that they mean the same thing? Of course, not! The only way to precisely characterize the relation between, for example, the type Person in system A and the homonymous type in system B is to find out the relation between their respective referents in their respective underlying conceptualizations. If that relation happens to be one of identity, then we can code the meta-properties of that relation (i.e., reflexivity, symmetry, transitivity and Leibniz’s Law [11]) in the representation of that relation in the corresponding system. If that relation, however, turns out to be a different one (e.g., specialization, historical dependence, existential dependence, parthood [11]), we can also code its representation accordingly.

Semantic interoperability can, thus, be characterized in the following way: two systems A and B semantically interoperate if the coded relations connecting the information structures of A and B: (i) preserve the semantics of the referents represented in those structures; (ii) reflect the real-world meta-properties of the represented relations; and (iii) yield a resulting information structure that constraints the possible states of the resulting system to the intended ones, i.e., to those that represent intended state of affairs according to the conceptualizations underlying A and B.

Figure 2. Problem of semantic interoperability.
As this example illustrates, in order to safely interoperate systems A and B, we need to safely integrate the information structures of A and B. In order to do that, we need a set of conceptual tools that help us to: (i) produce ontologically consistent information structures; (ii) uncover the worldview embedded in existing information structure; (iii) clarify the nature of the notions constituting that worldview; and (iv) calculate the relations between notions constituting different worldviews. These tasks of domain analysis, conceptual clarification and meaning negotiation are the very business of the discipline of Formal Ontology.

4. NO ONTOLOGY WITHOUT ONTOLOGY

As explained in [1], one of the essential FAIR principles is interoperability, which means to guarantee that:

1. I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation;
2. I2. (meta)data use vocabularies that follow FAIR principles;
3. I3. (meta)data include qualified references to other (meta)data.

Item I2 refers to vocabularies. However, given the semantic interoperability characterized in the previous section, these cannot only be vocabularies. Vocabularies are terminological resources and the only way to safely make references to other (meta)data artifacts (i.e., the only way to satisfy item I3) is by precisely clarifying and characterizing the nature of the relations between the referents represented by these artifacts. For this reason, at the bare minimum, we need more than merely terminological resources, but formal, shared and explicit representations of conceptualizations, or, what the area of knowledge representation has conventionally called ontologies. This desiderata is reflected in item I1, which also requires the use of broadly applicable knowledge representation languages for that. The immediate question that comes to mind is then: what are the criteria that a representation language must have in order to satisfy requirements I2 and I3? Those attributes described in I1 are, perhaps, necessary but they are clearly not sufficient! For example, First-Order Logics (FOL) is, historically, the most used language for Knowledge Representation. It is also, obviously, formal, accessible, shared and broadly applicable. However, as explained in depth in [12, 13], it is not a suitable language for systematically addressing requirements I2 and I3.

The reason follows directly from the definition of semantic interoperability I previously defended. In order to address these three interoperability requirements, we need a language that support us in: (a) systematically making ontologically consistent representation choices; (b) making explicit the ontological nature of the elements represented, i.e., the ontological commitment that is being made; and (c) identifying and characterizing the nature of the relations between real-world entities represented in these different data artifacts. In order words, we need a language that is truly ontological in nature, i.e., a language that explicitly commits to a foundational ontology [14].

As previously discussed, providing formal theories for addressing these challenges is the very business of Formal Ontology. This discipline aims at developing formal theories dealing with general aspects of reality
such as identity, dependence, parthood, truthmaking, causality, etc. These domain-independent theories can then be used to investigate and articulate representations of conceptualizations across all domains. A foundational ontology is a particular consistent system of such ontological theories.

FOL is not a true ontological language in this sense. In fact, it is exactly its ontological neutrality that allows it to be attractive as a formalism that can be employed to a large variety of cases, ranging from the foundation of mathematics to representing particular domains. However, also because of its ontological neutrality, users have no support when choosing how to better represent elements constituting a domain conceptualization (a point beautifully demonstrated in [15]); there is also no support for making explicit the ontological commitments that the user thinks he or she is making. To put it boldly, FOL allows one to represent almost everything, including the things one should not represent! As discussed by [16] almost three decades ago:

“Formal semantics of current knowledge representation languages usually account for a set of models which is much larger than the models we are interested in, i.e., real-world models. As a consequence, the possibility to state something which is reasonable for the system but not reasonable in the real world is very high. What we need, instead, is a semantics which is not neutral with respect to some basic ontological assumptions.”

Another alleged candidate for addressing I1 is the Ontology Web Language (OWL). However, as explained in [12, 13], the acronym actually hides a misnomer: there is no Ontology in OWL. OWL is a logical language that can be used to produce logical specifications. As such it inherits all the problems of FOL with the additional non-trivial aspect of having a much lower expressivity than FOL. As a result, as demonstrated in [17, 18], there are many critical interoperability problems that can go undetected when integrating information structures represented in languages such as OWL.

An example of a language that is truly ontological in this sense is OntoUML [11, 14]. OntoUML has been designed to conform to the Unified Foundational Ontology (UFO) such that the modeling primitives of this language reflect the ontological distinctions put forth by UFO. Moreover, the grammar of this language includes formal constraints that reflect the axiomatization of UFO, i.e., the grammatically valid models of OntoUML are those that respect the axiomatization of the ontological theories comprising UFO.

Over the years, OntoUML has been successfully employed in academic, industrial and governmental settings to create conceptual models in a number of different domains, ranging from Geology and Biodiversity Management, to Telecommunications and Bioinformatics, among many others [14]. In fact, research shows that it is among the most used ontology-driven conceptual modeling languages in the literature [19]. Moreover, empirical evidence shows that it significantly contributes to improving the quality of domain representations without requiring an additional effort to produce them [20]. Moreover, as shown by [19], UFO is the second-most used foundational ontology in conceptual modeling and the one with the fastest adoption rate.
By building on ontological semantics of this language, the OntoUML community has developed several methodological and computational tools. Regarding the former, I can cite a pattern grammar for the language comprising a library of ontological design patterns and their relations [21, 22], as well as a library of ontological anti-patterns [10]. Regarding the latter, we can mention computational support for pattern-based ontology construction [22, 23], formal ontology verification, ontology verbalization, proactive anti-pattern detection and rectification [10], and ontology validation by visual simulation. In particular, regarding this last point, as shown in [9], this computational support allows the user to exactly check if the states admitted by a particular information structure (a conceptual model, an ontology) are the ones representing the state of affairs intended by its underlying conceptualization. Finally, as discussed in [14], there are many proposals for model transformation mapping OntoUML models to a variety of languages, including OWL [24]. The idea, also strongly defended here, is that a truly ontological language must be used to address the requirements of domain and comprehensibility appropriateness [11], and semantic interoperability. However, from these models, one can generate several operational (or codification) ontologies addressing non-functional requirements (e.g., efficient computational reasoning, executability) for different classes of applications.

In Figure 3, I present the models of Figure 2 now in OntoUML®. In the model on the left, the language makes a clear distinction between the kinds of objects that exist in this domain, namely, people and organs. Kinds are types that necessarily classify their instances, being responsible for their principle of identity, individuation and persistence [25]. The kind Organ is specialized in the subkinds Heart and Brain. Subkinds are also types that necessarily classify their instances: all hearts (brains) are necessarily hearts (brains). Instances of Person can contingently instantiate the type Living Person and the type Deceased Person. These types are mutually disjoint and they exhaust the extension of the type Person. Instances of Person can also contingently instantiate the types Surgeon, Donor and Donee. However, people move from Living Person to Decease Person due to a change in their intrinsic properties, they move in an out of the extension of Surgeon, Donor and Donee due to a change in relational properties. In this case, by being associated to the relational context of a Transplant, the former contingent types are called Phases and the latter are called Roles. Entities such as Transplants are existentially dependent on a multitude of individuals, and thus, connecting them. These are called Relators [26]. Finally, the model can establish the distinction between a mandatory part (here, every person must have as a part an individual of the type Heart, which can change from situation to situation) and an essential part (here, every person must have as a part a specific instance of Brain, which remains the same from situation to situation) [27]. Now, in the model of the right, we have examples of these different types of ontological categories as well, with the addition of a classification type [28, 29]. A classification type is a higher-order type in OntoUML, i.e., it is a type whose instances are types. In this case, its instances are different types of Transplants.

By using a language that makes explicit the ontological categories to which each of these types and relations belong, we can clearly analyze the connection between the categories in these two models. For example, Person in system A (Person-A) cannot be identical to Person in system B (Person-B) because the

---

® For a full presentation and formal characterization of the OntoUML constructs, one should refer to [9,10,11].
former is a kind of entity and the latter is a phase of Human Beings (for example, suppose that legally speaking Human Beings that loose their cognitive capacities are no longer instances of Person). In fact, Person-A is identical to Human Being in system B and, hence, the relation between Person-A and Person-B is not one of identity but one of generalization. Analogously, the relation between Transplant-A and Transplant-B is not one of identity. The instances of Transplant-A are individual transplants that occur in particular time and space; the instances of Transplant-B are types of transplant. So the relation between them is one of instantiation, i.e., the instances of Transplant-A are instances of instances of Transplant in the sense of system B (Transplant-B)!

In summary, the “I” (interoperability) of FAIR is only possible with the support of information structures that are ontologically consistent and that make explicit the ontological commitments that they inevitably make. We need more than vocabularies. We need good domain ontologies. To construct these ontologies, however, we need engineering support based on Ontology, as a discipline with its numerous and mature theories and methods. Or, as beautifully put by the philosopher Achille Varzi: “No ontology without Ontology”! [30]
REFERENCES

[1] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, … & B. Mons. The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3(2016), Article No. 160018. doi: 10.1038/sdata.2016.18.

[2] C.S. De Souza. The semiotic engineering of human-computer interaction. Cambridge, MA: MIT press, 2005.

[3] G. Guizzardi. Ontological patterns, anti-patterns and pattern languages for next-generation conceptual modeling. In: Conceptual Modeling – the 33rd International Conference (ER 2014), 2014, pp. 13–27. doi: 10.1007/978-3-319-12206-9_2.

[4] Y. Wand & R. Webe. An ontological model of an information system. IEEE Transactions on Software Engineering 16(11)(1990), 1282–1292. doi: 10.1109/32.60316.

[5] N. Guarino. Formal ontology in information systems. In: Proceedings of the First International Conference (FOIS’98), 1998, pp. 337. doi: 10.1071/CP12149.

[6] G.H. Mealy. Another look at data. In: AFIPS ’67 (Fall) Proceedings of the November 14–16, 1967, Fall Joint Computer Conference, pp. 525–534. doi: 10.1145/1465611.1465682.

[7] G. Guizzardi & J. Mylopoulos. Taking it to the next level: Nicola Guarino, formal ontology and conceptual modeling. In: Ontology Makes Sense - Essays in Honor of Nicola Guarino. Amsterdam: IOS Press, 2019, pp. 223–241. doi: 10.3233/978-1-61499-955-3-223.

[8] A. Collier. Critical realism: An introduction to Roy Bhaskar’s philosophy. Brooklyn, NY: Verso, 1994. isbn: 9780860916024.

[9] A.B. Benevides, G. Guizzardi, B.F.B. Braga & J.P.A. Almeida. Validating modal aspects of ontouml conceptual models using automatically generated visual world structures. Journal of Universal Computer Science 16(20) (2010), 2904–2933. doi: 10.3217/jucs-016-20-2904.

[10] T.P. Sales & G. Guizzardi. Ontological anti-patterns: Empirically uncovered error-prone structures in ontology-driven conceptual models. Data & Knowledge Engineering 99(2015), 72–104. doi: 10.1016/j.datak.2015.06.004.

[11] G. Guizzardi. Ontological foundations for structural conceptual models. PhD thesis, University of Twente, Enschede, 2005. Available at: http://doc.utwente.nl/50826/ (accessed 17/08/2019).

[12] G. Guizzardi. On ontology, ontologies, conceptualizations, modeling languages, and (meta)models. In: Databases and Information Systems IV - Selected Papers from the Seventh International Baltic Conference (DB&IS), 2006, pp. 18–39.

[13] G. Guizzardi. Theoretical foundations and engineering tools for building ontologies as reference conceptual models. Semantic Web 1(1–2)(2010), 3–10. doi: 10.3233/SW-2010-0015.

[14] G. Guizzardi, G., Wagner, J.P.A. Almeida & R.S.S. Guizzardi. Towards ontological foundations for conceptual modeling: The Unified Foundational Ontology (UFO) story. Applied Ontology 10(3–4)(2015), 259–271. doi: 10.3233/OA-150157.

[15] N. Guarino. A concise presentation of ITL. In: Processing Declarative Knowledge, International Workshop (PDK’91), 1991, pp. 141–160. doi: 10.1007/BFb0013526.

[16] N. Guarino. The ontological level. In: R. Casati, B.B. Smith & G. White (eds.) Philosophy and the Cognitive Sciences. Vienna: Holder-Pichler-Tempsky, 1994, pp. 443–456.

[17] G. Guizzardi. The role of foundational ontologies for conceptual modeling and domain ontology representation. In: The 7th International Baltic Conference on Databases and Information Systems, 2006, pp. 17–25. doi: 10.1109/DBIS.2006.1678468.

[18] G. Guizzardi, F.A. Baião, M. Lopes & R. de Almeida Falbo. The role of foundational ontologies for domain ontology engineering: An industrial case study in the domain of oil and gas exploration and production.
Ontology, Ontologies and the “I” of FAIR

International Journal of Information System Modeling and Design 1(2)(2010), 1–22. doi: 10.4018/jismd.2010040101.

[19] M. Verdonck. Ontology Driven Conceptual Modeling: Model Comprehension, Ontology Selection, and Method Complexity. PhD thesis, University of Ghent, 2018.

[20] M. Verdonck, F. Gailly, R. Pergl, G. Guizzardi, B. Martins & O. Pastor. Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study. Information Systems 81(2019), 92–103. doi: 10.1016/j.is.2018.11.009.

[21] E. Zambon & G. Guizzardi. Formal definition of a general ontology pattern language using a graph grammar. In: Proceedings of the 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), 2017, pp. 1–10. doi: 10.15439/2017f001.

[22] F.B. Ruy, G. Guizzardi, R. de Almeida Falbo, C.C. Reginato & V.A. Santos. From reference ontologies to ontology patterns and back. Data & Knowledge Engineering 109(2017), 41–69. doi: 10.1016/j.datak.2017.03.004.

[23] F.B. Ruy, C.C. Reginato, V.A. dos Santos, R. de Almeida Falbo & G. Guizzardi. Ontology engineering by combining ontology patterns. In: P. Johannesson, M. Lee, S. Liddle, et al. (eds) Conceptual Modeling. ER 2015. Cham: Springer, 2015, pp. 173–186 doi: 10.1007/978-3-319-25264-3_13.

[24] V. Zamborlini & G. Guizzardi. On the representation of temporally changing information in OWL. In: Proceedings of the 14th IEEE International Enterprise Distributed Object Computing Conference (EDOCW), 2010, pp. 283–292. doi: 10.1109/edocw.2010.50.

[25] G. Guizzardi. Logical, ontological and cognitive aspects of object types and cross-world identity with applications to the theory of conceptual spaces. In: Applications of Conceptual Spaces, 2015, pp. 165–186. doi: 10.1007/978-3-319-319-15021-5_9.

[26] N. Guarino & G. Guizzardi. Relationships and events: Towards a general theory of reification and truthmaking. In: AI*IA 2016: Advances in Artificial Intelligence, 2016, pp. 237–249. doi: 10.1007/978-3-319-49130-1_18.

[27] G. Guizzardi. Modal aspects of object types and part-whole relations and the de re/de dicto distinction. In: International Conference on Advanced Information Systems Engineering, 2007, pp. 5–20. doi: 10.1007/978-3-540-72988-4_2.

[28] G. Guizzardi & G. Wagner. Some applications of a unified foundational ontology in business modeling. In: Business Systems Analysis with Ontologies, 2005, pp. 345–367. doi: 10.4018/9781591403395.ch013.

[29] V.A. Carvalho, J.P.A. Almeida & G. Guizzardi. Using a well-founded multi-level theory to support the analysis and representation of the powertype pattern in conceptual modeling. In: International Conference on Advanced Information Systems Engineering, 2016, pp. 309–324.

[30] A. Varzi. Carnapian engineering. In: S. Borgo, R. Ferrario, C. Masolo et al. (eds.) Ontology Makes Sense. Essays in Honor of Nicola Guarino. Amsterdam: IOS Press, 2019, pp. 3–23.