

The Ontology for Conceptual Characterization of Ontologies

Beatriz Franco Martins^{1(\boxtimes)}, Renata Guizzardi², José Fabián Reyes Román¹, Moshe Hadad³, and Oscar Pastor^{1(\boxtimes)}

¹ Universitat Politècnica de València, VRAIN, Camino de Vera, 46022 Valencia, Spain

{bmartins, jreyes}@vrain.upv.es, opastor@dsic.upv.es

² University of Twente, Twente, The Netherlands

r.guizzardi@utwente.nl

³ Accenture LTD, Cyber R&D Lab, Ha-Menofim St 2, 4672553 Tel Aviv-Yafo, Israel moshe.hadad@accenture.com

Abstract. Ontologies as computational artifacts have been seen as a solution to FAIRness due to their characteristics, applications, and semantic competencies. Conceptualizations of complex and vast domains can be fragmented in different ways and can compose what is known as ontology networks. Thus, the ontologies produced can relate to each other in many different ways, making the ontological artifacts themselves subject to FAIRness. The problem is that in the Ontology Engineering Process, stakeholders take different perspectives of the conceptualizations, and this causes ontologies to have biases that are sometimes more ontological and sometimes more related to the domain. Besides, usually, Ontology Engineers provide well-grounded reference ontologies, but rarely are they implemented. At the same time, Domain Specialists produce operational ontologies storing large amounts of valid data but with naive ontological support or even without any. We address this problem of lack of consensual conceptualization by proposing a reference conceptual model (O4OA) that considers ontological-related and domainrelated perspectives, knowledge, and commitment necessary to facilitate the process of Ontological Analysis, including the analysis of ontologies composing an ontology network. Indeed, O4OA is a (meta)ontology grounded in the Unified Foundational Ontology (UFO) and supported by well-known ontological classification standards, guides, and FAIR principles. We demonstrate how this approach can suitably promote conceptual clarification and terminological harmonization in this area through our framework proposal and its case studies.

Keywords: Interoperability \cdot Reuse \cdot FAIR \cdot Ontological Analysis \cdot Ontology Networks

The authors are grateful to the members of the PROS Center Genome group and Ontology from UPV and Conceptual Modeling Research Group (NEMO) from UFES for fruitful discussions.

[©] The Author(s), under exclusive license to Springer Nature Switzerland AG 2023 J. P. A. Almeida et al. (Eds.): ER 2023, LNCS 14320, pp. 105–124, 2023. https://doi.org/10.1007/978-3-031-47262-6_6

1 Introduction

The FAIR initiative proposes a series of principles to which data management practices should adhere to be considered adequate to meet the challenges of our times. These principles underlie the name of the initiative since FAIR means to be Findable, Accessible, Interoperable, and Reusable [66]. Ontologies as computational artifacts have been seen as a solution to FAIRness due to their characteristics, applications, and semantic competencies. In other words, it is possible to meet the aforementioned principles by using ontology to support data management. Consequently, more and more organizations are looking for ontology-based solutions to achieve these features. The usefulness of ontologies is vast, such as providing conceptual support for architectures (such as data mesh, data lake structuring, and big data solutions), and facilitating human-computer interaction through well-founded conceptual models, among others. All uses of ontologies have in common is the need to interoperate and reuse conceptualizations and data. This common trait requires that the semantics used be clear, and consensual; even more, these characteristics must last throughout the life cycle of the systems that use them. In other words, ontologies that describe complex, vast, and vital knowledge domains such as the cybersecurity domain, and the genetic domain, among others, require a suitable environment for them to comply with the FAIR Principles [66] and be effective in being FAIR [43]. In summary, ontological artifacts themselves are also subject to FAIRness.

Some applications lead to conceptualizations of complex and vast domains that can be fragmented in different ways, thus composing what is known as **Ontology Networks.** The capacity of ontologies to allow modelers to articulate abstractions of a particular state of affairs in reality [25] provides new possibilities for semantic interoperability and data reuse for more extensive and complex domains. However, our research identifies that those domains have characteristics that potentialize semantic misinterpretation that may occur when it is necessary to interoperate conceptualizations. Besides, ontologies covering these kinds of domains deal with data whose sources are strongly embraced by their community. The problem is that in the Ontology Engineering process, stakeholders take different conceptual perspectives, and this causes ontologies to have biases that are sometimes more ontological and sometimes more domain in nature. Indeed, the way domain specialists and ontology engineers seek to achieve FAIRness lacks a more robust semantic bond. Stakeholders usually adopt different perspectives (regarding their cognitive process - ontological commitment [23, 24]) even about the same concept and its surroundings. This is why Ontology Engineers provide well-grounded reference ontologies, which are rarely implemented, while Domain Specialists produce operational ontologies storing large amounts of valid data but with naive ontological support or even without any. Actually, the Ontological Perspective must always comply with the Domain Perspective, throughout the whole conceptualization life cycle. In other words, ontology engineers must capture the domain notions provided by the domain specialists, returning them with conceptualization solutions through well-founded ontological artifacts (e.g., documents, models, and implementations) to support managing their data [51].

We address this problem of lack of consensual conceptualization by proposing a reference conceptual model named **Ontology for Ontological Analysis** (O4OA). O4OA considers, at the same time, the ontological-related and the domain-related perspectives (along with their respective knowledge and commitment necessary to facilitate the process of ontological analysis). This is particularly useful when considering the analysis of the ontologies pertaining to an Ontology Network, which need to maintain consistency among many models to meet FAIR principles. By doing so, these ontology networks may serve the purpose of interoperating data and conceptualizations to their full potential. The presentation of the O4OA is the **main contribution** of this paper. The O4OA Reference Ontology is represented in OntoUML [6], along with its constraints formalized using OCL¹ rules. We implemented the O4OA as a REST-API over a NoSQL database [48] to support semi-automated ontological analysis.

We have organized the remaining of this paper as follows way: Sect. 2 walks through the FAIR Principles, showing the importance of homogenizing ontological artifacts' characterization to achieve FAIRness; Sect. 3 describes O4OA; Sect. 4 presents the verification and validation of O4OA; Sect. 5 discusses our proposal in face of related works; Sect. 6 concludes the paper, and discusses some further research directions.

2 Ontologies and FAIR Principles

The FAIR Principles proposed in [66] clarify data management and stewardship, providing a set of best-practice indicators to allow these processes to be effective. FAIR stands for Findable, Accessible, Interoperable, and Reusable [8,54]. The ontological notion that each individual "thing" has its own identity (Identity Principle) encompasses exactly the "F" principle, in a way that such identity serves the purpose of identification and, thus, allowing an object to be findable. Moreover, the notion of *Rigidity* guarantees that these individuals have a perennial identity, keeping the same *Identity Across Possible Worlds* [46]. The *Identity Principle* goes further from the notion of identification provided by computational artifacts because this kind of identification system only has a programmatic function. On the contrary, foundational ontologies like UFO [28,35], DOLCE [10], among others, can provide computational artifacts with that ontological identity support beyond their processable identification system. UFO provides *Identity Principle* and *Rigidity* through a clear definition of what is a Kind [33]. The "A" principle, best practices address this by exploring the (meta)characteristics ontologies must have in order to guarantee that the data it classifies is truly accessible. Examples of these approaches are [2, 52], while [3] addresses mainly quantitative motivations, not exploring (meta)characteristics. Thus, as denoted in [57], for the security domain, achieving public availability and findability for domain ontologies is still an open issue.

¹ https://www.omg.org/spec/OCL/About-OCL/.

Reference domain ontologies [30] grounded over foundational ontologies provide conceptualizations that encompass the "I" and "R" principles. Moreover, implemented versions of these ontologies are computational artifacts that carry the necessary elements to be processable and semantically precise. However, this is only possible when ontologies fulfill the principles of being well-defined and well-grounded [51]. Guizzardi discusses this perception in depth in [34]. Mainly, well-founded ontologies are able to provide real-world semantics and are more prone to maintain consistency, making explicit the commitments of different conceptualizations.

However, Domain Ontologies (implemented or not) usually fail to accomplish interoperability requirements, mainly due to the different perspectives that Ontology Engineers and Domain Specialists have about standards and norms. These different perspectives are often a source of misinterpretations because they communicate through natural language, which is inherently ambiguous, besides being often governed by political decisions that rarely relate to semantics. Another problem with this divergence of perspective is embedding ontologies with different biases. Domain ontologies developed over the Ontology Engineers' bias usually are well-grounded reference ontologies, but are rarely implemented and the validation with data is limited. Instead, domain ontologies (following Domain Specialists' bias) are usually operational ontologies (i.e., implementations) storing large amounts of valid data but with naive ontological support or even without any [49]. Besides, in both cases, the lack of ontological grounding is a common issue [16, 57]. We address this problem by providing a stable environment for ontological analysis through our proposal of an ontology and an associated framework and computational tool. We present the ontology in the next section. The framework is called **The Framework For Ontologies Clas**sification (F4OC) [50, 51], and it uses O4OA to classify and analyze the ontologies meta-characteristics based on knowledge domain requirements. Finally, a computational tool (semi)automates the use of the framework.

3 A (Meta)Ontology to Describe Ontologies

The Ontology For Ontological Analysis (O4OA) models the foundational and domain-related concepts and relations that are necessary to facilitate the process of *Ontological Analysis* [22]. The goal (purpose) of O4OA is to clarify and homogenize the necessary (meta)ontological requirements, data, and characteristics to help stakeholders achieve awareness and common sense about conceptualizations (ontologies). Because O4OA covers the perspectives of both stakeholders (Ontology Engineer and Domain Specialist), it addresses the "Interoperability" and "Reusability" principles of FAIR, since it serves the purpose of correcting misalignment and miscommunication between the conceptualizations of each of these perspectives.

3.1 Methodology, Stakeholders and Research Questions

We advocate that every ontology must be developed in light of the best practices within the Ontology Engineering Process. Besides, we strongly recommend the adoption of a well-known methodology to drive the process; thus, we adopt the SABiO methodology [1]. According to it, we must define the ontology purpose and identify its users, i.e. their stakeholders. Therefore and taking into account the discussion in Sect. 2, two key responsibilities are part of the pursuit FAIR w.r.t. to the Ontology Engineering Process: While Domain Specialists are concerned with identifying the relevant knowledge aspects that are part of a conceptualization, Ontology Engineers aim at representing this ontology in a way that it expresses this knowledge with real-world semantics to be interpreted unequivocally, either by humans or by computational assets. Then we define the Competence Questions (CQs)² that are the pathway to define the ontology scope and provide its evaluation capabilities; complying with the stakeholder's expectations and requirements [20, 21].

This set of CQs contemplates a cross-perspective of ontological and domainrelated perspectives, extending them to consider ontology networks. In order to formulate these questions, and considering that they are the requirements engineering guidelines, we conducted the O4OA elicitation process in partnership with a multidisciplinary group of stakeholders and attended online meetings³, providing different contributions. Our proposal is domain-agnostic, but our case studies are about the Cybersecurity domain; therefore, we are receiving advice from Cybersecurity specialists of our research group and others who are members of the project consortium we participate in⁴. The group comprises Cybersecurity Domain Specialists, Ontology Engineers, and Literature Review Specialists, among others. Also, it is essential to clarify that Ontology Engineers took on different roles, sometimes eliciting requirements and sometimes as project clients, depending on the context discussed and their roles in the group.

From the defined scope, purpose, commitment, and competence questions, and knowing the involved stakeholders, we proceeded with the ontology O4OA engineering process according to SABiO. We represent our proposal using the OntoUML language, which provides grounding over UFO.

3.2 Conceptual Characterization of Ontologies

In order to develop the Ontology Engineer perspective, we searched within the state of the art in ontology engineering to find the meta-features used to characterize an ontology. We found vast works within the context of ontology classification, such as the works [15,19,38,45,47,53,61,62,64]. These works use several

² Readers may find the complete description of O4OA competence questions at the following repository: https://bfmartins.gitlab.io/o4oa/.

³ Part of the elicitation process happened during the COVID-2019 pandemic, so the remote strategy was mandatory.

⁴ Our research is part of a research consortium to develop well-grounded knowledge graphs through a comprehensive solution within a project in collaboration with teams from several academic institutions, and Accenture LTD.

levels of abstraction to classify ontologies, for instance, according to the degree of formalization and/or axiomatization of ontologies, their applicability, generality, structure, and development among others. We adopt the most relevant and comprehensive classification criteria as the referential base for O4OA, they are [17, 18, 22, 30, 65]. However, we advocate that these dimensions must encompass a systematic ontology classification approach to guarantee the FAIRness of ontological artifacts. Therefore, O4OA uses a holistic approach that considers a set of dimensions to characterize ontologies.

The first dimension considers a classification based on the *level of applicability* proposed in [30]. This classification allows us to differentiate when an ontology is an "explicit and formal representation of a portion of reality for knowledge sharing" or an "implementation of this representation for knowledge computational management", i.e., and if it is a Reference or an Operational Ontology.

The second dimension deals with the *level of generality* (sometimes called knowledge kind) of ontologies that refer to a level of dependence on a specific point of view. Many proposals target this dimension, such as [15,38,62]; however, the proposal most accepted by the community is Guarino's [22], which complements the proposal of Mizoguchi and Ikeda [53]. This classification characterizes conceptualizations as *Foundational Ontologies* (which are independent of a particular problem or domain and express very general concepts and their relations like *things* and their properties, *events, relations*, etc. They are also known as High-level Ontologies or Upper Ontologies). Already, Non-Foundational Ontologies are *Domain Ontologies* (which provide conceptualizations about domain tasks, processes, and activities), and *Application Ontologies* (which encompasses both contexts of Domain and Task Ontologies). Another widely accepted classification describes the *Core Ontologies* [65]⁵.

Figure 1 shows the classification approach adopted in O4OA, in which we describe the classification levels using OntoUML <<*subkind>>*, considering the aforementioned classification describes types of ontologies.



Fig. 1. Fragment of the O4OA as a (meta)ontology – Classifications according to [22, 30, 65].

⁵ Which is more general than Domain, Task, and Application ontologies, but more specific than Foundational Ontologies.

We also consider the classification provided by Gomés-Peréz and Corcho [18] as additional dimensions for ontologies classification because it analyzes the ontologies based on their axiomatization level (and considers the limitations of the language) in order to identify its computational limitations when a conceptualization becomes an implemented ontology (i.e., an operational ontology). They divide ontologies by considering the expressiveness of the language used into two aspects: Lightweight and Heavyweight ontologies. A bi-dimensional classification [17], based on [63] and [18], provides a link between the axiomatization and formal levels, focusing on the approach and expressiveness of the language. In Subsect. 3.4 we detail how languages and ontologies are related, as well as the relational aspects that rely on this classification.

We opt for these works because are the most accepted classifications used by the Ontology Engineering community and cover the set of CQs related to the ontology engineering perspective (see previous section). The preference for these classification dimensions instead of others is based on the fact that they already combine the necessary meta-features for FAIRness and because other dimensions are not frequently used. For instance, some works provide a classification based on the nature of the real-world issue [45], the development method [61], and other bi-dimensional classifications [19,47]. However, due to their limited use and to avoid increasing the complexity of the proposed conceptualization, we do not use these additional classifications.

3.3 Domain Cloud of Concepts in Conceptual Characterizations

The O4OA responds to the domain-related CQs questions in terms of ontological artifacts. Thus, we center on the concepts belonging to a conceptualization that must be represented and described. However, to understand this, we need first to clarify the philosophical grounding of what encompasses a conceptualization, and precisely distinguish what a concept is (as an abstraction) and what is the concept representation (as an artifact).

In the philosophical context, a *Concept* is basically a building block of thoughts⁶ and can be seen as a mental representation. UFO deals with this philosophical notion of what is a concept as *Tropes* [27,28]. However, we need to define artifacts belonging to ontologies used to represent concepts. Figure 2 presents the relation between concepts (Concept) and documentation sources (Source). In this situation, Document is defined as a << category >> because it aggregates properties of individuals with different identity principles. Indeed, policies, standards, and any literature documentation exist with no dependence, having their own identity. Sources, as well as any other sort of element that provides relevant information for ontologies, are fluid (<< rolemixin>>). We also use Term as a syntactical artifact (an *Object Kind*) used to describe the notion of a *Concept* (as *Trope Kind*), thus we call Concept the role that a term assumes when is defined in an ontology. As a matter of fact, the notions of *Source* and *Term* as roles are relational-dependent [36]. Thus, we represent a Concept Definition

⁶ https://plato.stanford.edu/entries/concepts/.

present in an ontology (Domain Description) as a building block used to clarify grammatically (terminologically) the notion of a *Concept* (as *Trope Kind*) according to some source of information. Note that we use the *OntoUML notion* of *Part/Whole* through the relation componentOf to represent the definitions that compose each domain description. We present more details about domain descriptions in Subsect. 3.4.



Fig. 2. Fragment of the O4OA as a (meta)ontology – Domain definitions.

3.4 Conceptual Characterization of Ontology Networks

From the conceptual characterization of ontologies as artifacts presented in Subsect. 3.2, we can extend it to identify the relationships among these ontologies (presented in Sect. 3.1). These are key elements regarding FAIR Principles, especially regarding the "I" and "R". Indeed, ontologies can relate in different ways in networks [1,14], as demonstrated in initiatives like [9,39], for example. Aside from that, ontology networks are not necessarily a set of isolated ontologies grouped together, merely because they act in a domain subdivided into smaller parts (subdomains). Instead, how ontologies relate to each other directly depends on how their building blocks relate; in this case, we are talking about relational (meta)characteristics that promote FAIRness.

Ontologies (meta)characteristics (i.e., their purpose, scope, generality, etc.), together with the definitions that compose a domain description applicable to these ontologies (and consequently its foundation), define the relationships present in this network [50]. The Applicability Level of ontologies goes beyond only classifying whether an ontology is implemented. This information follows the notion that a Reference Ontology should be a conceptualization that is constructed to make the best possible description of the domain concerning a certain level of generality and point of view and that an Operational Ontology is the actionable version of a Reference Ontology that uses the most appropriate language in order to guarantee desirable computational properties without compromising the previously defined ontological commitments [23,30]. Therefore, there should not be any operational ontology without the existence of a previous reference ontology in which concepts and their relationships are *well-defined*. Figure **3** presents Reference and Operational Ontologies, their roles, and their relation.



Fig. 3. Fragment of the O4OA as a (meta)ontology – Applicability Level.

Regarding the Application Level, a Conceptualization is a reference ontology that is represented through a Modeling Language; and an Implementation is an operational ontology that works through an Implementation Language. We use the *Relator Pattern* [32] to represent the relational aspects that appear in the characterization of ontologies. According to its Applicability Level, they are Domain Description and the Ontology Schema. Thus, the notion that an ontology is or has an implemented version (implementationFor, a <<material>> relation) derives from the fact that reference ontologies provide ontological support for ontological schemes. This relation allows us to evaluate the relational characteristics a Reference Ontology can provide to its implementations. Besides, this approach can also help ontology engineers deal with implementation language limitations by knowing which ontological aspects can (or can not) be implemented without losing ontological decidability.

Incidentally, a domain description is thought in some representation language (the Modeling Language role), usually an Ontology-Driven Modeling Language (ODML). Besides, as metamodels specify languages, an Ontology-Driven Metamodel specifies an ODML. From this perspective, an ontology drives an ODML, constraining philosophically its metamodel, denoting the specifys <<material>> relation, and defining the <<Relator>> Ontology-driven Language Specification as depicted in Fig. 4.



Fig. 4. Fragment of the O4OA as a (meta)ontology – Ontology-Driven Modeling Languages.

Any ontology can drive ODMLs. For instance, OASIS is an Ontology-Driven Domain Specific Language grounded over O^3 [59,60]; likewise, OntoUML is a Foundational Ontology-Driven Language grounded over UFO. In this case, an ontology is considered *Well-grounded*, if it is represented through a Foundational Ontology-Driven Language (as OntoUML, for instance), i.e. this language, is the bearer of the Ontological Foundation for the conceptualization. Thus, an indirect grounding is provided through the Ontology-driven Language Specification as depicted in Fig. 5.



Fig. 5. Fragment of the O4OA as a (meta)ontology – Ontologies driving languages.

The notion of *Well-grounded* ontologies is based on the fact that the support of a Foundational Ontology helps to avoid semantic interoperability problems in more specific ontologies [26]. In other words, Foundational Ontologies are fundamental for Ontology-Driven Conceptual Languages used to produce Domain, Task, and Application, and Core Ontologies, as well as providing ontological analysis for not grounded conceptualizations. Therefore, we advocate that ontologies must be evaluated according to their grounding, separating ontologies that are driven by foundational ontologies (i.e., well-grounded) from ontologies without this support (i.e., not grounded) [51].

The classification according to the ontology generality level provides us the ability to study the impact that the lack of ontological foundation can produce when it is necessary to interoperate concepts of this type of ontologies and at the same time guarantee FAIRness when we put attention on this relationship among a Foundational Ontology, and the ontologies grounded by it. Therefore, we use O4OA to describe the grounding ontologies relationship. With this respect, we define the groundedOver <<material>> relation established through the Foundational Ontologies role, i.e. concepts defined in a non-foundational ontology specialize from more general conceptual (philosophical) notions from a Foundational Ontology, defining well-grounded ontologies and allowing stakeholders to make solid semantic considerations.

Figure 6 shows the (groundedOver) relation we define, as well as it also describes how the classification differentiates the ontologies through some of its characteristics. We use the *Relator Pattern* to describe how Foundational Ontologies ground the non-foundational ontologies.



Fig. 6. Fragment of the O4OA as a (meta)ontology – Well-grounded ontologies.

Still chasing the FAIRness, O4OA characterizing ontologies according to their generality level provides another important feature. An ontology can reuse other ontologies; in this case, we are dealing directly with the "R" principle of FAIR. Different types of reuse can appear in this relation, depending on how the Reuser Ontology lays hold of and uses concepts of the Reused Ontology. The most usual reuse happens when concepts defined in an ontology can specialize into concepts defined in another ontology. Apart from this, concepts defined in non-foundational ontologies can specialize foundational notions (thought a groundingOver relation). Additionally, the reuse of ontologies can occur through the addition of stronger ontological grounding, however, maintaining the alignment of the domain definitions already adopted. This situation happens when the domain perspective about the definitions present in the related conceptualization is aligned, but the ontological perspective must be reinforced. In other words, this happens when the reused ontology lacks an ontological foundation and requires the grounding provided by a Foundational Ontology or the use of an ODML (provided by the reuser ontology), [13] is an example of this reuse.

Under the umbrella of ontology networks, stakeholders usually confuse the reuse of ontologies with the notion that ontologies can be composed of other ontologies. This is because the notion of a Whole/Part can be seen as a larger ontology using smaller ontologies, but this is not the same. This issue can be aggravated when these relationships occur simultaneously. For instance, UFO is composed of UFO-A, UFO-B, and UFO-C sub-ontologies, and at the same time, UFO-B and UFO-C reuse UFO-A. Indeed, the reuse of ontologies denotes an *Intersection* among ontologies while Whole/Part follows the *Weak Supplementation Pattern*, which states that every whole must be composed by at least two parts [29,31]. Figure 7 depicts the reuse of ontologies and how ontologies can be composed by other ontologies (sub-ontologies).



Fig. 7. Fragment of the O4OA as a (meta)ontology – Reuse a Whole/Part.

3.5 A Semi-automatic Support for Ontological Analysis

Given the O4OA reference model, we implemented an operational version and a frontend solution to provide easy, responsive, and multiplatform access. This operational version uses *Microservices Architecture* and composes data storage with *MongoDB*, a *REST-AP* made with *NodeJS* and *Express*, and the responsive frontend prototype implemented with *Angular Material*, all in *Docker* containers. We manually added the data collected about the ontologies belonging to our case studies [49]. Regarding the domain perspective, we loaded, verified, and validated the terminology of the domain in a study by using well-established standards. In this case, as our target domain is cybersecurity, our referential sources are the ISO/IEC 27032:2012 [40] and the ISO/IEC 27000:2018 [41]. Additionally, we use this referential to compare the cybersecurity terminology definitions from other sources, such STIX, MAEC, NIST, ITU, among others.

Up to now, we have assessed 161 concepts in the Cybersecurity domain, and many others obtained from the associated foundational and domain-correlated ontologies studied. Associated with these concepts (in the cloud of concepts), we registered 73 reliable sources, providing a burst of possible usage definitions in ontologies of this and its related domains. For instance, taking the concept of Risk, we found 18 definitions of what it is a Risk. Besides, we also found many other risk-related definitions, the ones for concepts such as *Level of Risk*, Residual Risk, Risk Criteria, and other 11 associated ones. In fact, this is a (regular expressions) recursive process because the O4OA operational ontology version is a graph. Although it is a syntactic process, these kinds of findings open the opportunity for the next step of research, which is reasoning the semantics of concepts, context, and ontological commitments⁷. This is possible because the concepts of Foundational Ontologies and their definitions are also registered in our data storage (as meta characteristics) through using ODML and grounding by specialization. We present findings about the analysis of the Risk and the analysis of the *Vulnerability* in the works [51] and [50] respectively, as part of our case studies.

The frontend solution of the tool is still under development, so we consider it as an *alpha* version. However, it has already demonstrated its potential in facilitating access to data and supporting the ontological analysis process with O4OA. For instance, Fig. 8 shows that we can trace the concept of *Vulnerability* from its definitions until the ontologies use them. We are working on a better graph presentation (dynamic) to allow dynamic navigation in the cloud of concepts as well as in ontology (meta)characteristics⁸.

⁷ The adoption of microservices allows API scaling adding reasoning capabilities, for this future possibilities.

⁸ It is important to point out that we adopted an *Agile Development* approach in order to provide fast initial results meanwhile being scaled.



Fig. 8. Export image of the O4OA tool - fragment of the Vulnerability definitions.

O4OA deals with the (meta)characteristics of both the ontological and domain views, clarifying their relationship. Therefore, this allows tracing of how concepts are represented or implemented in ontologies that go beyond the presentation of the sub-ontologies walk graph, exemplified in Fig. 8. This allows navigating a graph starting from any concept within the cloud of concepts to the ontologies that use them, including access to the definitions adopted in each case. In fact, this is a feature already available in the API we developed and that will possibly gain relevance with the use of a graph-enhanced presentation in the frontend. Likewise, we can navigate through the ontological relations to find out the ontological grounding supporting a concept, even when it appears in different ontologies, and even compare it in one well-grounded ontology with another imprecise one, for instance.

4 Evaluation

According to SABiO, during the evaluation phase of the development process, the proposed CQs must be confronted with the ontology developed to guarantee that it complies with the requirements defined. Additionally, it is required that the reference model be analyzed through processes of model instantiation in order to explore possible issues or unexpected possibilities scenarios (branches or worlds). We adopted the Alloy analyzer tool [42] applying the OntoUML notions present in the work [7] to proceed with the validation; besides, this analysis is being performed concurrently with the development of the operational version of the ontology. Due to the O4OA model characteristics (size and complexity) and design decisions, we fragment the analysis, running the instantiation of each model package in an individual and modular way. In the validations process, we elicit the set of additional constraints (in addition to those already present in OntoUML) required, and we also check model cardinalities to ensure correct semantics. For example, when analyzing the instantiation of the contents of the **Reuse** package, because the reused and reuser ontology roles are not *disjoint*, we had to add a constraint to avoid cyclic reuses, i.e., a Transitive Closure predicate for the relations **reuses**. Note that some required constraints must be implemented directly in the persistence, while others in the API. See the complete model evaluation details in the repository of the ontology.

We also developed a framework for classifying and characterizing ontologies, which is composed of the presented reference ontology, its version implemented in NoSQL, and a prototype API that manipulates and manages the (meta)data obtained in the application of the framework. We proposed a sequence of five ontological-related and domain-related steps to identify and catalog the (meta)data regarding the ontology and the domain perspectives, respectively. As the framework is based on O4OA and was formulated to ensure compliance with FAIR principles, we obtained promising results in our case studies; for example, those we have presented in [49–51].

It is important to point out that although our case studies are within the Cybersecurity domain, O4OA and its associated framework are agnostic, allowing their use in any domain of knowledge. Indeed, in domains covering vast knowledge, which are extremely regulated (normalized) or complex, our proposal demonstrates its advantages more than those in lighter domains. This is because the complexity and expense of ontological analysis grow proportionally as the domain gets more vast and complex. We observe these phenomena during the course of the study done within the Cybersecurity domain, which has several of these characteristics; it is vast, highly regulated by norms and standards, constantly evolving, and difficult in its own right. One evidence of this is present in the work of [56], an example of which the O4OA-based framework can homogenize and contribute to the process of ontological analysis.

5 Related Works

Several initiatives deal with Ontology-Driven Interoperability (ODI), especially in areas of Internet of Things (IoT) and Web of Things (WoT), such as [5, 55]. Their related ontologies SSN [4], oneM2M [58], and SAREF [11] are W3C standards. As in Cybersecurity, IoT and WoT are complex domains where stakeholders must commit agreement. However, these initiatives differ from ours. The first distinction is in the domain itself; while they deal with the core characteristics in the IoT/WoT domain⁹, O4OA deals with (meta)characteristics present in any kind of ontological artifact created to represent any domain. Besides, O4OA rationalizes the notion of FAIRness over ontological analysis processes, while such ontologies rationalize ODI into their domain. Second, although they are well-example initiatives in the reuse of ontologies in themselves, they do not deal with the notion of a broad cloud of concepts (and their details) nor relations among ontologies in any networks. Indeed, they are data interoperability providers for IoT/WoT while O4OA is an interoperability provider for any ontologies¹⁰. Lastly, IoT/WoT ontologies have the same issues we detected in the cybersecurity ontologies, detailed in [49]; notably, lack of a grounding, making them require adaptations to interoperate or have proper reuse, with no

 $^{^{9}}$ SSN, oneM2M, and SAREF are *Core Ontologies* in the sense of [22,65].

¹⁰ In O4OA, the relations and concepts of ontologies are data instances.

assuring semantic (grounding). The work [5] runs ODI by making ontological analysis and goes in line with the notion of FAIRness (like O4OA) under the ODI viewpoint (ontological perspective), but there is no mention of important domain-dependent aspects, i.e., domain (meta)characteristics (domain perspective). Instead, O4OA is domain-agnostic but not domain-indifferent since the purpose of performing an ontological analysis is to elicit knowledge in a consensual, reproducible, traceable, and formal way. Indeed, ODI is among many uses where ontological analysis is a key contributor.

The Ontology Metadata Vocabulary (OMV) [37] is a proposal for describing ontologies and related entities, and this is the only approach similar to ours that we could find in the state of the art. It focuses on metadata of ontologies intending to be the standard covering this domain. The proposal has demonstrated usefulness in initiatives such as [12]. The approach distinguishes between an ontology base (a conceptualization) and an ontology document (a realization of a conceptualization - an implementation). The ontology covers some of the metadata that is part of the FAIRness discussion, such as language, licensing, and quantitative data (number of classes, properties, and axioms). OVM also uses Guarino's classification [22] to classify ontologies. In this respect, OVM is similar but lighter than our proposal; however, as an ontology, OVM in itself is not FAIR. Besides, it does not have the support of a prior reference ontology; indeed, it is an ontology implemented in OWL without using any foundation ontology for grounding. Conversely, O4OA is grounded on UFO, has a well-defined reference model written in OntoUML, and is implemented in a NoSQL database; besides, it supports our framework proposal following a solid methodological approach (namely, SABiO).

The work [44] presents a study of the metadata of a vast number of ontologies. In this work, some works were more relevant in terms of being available to describe ontological metadata, such as Dublin Core, Ontology Metadata Vocabulary, VoID, etc. The study compared of these works and their implementations, demonstrating the lack of foundational grounding as an issue. This confirms the claim that "surprisingly, both in research and industry, ontologies as computational artifacts are very often built without the aid of any framework of this kind (citing our proposal), favoring recurrent modeling mistakes and gaps" [56].

6 Conclusions

The main contribution of this paper is proposing the Ontology for Ontological Analysis, a (meta)ontology that classifies and characterizes ontologies from their (meta)characteristics. We use the SABiO approach and OntoUML language (grounded in UFO) to develop O4OA. Additionally, our proposal is based on a series of well-established ontology classifications, as well as the best practices supported by FAIR Principles. This ontology is implemented and guides a framework for classifying and characterizing ontologies, providing a clear and reproducible environment that helps the Ontological Analysis Process. Our proposal stands out because in itself it is ontology-driven, well-defined, and wellgrounded, i.e., it is FAIR. Firstly, O4OA provides a conceptual analysis of the nature of the different (meta)characteristics present in the distinct stakeholders' perspectives by using of UFO's foundational categories present in OntoUML. We systematize the process of ontology classification using the most recognized works and with the best coverage of classification dimensions. Furthermore, this ontology is a reference model to study, manage, and maintain ontologies that describe real-world complex, extremely regulated, and data-sensible domains.

Secondly, as a (meta)ontological reference model, O4OA can provide operational versions to track, analyze, and provide reasoning about ontologies belonging to ontology networks. This kind of approach is a fertile field for managing conceptualizations that support industrial architectures such as data mesh, data lake structuring, big data solutions, facilitating human-computer interaction among enterprise stakeholders and teams, and many others.

Thirdly, O4OA establishes a common, stable, and systematic environment for improving communication among ontology engineers and domain experts, avoiding misinterpretations, misunderstandings, structural issues in ontologies, and communication problems that interfere with FAIRness.

Fourth, in addition to these uses, the O4OA prototype tool we are developing to support ontological analysis has been presenting interesting results despite its ongoing development. Its already-built features demonstrate its potential in allowing management and clarification of cloud-of-concepts in ontologies in a semi-automated way. Besides our implementation proposal, other initiatives can emerge; for instance, providing Analytics in a logical language such as Common Logic, OWL, or other reasoners. Moreover, these operational versions can evolve to provide reasoning, tools, and other features or automation. In particular, we intend to develop a web solution encompassing these resources.

Finally, we intend to define the full set of axioms (we did not include any axiom in this paper because of space limitations) to complete the formalization of O4OA. We plan to improve the evaluation of the ontology in other realworld industrial scenarios, including one focusing on the notion of **Ontological Technical Depth**, to prove that ontologies are promising for practical use. We also intend to strengthen the connection between the work developed in the Cybersecurity and Software Engineering domains to bring teams working in both areas closer to improve security in software systems. The objective is to promote practical results in industrial development environments. We also pretend to extend the number of (meta)characteristics covered in O4OA and provide Analytics with them.

Acknowledgments. This work has financial support of Accenture LTD with project Digital Knowledge Graph – Adaptable Analytics API, Generalitat Valenciana with the project CoMoDiD (CIPROM/2021/023), Spanish State Research Agency with projects DELFOS (PDC2021-121243-I00), SREC (PID2021-123824OB-I00), MICIN/AEI/10.13039/501 100011033, and co-financed by the European Union Next Generation EU/PRTR with ERDF.

References

- 1. de Almeida Falbo, R.: Sabio: systematic approach for building ontologies. In: Onto.Com/odise@ Fois (2014)
- Amdouni, E., Bouazzouni, S., Jonquet, C.: O'faire: ontology fairness evaluator in the agroportal semantic resource repository. In: The Semantic Web: ESWC 2022 Satellite Events: Hersonissos, Crete, Greece, 29 May-2 June 2022, Proceedings, pp. 89–94. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-11609-4_17
- Amdouni, E., Jonquet, C.: FAIR or FAIRer? an integrated quantitative FAIRness assessment grid for semantic resources and ontologies. In: Garoufallou, E., Ovalle-Perandones, M.-A., Vlachidis, A. (eds.) MTSR 2021. CCIS, vol. 1537, pp. 67–80. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-98876-0_6
- Atkinson, R., García-Castro, R., Lieberman, J., Stadler, C.: Semantic sensor network ontology. Technical Report. OGC 16–079, World Wide Web Consortium (2017)
- Bauer, M., et al.: Towards semantic interoperability standards based on ontologies. In: AIOTI White paper (2019)
- Benevides, A.B., Guizzardi, G.: A model-based tool for conceptual modeling and domain ontology engineering in OntoUML. In: Filipe, J., Cordeiro, J. (eds.) ICEIS 2009. LNBIP, vol. 24, pp. 528–538. Springer, Heidelberg (2009). https://doi.org/ 10.1007/978-3-642-01347-8_44
- Benevides, A.B., Guizzardi, G., Braga, B.F.B., Almeida, J.P.A., et al.: Validating modal aspects of ontouml conceptual models using automatically generated visual world structures. J. Univ. Comput. Sci. 16(20), 2904–2933 (2010)
- Boeckhout, M., Zielhuis, G.A., Bredenoord, A.L.: The fair guiding principles for data stewardship: fair enough? Eur. J. Hum. Genet. 26(7), 931–936 (2018)
- Borges Ruy, F., de Almeida Falbo, R., Perini Barcellos, M., Dornelas Costa, S., Guizzardi, G.: SEON: a software engineering ontology network. In: Blomqvist, E., Ciancarini, P., Poggi, F., Vitali, F. (eds.) EKAW 2016. LNCS (LNAI), vol. 10024, pp. 527–542. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49004-5_34
- Borgo, S., Masolo, C.: Ontological Foundations of DOLCE, pp. 279–295. Springer, Dordrecht (2010). https://doi.org/10.1007/978-90-481-8847-5_13
- Daniele, L., Garcia-Castro, R., Lefrançois, M., Poveda-Villalon, M.: Smart applications reference ontology (saref) (2019). Accessed Aug 2023
- 12. d'Aquin, M., et al.: Watson: a gateway for the semantic web (2007)
- Duarte, B.B., Falbo, R.A., Guizzardi, G., Guizzardi, R.S.S., Souza, V.E.S.: Towards an ontology of software defects, errors and failures. In: Trujillo, J.C., et al. (eds.) ER 2018. LNCS, vol. 11157, pp. 349–362. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00847-5_25
- 14. Euzenat, J.: Revision in networks of ontologies. Artif. Intell. 228, 195-216 (2015)
- Fensel, D.: Ontologies, pp. 11–18. Springer, Heidelberg (2001). https://doi.org/10. 1007/978-3-662-04396-7_2
- García S, A., Guizzardi, G., Pastor, O., Storey, V.C., Bernasconi, A.: An ontological characterization of a conceptual model of the human genome. In: Intelligent Information Systems: CAiSE Forum 2022, Leuven, Belgium, 6–10 June 2022, Proceedings, pp. 27–35. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-07481-3_4

- 17. Giunchiglia, F., Zaihrayeu, I.: Lightweight ontologies. University of Trento, Technical report (2007)
- Gómez-Pérez, A., Corcho, O.: Ontology languages for the semantic web. IEEE Intell. Syst. 17(1), 54–60 (2002)
- Gomez-Perez, A., Fernández-López, M., Corcho, O.: Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web. Springer, Heidelberg (2004). https://doi.org/10.1007/b97353
- 20. Gruninger, M.: Methodology for the design and evaluation of ontologies. In: International Joint Conference on Artificial Intelligence (1995)
- 21. Gruninger, M.: Designing and evaluating generic ontologies. In: 12th European Conference of Artificial Intelligence, vol. 1, pp. 53–64. Citeseer (1996)
- Guarino, N.: Formal ontology in information systems. In: Proceedings of the 1st International Conference, pp. 6–8. IOS Press, Trento (1998)
- Guarino, N.: The ontological level. In: Philosophy and the Cognitive Sciences (1994)
- Guarino, N.: The ontological level: Revisiting 30 years of knowledge representation. In: Conceptual Modeling: Foundations and Applications, pp. 52–67 (2009)
- Guarino, N., Poli, R.: The role of formal ontology in the information technology. Int. J. Hum Comput Stud. 43(5–6), 623–965 (1995)
- 26. Guizzardi, G.: The role of foundational ontology for conceptual modeling and domain ontology representation, keynote paper. In: 7th International Baltic Conference on Databases and Information Systems (DB&IS). IEEE Press, Vilnius (2006)
- Guizzardi, G., Masolo, C., Borgo, S.: In the defense of a trope-based ontology for conceptual modeling: an example with the foundations of attributes, weak entities and datatypes. In: 25th International Conference on Conceptual Modeling, Berlin (2006)
- 28. Guizzardi, G.: Ontological Foundations for Structural Conceptual Models. CTIT, Centre for Telematics and Information Technology (2005)
- Guizzardi, G.: Modal aspects of object types and part-whole relations and the de re/de dicto distinction. In: Krogstie, J., Opdahl, A., Sindre, G. (eds.) CAiSE 2007. LNCS, vol. 4495, pp. 5–20. Springer, Heidelberg (2007). https://doi.org/10.1007/ 978-3-540-72988-4_2
- Guizzardi, G.: On ontology, ontologies, conceptualizations, modeling languages, and (meta) models. Front. Artif. Intell. Appl. 155, 18 (2007)
- Guizzardi, G.: The problem of transitivity of part-whole relations in conceptual modeling revisited. In: van Eck, P., Gordijn, J., Wieringa, R. (eds.) CAiSE 2009. LNCS, vol. 5565, pp. 94–109. Springer, Heidelberg (2009). https://doi.org/10.1007/ 978-3-642-02144-2_12
- 32. Guizzardi, G.: Ontological patterns, anti-patterns and pattern languages for nextgeneration conceptual modeling. In: Yu, E., Dobbie, G., Jarke, M., Purao, S. (eds.) ER 2014. LNCS, vol. 8824, pp. 13–27. Springer, Cham (2014). https://doi.org/10. 1007/978-3-319-12206-9_2
- 33. Guizzardi, G.: Logical, ontological and cognitive aspects of object types and crossworld identity with applications to the theory of conceptual spaces. In: Zenker, F., Gärdenfors, P. (eds.) Applications of Conceptual Spaces. SL, vol. 359, pp. 165–186. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15021-5_9
- Guizzardi, G.: Ontology, ontologies and the "I" of FAIR. Data Intell. 2, 181–191 (2020)
- 35. Guizzardi, G., Ferreira Pires, L., van Sinderen, M.: An ontology-based approach for evaluating the *domain appropriateness* and *comprehensibility appropriateness* of modeling languages. In: Briand, L., Williams, C. (eds.) MODELS 2005.

LNCS, vol. 3713, pp. 691–705. Springer, Heidelberg (2005). https://doi.org/10. 1007/11557432_51

- Guizzardi, G., Wagner, G.: What's in a relationship: an ontological analysis. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) ER 2008. LNCS, vol. 5231, pp. 83–97. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87877-3_8
- Hartmann, J., Sure, Y., Haase, P., Palma, R., Suarez-Figueroa, M.: OMV-ontology metadata vocabulary. In: ISWC, vol. 3729 (2005)
- Hele-Mai, H., Tanel-Lauri, L.: A survey of concept-based information retrieval tools on the web. In: Proceedings of the 5thEast-European Conference AD BIS, pp. 29–41 (2001)
- Hemberg, E., et al.: Linking threat tactics, techniques, and patterns with defensive weaknesses, vulnerabilities and affected platform configurations for cyber hunting (2021)
- ISO Central Secretary: Information technology security techniques guidelines for cybersecurity. Standard ISO/IEC 27032:2012. International Organization for Standardization, Geneva (2012)
- 41. ISO Central Secretary: Information technology security techniques information security management systems - overview and vocabulary. Standard ISO/IEC 27000:2018–02, International Organization for Standardization, Geneva (2018)
- 42. Jackson, D.: Software Abstractions: Logic, Language, and Analysis. MIT press, Cambridge (2012)
- Jacobsen, A., et al.: FAIR principles: interpretations and implementation considerations. Data Intell. 2(1-2), 10-29 (2020). https://doi.org/10.1162/dint_r_00024
- Jonquet, C., Toulet, A., Dutta, B., Emonet, V.: Harnessing the power of unified metadata in an ontology repository: the case of agroportal. J. Data Semant. 7(4), 191–221 (2018)
- Jurisica, I., Mylopoulos, J., Yu, E.: Using ontologies for knowledge management: an information systems perspective. In: Proceedings of the Annual Meeting-American Society For Information Science, vol. 36, pp. 482–496. Information Today; 1998 (1999)
- 46. Kripke, S.A.: Naming and Necessity. Harvard University Press, Cambridge (1980)
- Lassila, O., McGuinness, D.: The role of frame-based representation on the semantic web. Linköping Electron. Articles Comput. Inf. Sci. 6(5), 2001 (2001)
- Martins, B.F., Serrano, L., Reyes, J.F., Panach, J.I., Pastor, O.: Towards the consolidation of cybersecurity standardized definitions: a tool for ontological analysis. In: Proceedings of the XXIV Iberoamerican Conference on Software Engineering, CIbSE 2021, San José, Costa Rica, 2021, pp. 290–303. Curran Associates (2021)
- Martins, B.F., Serrano, L., Reyes, J.F., Panach, J.I., Pastor, O., Rochwerger, B.: Conceptual characterization of cybersecurity ontologies. In: Grabis, J., Bork, D. (eds.) PoEM 2020. LNBIP, vol. 400, pp. 323–338. Springer, Cham (2020). https:// doi.org/10.1007/978-3-030-63479-7_22
- Martins, B.F., Reyes Román, J.F., Pastor, O., Hadad, M.: Improving conceptual domain characterization in ontology networks. In: Research Challenges in Information Science: 17th International Conference, RCIS 2023, Corfu, Greece, 23–26 May 2023, Proceedings, pp. 187–202. Springer, Heidelberg (2023). https://doi.org/10. 1007/978-3-031-33080-3_12
- Martins, B.F., et al.: A framework for conceptual characterization of ontologies and its application in the cybersecurity domain. Softw. Syst. Model. 21(4), 1437–1464 (2022)

- Mazimwe, A., Hammouda, I., Gidudu, A.: Implementation of fair principles for ontologies in the disaster domain: a systematic literature review. ISPRS Int. J. Geo Inf. 10(5), 324 (2021)
- Mizoguchi, R., Ikeda, M.: Towards ontology engineering. J.-Jpn. Soc. Artif. Intell. 13, 9–10 (1998)
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L.O.B., Wilkinson, M.D.: Cloudy, increasingly fair; revisiting the fair data guiding principles for the European open science cloud. Inf. Serv. Use 37(1), 49–56 (2017)
- Murdock, P., et al.: Semantic interoperability for the Web of Things. Ph.D. thesis, Dépt. Réseaux et Service Multimédia Mobiles (Institut Mines-Télécom-Télécom (2016)
- Ítalo Oliveira, E.G., et al.: Boosting D3FEND: ontological analysis and recommendations. In: Formal Ontology in Information Systems. IOS Press, Nieuwe Hemweg (2023)
- Oliveira, Í., Fumagalli, M., Prince Sales, T., Guizzardi, G.: How FAIR are security core ontologies? a systematic mapping study. In: Cherfi, S., Perini, A., Nurcan, S. (eds.) RCIS 2021. LNBIP, vol. 415, pp. 107–123. Springer, Cham (2021). https:// doi.org/10.1007/978-3-030-75018-3-7
- 58. oneM2M Partners: ONEM2M TECHNICAL SPECIFICATION. Technical Report. TS-0012-V3.7.3, oneM2M (2019). https://www.onem2m.org/images/pdf/ TS-0012-Base_Ontology-V3_7_3.pdf. type 1 (ARIB, ATIS, CCSA, ETSI, TIA, TSDSI, TTA, TTC)
- Pastor, O.: Diseño y Desarrollo de un Entorno de Producción Automática de Software basado en el modelo orientado a Objetos. Ph.D. thesis, Tesis doctoral dirigida por Isidro Ramos, DSIC, Universitat Politècnica de València (1992)
- 60. Pastor, O., Gómez, J., Insfrán, E., Pelechano, V.: The OO-method approach for information systems modeling: from object-oriented conceptual modeling to automated programming. Inf. Syst. 26(7), 507–534 (2001). https://doi.org/10. 1016/S0306-4379(01)00035-7. http://www.sciencedirect.com/science/article/pii/ S0306437901000357
- Simperl, E., Bürger, T., Hangl, S., Wörgl, S., Popov, I.: Ontocom: a reliable cost estimation method for ontology development projects. J. Web Semant. 16, 1–16 (2012)
- Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: principles and methods. Data Knowl. Eng. 25(1–2), 161–197 (1998)
- Uschold, M., Gruninger, M.: Ontologies and semantics for seamless connectivity. ACM SIGMOD Rec. 33(4), 58–64 (2004)
- Uschold, M., Gruninger, M., et al.: Ontologies: principles, methods and applications. Technical Report-University of Edinburgh Artificial Intelligence Applications Institute Aiai TR (1996)
- Van Heijst, G., Schreiber, A.T., Wielinga, B.J.: Using explicit ontologies in kbs development. Int. J. Hum.-Comput. Stud., 183–292 (1997)
- Wilkinson, M.D., et al.: The fair guiding principles for scientific data management and stewardship. Sci. Data 3(1), 1–9 (2016)