

Ontology Summit 2021 Communiqué: Ontology Generation and Harmonization

Second Draft: 9 June 2021

Abstract: This Communiqué surveys the landscape of ontologies throughout their lifecycle, including generation and development, harmonization with both human and machine environments, and sustainability. [The abstract will be completed later.]

Keywords: Ontology, Machine Learning, Definitions, Sustainability

1. Introduction

Ontologies are proliferating, producing a complex landscape of many types, roles and uses for many purposes. Ontologies can be extracted, learned, modularized, interrelated, transformed, analyzed, and harmonized as well as developed in a formal process; and there are now many ways that ontologies interact with other technologies, including but not limited to, generation by machine learning tools, serving as the basis for machine learning so as to improve the quality and explainability of its results, and integration into machine learning architectures. The variety has made it difficult to communicate about and between ontologies. There is a need for harmonization and better definitions of terms in ontologies.

The Ontology Summit 2021 examined the topic of ontology generation and harmonization at a series of virtual sessions from February to May 2021. This Communiqué synthesizes and summarizes the findings of this series. The Ontology Summit 2021 was organized into four tracks and this Communiqué is also organized based on these tracks. The Ontological Landscape track provided an overall framework for different kinds of ontology and for the rest of the summit tracks. The results of this track are summarized in Section 2.

Definitions are an essential part of ontologies, but they are often poorly written either because of a lack of experience with writing definitions or because of a lack of emphasis on properly defining terms during the ontology development process. The notion of a definition and its role in ontology development and harmonization are presented in Section 3.

The recent development of much more effective machine learning techniques has made it possible to extract ontologies from source documents. While such techniques are effective, they are generally

29 opaque and do not lend themselves to explainability. This issue was examined in a previous Ontology
30 Summit [Ontolog:19b]. The architectures that have been developed for integrating ontologies with
31 machine learning, and more generally integrating symbolic (which includes ontologies) and
32 subsymbolic (such as machine learning) techniques, are presented in Section 4.

33 Furthermore, it isn't enough to construct ontologies with well developed definitions and
34 documentation, they must be sustained over time. There is much more to sustainability than just
35 ensuring that it has a continuing source of funding. The results of the Sustainability of Ontologies track
36 are presented in Section 5.

37 Section 6 summarizes the harmonization challenges that were identified by the Ontology Summit. The
38 Communiqué ends with a conclusion and acknowledgments.

39 **2. The Ontological Landscape**

40 This section surveys the many aspects and dimensions of ontologies and their purposes. This is a topic
41 that was studied and surveyed in the first two Ontology Summits in 2006 and 2007. The first of these
42 was the Upper Ontology Summit [Ontolog:07h], and the second was on "Ontology, Taxonomy,
43 Folksonomy: Understanding the Distinctions" [Ontology.Summit:06]. A number of dimensions were
44 identified in these two summits, and some of the most prominent ontologies of the time were classified
45 along the dimensions [Baclawski:07o,Baclawski:07p,Baclawski.Duggar:07]. The most common
46 classification of ontologies is by their level of generality, and Section 2.1 discusses this dimension of
47 the landscape. One of the primary purposes of ontologies is for communication, and Section 2.2
48 surveys this aspect of the landscape. Section 2.3 discusses the different approaches to what entities of
49 an ontology represent. Sections 2.4 and 2.5 survey two more dimensions: attitudes toward realism and
50 representation of uncertainty. This section ends with a short discussion of ontological commitment in
51 Section 2.6. However, there are still other dimensions in the landscape, and these are surveyed later on
52 in Sections 3, 4 and 5.

53 **2.1 Generality**

54 Ontologies are most often classified by their level of generality. The most generic ontologies are called
55 foundation ontologies, top-level ontologies, and upper ontologies. The most specific ontologies are
56 called application ontologies because they are generally associated with a specific application or
57 narrow range of applications. In between these two extremes, there are reference ontologies and
58 domain ontologies. Reference ontologies are more specific than foundation ontologies but are not
59 limited to a particular domain. Domain ontologies are limited to a single domain, but domains can
60 form hierarchies with many levels of generality and the domain ontologies may also have many levels
61 of generality. [Schneider:21]

62 2.2 Communication

63 One of the main purposes of ontology is to improve communication. Communication can be between
64 people, between people and machines, and between machines. Machines operate best when the terms
65 are precise and logical so that there is no ambiguity. Human language, on the other hand, is much
66 richer with large numbers of types of figures of speech and inherent ambiguities. [Baclawski:21w] Far
67 from being a flaw of human language, this richness is its strength. Humans deal with the ambiguity of
68 language by both using the context of a dialog and asking for clarification. [Sowa:21] Accordingly,
69 ontologies should recognize the distinctions between the needs of humans and machines. One means
70 of accomplishing this is to have a "language interface" that mediates between the human and machine
71 terminology. The language interface is an important feature of modern software engineering processes,
72 and ontology development can also benefit from these practices. [Bennett:21], [Woods.Low:21]

73 One of the functions of a definition is to adjust the readers' or systems' inferential competences, i.e.,
74 what they infer when encountering the term that is defined [Seppälä.et.al:16f]. For a theoretical
75 explanation of the functions of definitions in ontologies backed by empirical neuropsychological
76 studies see [Seppälä.et.al:16].

77 Another effective means for communication is to base the ontology on the events and states of interest
78 in a domain, as well as the arguments (subjects, objects, adverbs, etc.) of those events (as per neo-
79 Davidsonian semantics). This allows the ontology to be deeply informed by narrative and event
80 linguistic theory. [Westerinen:21]

81 For a computable ontology using OWL, Common Logic or other logic based ontology representation
82 languages, the identifiers (for elements of the ontology) are, in fact, symbols of the signature of the
83 (logical) language. However, humans are already familiar with such natural language terms or phrases,
84 sometimes in highly emotional and biased ways. As a result, humans can easily forget that the
85 identifiers need to be treated as symbols whose interpretation is dictated by the formalizations used to
86 define the identifiers. This problem is complicated when an ontology element is not adequately defined
87 to meet the intended interpretations. This can happen when the ontology developer implicitly relies on
88 the expectation that a user of the element will 'understand' how the element is to be interpreted, based
89 on an expected understanding of the natural language term or phrase similar to that of the ontology
90 developer but which may not be shared by all of the users of the ontology. [Schneider:21]

91 Well defined concepts are an essential ingredient in communication and play a central role in semantic
92 interoperability. Both communication and interoperability are based on a common understanding of
93 concepts, services, information and contributing data. Some of this was discussed at the Ontology
94 Summit 2007, whose theme was "Ontology, taxonomy and folksonomy: Understanding the
95 distinctions". [Ontolog:07h] This summit discussed the spectrum of useful semantic artifacts starting
96 with an implicit preference for strong formality found in ontologies. But a range of resources exist such
97 as folksonomies (such as simple user-defined keywords lists useful to annotate resources on the Web)
98 along with taxonomies, conceptual models and what are considered weak semantic resources of
99 controlled vocabularies such as Medical Subject Headings (MeSH). Vocabularies like MeSH reflect
100 association of terms and are part of language use which, like a model, impose some simplifying,
101 constraining, and organizing form on the fluxing complexity we call reality about which we

102 communicate. The impetus in these early days was on formalizing concepts in a logical language to
103 allow some degree of automated, interoperable processing of data. To achieve this goal definitions
104 should be represented in a formal, logical language. There was less attention at the Ontology Summit
105 2007 to standardization of other forms of semantic resources in pre-formalization phases. More
106 recently, there has been a greater recognition and appreciation of both formal ontologies and standard
107 vocabularies to support research communication, data sharing and interoperability. The effort to
108 control the meaning of terms in vocabularies requires a lifecycle of their own, like other digital data
109 lifecycles.

110 **2.3 Ontology Styles**

111 There are many different questions that must be answered that influence how an ontology is developed
112 and how it can be reused. These include:

- 113 • What is the overall domain or ‘system’ that is being addressed?
- 114 • What are the ‘things’ in that system?
 - 115 ◦ How are they distinguished?
 - 116 ◦ Are there basic relationships or interactions between the things?
- 117 • What level of specificity is needed?
- 118 • Are basic relationships such as whole-part needed?
- 119 • What does it mean to be a ‘thing’? [Bennett:21]

120 For an ontology, especially a very large one, to be consistent in the broader sense of this term, it must
121 not only be logically and conceptually consistent, it must also have a consistent style throughout. This
122 requires conformance to style guidelines, such as naming conventions and change management
123 procedures. [Uschold:21] This is related to the issue of sustainability that is discussed in more detail in
124 Section 5.

125 **2.4 Realism**

126 Another dimension of ontology is the degree of "realism". Philosophical realism is the thesis that there
127 is a reality that exists independently of people. An ontology is realist if it is based on this thesis. One
128 consequence of realism is that one should only define classes that have at least one instance.
129 Ontologies do not necessarily require realism; indeed, computer scientists' focus is generally more
130 pluralistic with respect to what entities may be relevant, whether they are ontological (in the
131 philosophical sense), conceptual, cognitive, hypothetical, etc. From this point of view in knowledge
132 representation and conceptual modeling there is no absolute reality. This allows for many different
133 "realities" and views. This has benefits for dealing with natural language, common sense, and other
134 human capabilities. Another consequence of a pluralistic attitude toward realism is that there is no
135 absolute truth. There are different notions of truth that underlie different reasoning mechanisms (i.e.,
136 logics) and that depend on context and situation. [Masolo:21] Different forms of logic are a frequent
137 topic of earlier Ontology Summits, including the ones in 2007 "Ontology, Taxonomy, Folksonomy:

138 Understanding the Distinctions" [Ontology.Summit:07], in 2010 "Creating the Ontologists of the
139 Future" [Ontolog:10b], and in 2017 "AI, Learning, Reasoning, and Ontologies" [Ontolog:17b]. One of
140 the most important examples of a different reasoning mechanism is probabilistic reasoning that is
141 discussed in the next section.

142 **2.5 Uncertainty**

143 Human understanding is probabilistic. To reflect our best understanding of the world, ontologies must
144 be, too. However, there is no generally agreed upon way to specify uncertainty in an ontology. For
145 example, the likelihood of a specific RDF statement being true can be defined as simply as adding
146 a :probability predicate as an RDF* edge property or to a reification of the RDF statement
147 [Westerinen:xx]. But, perhaps having flexibility in the definition of ontological probabilities is an
148 advantage. It may be that the notion of uncertainty, like other aspects of an ontology, will depend on
149 the particular domain, community and purpose. Rather than attempt to create a single reference
150 ontology for uncertainty, there should be a meta-theory for uncertainty that is modularized and
151 customizable for the particular task and purpose. [Breiner:21]

152 **2.6 Commitment**

153 Ontology development is similar to software development; indeed, ontologies may be regarded as a
154 kind of software. Consequently, many of the same issues arise. In particular, during the development
155 process, many design decisions must be made. These decisions vary with respect to how much of the
156 ontology is affected. Furthermore, different ontology development methodologies vary with respect to
157 what decisions are made, who makes the decisions, as well as when and how the decisions are made.
158 While modularity can help limit the scope of a decision, there will nevertheless be decisions that have
159 major consequences for the process and the artifacts that are developed. The design decisions also
160 affect how well an ontology can adapt to future requirements. These design decisions are commonly
161 referred to as "ontological commitments". When using this term in the context of ontology
162 development, one must be careful not to confuse the term with the philosophical notion of ontological
163 commitment, such as Quine's Criterion.

164 **2.7 Understanding and Analysis**

165 Understanding and analysis of the notions to be represented in an ontology are fundamental problems
166 in ontology development that are two of the main aspects of ontology development that distinguishes it
167 from software or even systems development. In ontology development there are (at least) two phases
168 to representing a notion or entity. First, there is understanding what the notion or entity is. The second
169 is the analysis phase in which the notion or entity is represented (or modeled) using the selected or
170 available constructs. For example, one may have selected OWL as the ontology language or one is
171 working within the context of a foundational, reference, or domain ontology.

172 During the 'understanding' phase the developer needs to survey relevant references, starting with a
173 'common' dictionary (e.g., Oxford or Cambridge) and moving on to more domain specific references
174 (e.g., ISO specifications), if they exist. Using a range of references can provide a contrast among the
175 different ways a notion or entity is understood as well as the contexts in which the notion or entity

176 occurs. These contrasts will greatly aid in ‘understanding’ the notion or entity and what may be needed
177 to be represented to meet the stated needs of the ontology being developed. Ideally, the understanding
178 phase should lead to a useful natural language definition, a critical element in representing a notion or
179 entity.

180 The next phase is the ontological analysis process. Note that the understanding and analysis phases are
181 usually performed in parallel rather than in series. An aid in ontological analysis can be a well
182 constructed foundational ontology (e.g., DOLCE, BFO, UFO), since such an ontology has already
183 incorporated ontological analysis during its creation. One common pitfall during the analysis phase is
184 ‘taxonomy seduction’. This is the tendency to place an entity into a taxonomy without completing an
185 understanding or analysis. Premature ‘classification’ of a notion or entity can bias an entire
186 development effort and/or require rework.

187 **3. Definitions**

188 A definition is a statement of the meaning or significance of an entity, including words, phrases, classes
189 and properties. Definitions can serve as links between humans, between human communities, between
190 humans and ontologies, as well as between different ontologies. Historical attempts to standardize
191 terms included creating core metadata models and common conceptual models for combining data into
192 a single representation. These however have largely failed to be adopted because of flawed
193 conceptualizations, lack of community agreement, and inadequate representation; and thus they have
194 resulted in silos. Some progress has been made leveraging best practices including the use of
195 ontological analysis and design. This section surveys the different notions and levels of formality of
196 definitions, with emphasis on practical methods to harmonize a variety of semantic resources.

197 Domain vocabularies vary significantly in quality and scope, often with alternative definitions for the
198 same term and definitions that have varying degrees of formality. These problems have been
199 recognized for a long time. It is said that when Confucius was asked what he would do if he was a
200 governor, he replied that he would "rectify the names" to make words correspond to reality.
201 Standardization of term meanings is challenging since there are many conflicting and overlapping
202 glossaries or data models that define domain terms in idiosyncratic ways. Completely "rectifying the
203 names" may be too ambitious a goal, but harmonization can be achieved, albeit with some effort.
204 Harmonizing terminology is underway in some domains, such as the cryosphere, which is concerned
205 with ice fields and glaciers.

206 Writing a good definition is not as easy as one would expect. It is even more difficult when one must
207 harmonize many ontological entities in a single definition. Indeed, one of the main problems with
208 harmonizing definitions for ontologies is precisely the common assumption that anyone can write
209 definitions and that they will then be harmonious. In fact, it requires training and experience.
210 Experiments have shown that even for well understood domains, the results of manual classification
211 tasks performed by domain experts are highly inconsistent. [Westerinen:21] So the first step in writing
212 definitions is to accept that one does not know how to do it. The next step is to start learning how. The
213 ontology community is fortunate in having available a set of principles and conventions to assist in

214 definition writing, editing, and validating, by drawing on existing definition writing principles and
215 guidelines in lexicography, terminology, and logic [Seppälä.Ruttenberg.Smith:17].

216 The rest of this section gives examples and lessons learned about developing and harmonizing
217 definitions.

218 **3.1 Definitions in the Environmental Sciences**

219 In this section, a specific example is presented in detail of a domain that has addressed semantic
220 harmonization. The domain is the environmental sciences, and the ontology is EnvO. [EnvO] EnvO is
221 a widely cited semantic resource for semantically controlled descriptions of environmental entities. For
222 example, the Darwin Core glossary uses EnvO in its habitat descriptions and was developed by
223 applying text-mining approaches to extract habitat information from the Encyclopedia of Life and
224 automatically create experimental habitat classes within EnvO.

225 Here is an example of an EnvO definition: A habitat is “An environmental system which can sustain
226 and allow the growth of an ecological population.”

227 EnvO’s initial focus was to represent biomes, environmental features, and environmental materials, and
228 the initial purpose was for genomic and microbiome-related investigations. However, the need for
229 environmental semantics is common to a multitude of fields, and EnvO's scope has steadily grown
230 since its initial description. As the scope has expanded, the ontology has been enhanced and
231 generalized to support its increasingly diverse applications (such as the Cryo (glaciers and ice fields)
232 and the Marine (ocean) realms) that are now examined in more detail.

233 The global cryosphere watch (GCW) has sponsored the effort to harmonize term definitions that are
234 already in use in the Cryo domain. Sometimes a single term will have dozens of different term
235 definitions. Consider the term "snow cover". The GCW harmonized definition is the following:

236 "An area density which inheres in snow distributed over an area of a landmass or other substrate."

237 Additional information about this term is provided by annotations. For example, one annotation states
238 that, in general, snow cover is a layer of snow on the ground surface and can be compared to the related
239 terms of snowfield and snowpack. Annotations can also be expressed using logical axioms. One such
240 annotation is the following:

241 "has part some (surface layer and (composed primarily of some snow))"

242 EnvO definitions conform to the Minimum Information for the Reporting of an Ontology (MIRO)
243 guidelines. [Matentzoglou.et.al:18]

244 The EnvO harmonization effort has had a number of accomplishments.

- 245 1. The GCW glossary analysis results were harmonized with EnvO and aligned with
246 corresponding terms in the Semantic Web for Earth and Environment Technology (SWEET)
247 ontology.
- 248 2. The alignment with SWEET has improved the definitions of SWEET terms.

3. EnvO and SWEET terms have been aligned with other OBO Foundry ontology terms.
4. A special envoPolar subset of EnvO has been crafted with relevant terms and axioms.
5. Attribution for all of these updates have been documented to the people or groups responsible for these changes.
6. Best practices for documentation were instituted, including annotating the time the definition was added, the orcid.org of the individual who created the update, and provenance information about definition derivation from EnvO to SWEET.
7. Templating methods were employed to accelerate class creation, and spreadsheets were used to help update ontologies with definitions. Specifically, the OBO Robot tools was used to facilitate EnvO collaboration. This tool guides users through the process of creating new terms and is intended to be used by non-ontologists. Robot organizes new term requests in a standardized google sheet template, and users can follow a step by step process to fill out the appropriate spreadsheet columns.

3.2 Lessons Learned and Best Practices

In this section, some of the best practices for harmonization of definitions are presented. For general advice about developing definitions, see the Guidelines for Writing Definitions in Ontologies. [Seppälä.Ruttenberg.Smith:17] The guidelines include advice such as: be brief, align, re-use, extend, and revise semantic resources. The following are other best practices and advice [Berg-Cross:21]:

1. Organize the terms into a concept system (network of concepts), and indicate the position of each term in the concept system. Somewhat informal conceptual models may help organize the networks during the early phases of ontology development.
2. To ensure that definitions are brief, put other, more encyclopedic information in annotation notes.
3. To structure taxonomies use the nearest generically super-ordinate concept, adding one or a few constraining characteristics.
4. Consider harmonizing the role of lexical modifiers as well as nouns and verbs. For example, words like alpine or liquid.
5. Reuse existing vocabularies such as Schema.org, DCAT2, VIVO, DDI, etc. However, doing so will usually involve specializing the more general term.
6. Reduce (conceptual) ambiguity by explicitly showing relationships between the terms.
7. Be sensitive to issues of granularity. Metonymy is the naming of a thing by something related to it. For example, using the name of the whole for a part or vice versa. Figures of speech such as metonymy are so commonplace that one may not be aware that one is using a figure of speech at all.
8. Avoid vague comparisons, such as using the word "similar" in a definition. There will usually be many similar terms and terms with overlapping meanings. Perform semantic analysis to

distinguish similar terms. This involves establishing relations among terms where appropriate, such as subtypes, part-whole relationships, roles, influences, production (output) relationships, etc.

9. The interface between the business (natural) language and the technical language (FOL) helps human-machine communication by making language more precise and less ambiguous. [Woods.Low:21] The impact of differences between the textual and logical definitions in ontologies was examined in [Seppälä.et.al:16f]. However, when communicating with natural language, one must be aware of how people categorize the world. Unlike the classes and properties of ontologies, human categories "shimmer". [Baclawski:21w]

10. Terms may be standardized but the meaning has to be too. [BergCross:21]

11. Definitions will vary according to context of use and target audience. [BergCross:21]

These lessons of the EnvO community can be useful for other communities. In Section 5 below, the issue of sustainability of ontologies is discussed in general.

4. Neuro-Symbolic Learning Ontologies

Symbolic reasoning has a long history, and continues to be an active area of research. Machine learning, also known as subsymbolic methods, is also a very active area of research. Although both are part of AI, they two areas have been developed under clearly distinct technical foundations and by separate research communities. The two areas have complementary strengths and weaknesses. As a result, finding ways for bridging the gap between symbolic and subsymbolic approaches to AI is a long-standing unresolved challenge, and integrating these two areas is now the subject of growing research interest in AI. This topic was addressed in the Ontology Summit 2017 "AI, Learning, Reasoning, and Ontologies" [Ontolog:17b], but new AI techniques, especially in machine learning, have since been developed so that revisiting this topic is certainly timely.

Neuro-symbolic learning aims to integrate neural learning with symbolic approaches typically used in computational logic and knowledge representation in AI. One benefit of such an integration is the development of effective knowledge extraction methods towards explainable AI, which was examined in the Ontology Summit 2019 [Lamb:21],[Ontolog:19b], but there are many other advantages. While there are significant benefits for tighter integration of neural and symbolic paradigms, it not known how best to integrate them, and many integration architectures have been proposed. Symbolic models can be the result of, or the basis for, different stages of a neural process. The following are some of the architectures of techniques that integrate symbolic and subsymbolic methods [Kautz:21]:

1. The simplest and most common architecture is one in which symbolic data (e.g., documents) are processed with symbolic techniques to produce vectors that are input to a subsymbolic module (e.g., a neural network). The vector output of the subsymbolic module is then interpreted in symbolic form using symbolic techniques.
2. Another architecture is a symbolic system that can invoke subsymbolic submodules. As far as the symbolic system is concerned, the submodules are just subroutines like any others. This is the architecture commonly used by self-driving vehicles.

3. One could, in principle, invert the roles of symbolic and subsymbolic in Architecture (2) above to get an architecture in which it is the subsymbolic system that is invoking the symbolic one. This architecture would be useful for dealing with complex decision making, since symbolic reasoners can perform combinatorial reasoning much more scalably and efficiently than subsymbolic systems.
4. Architectures (1), (2) and (3) above do not learn during normal processing. The symbolic and the subsymbolic modules have already been programmed and trained, respectively. Some recently developed architectures incorporate symbolic reasoning in the subsymbolic system by organizing the subsymbolic system according to symbolic rules. Examples of this architecture are tensor product representations and logic tensor networks. This architecture is useful for finding concept hierarchies such as generalization and part-whole hierarchies. In other words, this architecture can generate ontologies, or at least some aspects of ontologies. The forms of reasoning that can be incorporated include temporal logic, description logic, and first-order predicate logic. [Hitzler:21]
5. Another learning architecture is to input instances of logical inferences, expressed as input-output pairs, to train a subsymbolic model. This is primarily useful for mathematical problems, and it is surprisingly effective, although it will make mistakes sometimes. [Kapanipathi:21]
6. Another architecture uses subsymbolic methods, such as back-propagation, to train a symbolic system. This can be useful for question-answering systems. The back-propagation is invoked whenever the system makes a mistake.

Another reason why subsymbolic methods need knowledge based methods is the intuition, based on human behavior, that intelligence necessarily involves learning, knowledge from experience, and reasoning, which could be expressed as an equation like this:

$$\text{Intelligence} = \text{Learning from Data} + \text{Knowledge/Experience} + \text{Reasoning} \text{ [Sheth:21]}$$

This intuition has been used as the basis for Knowledge-Infused Learning (KIL). The KIL notion is a collection of architectures that range from shallow to deep infusion of knowledge, allowing one to match an architecture with a particular application. KIL has the potential to have impacts on robotics, cognitive science, autonomous vehicles, and personal assistants. [Sheth:21]

Yet another architecture is to start with a "seed" ontology and augment it incrementally using signal processing techniques from knowledge graphs extracted from various sources, such as documents. Signal processing is a very highly developed and active area with a large community, and this connection between this community and the ontology community could have many benefits [Majumdar:21],[Sowa:21].

The survey given in this section has attempted to give a rough organization of the large variety of neuro-symbolic architectures that are now being developed, but one can expect new architectures to be developed by researchers in this very active field.

359 5. Sustainability of Ontologies

360 Many organizations, including government agencies, standards bodies and commercial firms, use
361 ontologies and have developed tools for various ontological activities, such as creation, evolution,
362 mapping and other forms of harmonization. The EnvO example in Sections 3.1 and 3.2 above
363 illustrates how a community is addressing the sustainability of their ontology. The lessons that they
364 learned are valuable for other communities. This section discusses the issues involved in sustainability
365 of ontologies in general.

366 Sustainability involves addressing much more than simply ensuring sufficient funding. Achieving
367 sustainability fully requires building on a firm foundation. The most important aspects of such a
368 foundation are the following three "pillars" [Dickerson:21]

- 369 1. Economic Viability. This pillar is the most obvious. One must ensure there is sufficient
370 funding for maintaining the ontology for as long as its purpose remains relevant. The manner in
371 which resources are allocated and monitored determines whether economic viability is
372 achieved. Without proper oversight, economic viability cannot be maintained.
- 373 2. Social Equity. Ontologies can have subtle biases, especially when ontologies are generated
374 using machine learning. The problem is that data invariably has biases to a greater or lesser
375 extent, and machine learning techniques cannot find or correct them on its own. Combining
376 symbolic knowledge with machine learning can help to discover and mitigate bias, but the most
377 important problem is accepting that correcting bias is an important part of ontology
378 development and maintenance. Using standards and rigorous methodology can also assist in
379 ensuring social equity is adequately addressed. Standards help provide a common baseline for
380 involved parties, so a wider group of people are able to fully participate in development. A
381 thorough approach helps establish quality, which also encourages an unbiased approach.
382 [Dickerson: 21]
- 383 3. Environmental Protection. The term "environment" in this pillar refers to the human
384 environment that surrounds the specific community that developed an ontology. It is important
385 to recognize that communities and their ontologies do not exist in isolation. One must maintain
386 avenues of communication and cooperation with adjacent and other related communities.

387 Well designed definitions, documentation and harmonization, as discussed in Sections 3 and 4 above,
388 can contribute both to the Social Equity and the Environmental Protection pillars of sustainability.

389 The pillars of sustainability illustrate that there is much more to ontology development than simply
390 creating the ontology. The first round of ontology development is only the beginning; one must put it
391 into production. This challenge is similar to ones found in releasing software.

392 First, planning is needed to execute the release. This includes infrastructure concerns, engagement of
393 stakeholders, and scheduling associated tasks [Franch.Ruhe:16]. Once an ontology is released and is
394 actively being used by those outside the development group, issues usually arise. A development team
395 can only anticipate so many potential issues and scope an initial release to contain so much content.
396 Those issues must be managed appropriately or the post-release viability of the ontology will be in

jeopardy. Mechanisms must be implemented to facilitate revisions and as appropriate, expansions to the original model. Effective feedback channels allow for building a strong environment for an ontology to mature. Stakeholders can be resources to edit the content of an ontology as well as enforce equity through promotion of standards. Maintaining the technical infrastructure furthers the intellectual and collaborative infrastructure required to sustain ontologies for the long term.

As research methodology and scope have expanded with technological advances, the management of ontologies and other related semantic resources has become a critical and distinct component of the ontology lifecycle. The scale and diversity of new semantic resources, such as knowledge graphs, neuro-symbolic generated products and domain vocabularies, requires a reexamination of ontological engineering practices, and the various roles of ontologies in the overall semantic research enterprise.

6. Harmonization Challenges

This section briefly summarizes the challenges of ontology harmonization. The most common context for ontology harmonization challenges is during ontology development, but there are harmonization challenges throughout the ontology lifecycle.

- The use of natural language for the symbols of an ontology can result in confusion because of existing meanings that humans have for the symbols. [Section 2.2]
- Accordingly, ontologies should recognize the distinctions between the needs of humans and machines. [Section 2.2]
- One of the functions of a definition is to adjust the readers' or systems' inferential competences, i.e., what they infer when encountering the term that is defined. [Section 2.2]
- An effective means for communication is to use narrative and event linguistic theory. [Section 2.2]
- The effort to control the meaning of terms in vocabularies requires a lifecycle of their own, like other digital data lifecycles. [Section 2.2]
- During ontology development and maintenance, develop and conform to style guidelines. [Section 2.3]
- The logic underlying an ontology should be selected according to the requirements of the users of the ontology. It is especially important to include an appropriate notion of uncertainty. [Section 2.4 and 2.5]
- Be careful not to confuse ontological development design decisions with the philosophical notion of ontological commitment. [Section 2.6]
- Avoid taxonomy seduction, i.e., prematurely placing entities into taxonomies. [Section 2.7]
- To better understand the domain of an ontology, survey all existing relevant references. [Section 2.7]
- An aid in ontological analysis can be a well constructed foundational ontology. [Section 2.7]
- The first step in writing definitions is to accept that one does not know how to do it. The next step is to start learning how. [Section 3]

- Harmonization can be very time-consuming, so tools are important to simplify the effort required as well as to manage the effort over time. The challenge is to develop better tools for harmonization. [Section 3]
- Community agreement is needed for extensions and revisions. Good mechanisms for community discussion are important as are partner agreements with groups with domain vocabularies. A variety of tools are used for coordination and harmonization, such as Slack and Github, but community members are not necessarily skilled in the use of these tools. The challenge is to develop state-of-the-art tools and techniques for reaching agreement efficiently with a diverse community. [Section 3]
- Another challenge is maintaining access to the vocabulary for reuse or for alignment with ontologies. Some preliminary mechanisms have been established for this, but improving these mechanisms is needed. [Section 3]
- Conform to the MIRO guidelines. [Section 3.1]
- Follow the Guidelines for Writing Definitions in Ontologies. [Section 3.2]
- Organize terms into a concept system. [Section 3.2]
- Harmonize the role of lexical modifiers as well as nouns and verbs. [Section 3.2]
- Be sensitive to issues of granularity. [Section 3.2]
- Avoid vague comparisons. [Section 3.2]
- The interface between natural language and technical language helps human-machine communication. [Section 3.2]
- Terms may be standardized but the meaning has to be too. [Section 3.2]
- Use a neuro-symbolic learning architectures that incorporates symbolic reasoning in the subsymbolic system. [Section 4]
- Intelligence = Learning from Data + Knowledge/Experience + Reasoning [Section 4]
- Ensure that all three pillars of sustainability are well founded. [Section 5]
- The scale and diversity of new semantic resources, such as knowledge graphs, neuro-symbolic generated products and domain vocabularies, requires a reexamination of ontological engineering practices. [Section 5]

7. Conclusion

8. Acknowledgments

Certain commercial software systems are identified in this paper. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology (NIST) or by the organizations of the authors or the endorsers of this Communiqué; nor does it imply that the products identified are necessarily the best available for the purpose. Further, any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not

469 necessarily reflect the views of NIST or any other supporting U.S. government or corporate
470 organizations.

471 We wish to acknowledge the support of the ontology community, especially the invited speakers and
472 participants who contributed to the Ontology Summit. There were many invited speakers, some of
473 whom gave presentations at more than one session. The complete list of sessions, speakers, and links
474 to presentation slides and video recordings is available at <https://bit.ly/3gKukvC>

475 [References are in a separate document.]