Documentation of epistemic metadata by a mid-level ontology of cognitive processes

Martin Thomas Horsch[1], Björn Schembera[2]

[1]Norwegian University of Life Sciences, Faculty of Science and Technology, Department of Data Science, Drøbakveien 31, 1430 Ås, Norway

[2]University of Stuttgart, Faculty of Mathematics and Physics, Institute of Applied Analysis and Numerical Simulation, Pfaffenwaldring 57, 70569 Stuttgart, Germany

Abstract

Epistemic metadata, *i.e.*, knowledge and validity claims, provenance descriptions, and grounding of research conclusions, contain information that is relevant to substantiating the knowledge status of data. They constitute the annotation required for communicating interpretations that raw research outcomes are given by researchers, and for making these interpretations intelligible and acceptable to others. Therefore, epistemic metadata are that which needs to become FAIR whenever epistemic opacity and darkness of data are to be avoided. The present work addresses the challenge of documenting epistemic metadata in a way that integrates well into the pre-existing semantic landscape from national research data infrastructures and European digitalization platforms such as those related to EOSC and EMMC ASBL. It is explored how an ontology that formalizes cognition in a way inspired by the triadic structure of Peircean semiotics can advance FAIRness of epistemic metadata and support the development of materials digitalization platforms such as molecular modelling interoperability infrastructures.

1. Introduction

Any *data about data* are metadata; by *epistemic metadata*, we refer to data that document, or contribute to documenting, the knowledge status of other data. In other words, these are metadata that contain, denote, or express "epistemically relevant elements" [1] supporting some meaning or interpretation that has been given to other data. Epistemic metadata are that which would be needed to respond to the queries: What *knowledge claims* have been formulated or should be formulated on the basis of the given data, and what exactly is the relation between the knowledge claims, their proponents, and the data? This is very close to asking directly: How can or should I reuse the data?

Here, we argue that for this reason, epistemic metadata are central to digitalization and FAIR data management in engineering research and development: They are what is needed to avoiding epistemic opacity and darkness of data [2]. We also argue that there is a hierarchy of requirements, and that it is *first and foremost the knowledge claims* that need to become FAIR: The knowledge claims are that which is reused at the semantic level whenever, at a technical level, the data are being reused. Other epistemic metadata can support this reuse

The Eighth Joint Ontology Workshops (JOWO'22), August 15-19, 2022, Jönköping University, Sweden

martin.thomas.horsch@nmbu.no (M. T. Horsch); bjoern.schembera@mathematik.uni-stuttgart.de (B. Schembera)

0000-0002-9464-6739 (M. T. Horsch); 0000-0003-2860-6621 (B. Schembera)

in important ways, *e.g.*, regarding reliability or reproducibility [3]. Therefore, FAIRness of the *provenance and other epistemic metadata is secondary*; it is conditioned upon FAIRness of metadata that report what the knowledge claims are. Documentation requirements along these lines will need to be placed at the centre of metadata standardization efforts in Industry 5.0 related Horizon projects and disciplinary research data infrastructures that cover domains of knowledge from engineering. This will have to occur at the expense of schemas and practices that are ultimately based on the legacy paradigm of proper librarianship [4]. These legacy practices do not optimally support our contemporary paradigm of digitalization in engineering by Industry Commons whereby an automated and quantitatively reliable reuse is to be made of pre-existing research outcomes [5]. There, it is primarily the knowledge claims that need to become machine-actionable [6]. Documentation requirements that focus on the wrong sort of metadata – or that simply lack a clear focus on the right metadata, making them excessively complicated to use – could instead become a barrier to digitalization. In materials modelling, this has been the case with MODA [7], which it takes an immense effort to fill in (usually, by hand [8, pp. 79–374]), while it is hard to recognize any actual contribution toward making the research outcomes machine-actionable [ibid.].

To support advancing in the direction suggested above, we here report on work in progress toward documenting knowledge claims as well as provenance metadata using PIMS-II [9], a mid-level ontology of cognitive processes [10]. Since knowledge is an outcome of cognition, and the provenance of knowledge is always some (individual or collective, social or technical, *etc.*) cognitive process, this is a suitable framework for positioning epistemic metadata at the mid level of an ontology ecosystem [5]. The PIMS-II mid-level ontology was presented previously in a rather formal way, with a focus on its core system of axioms in first-order logic [9]. The present work approaches its subject with a more practical motivation; it states how we expect it to advance reusability of engineering data and explains how major design decisions were made.

2. Epistemic metadata

2.1. Pre-existing semantic landscape

Ontology of cognition is a very high-level, general topic. "Great philosophers" usually develop ontologizations of cognition as major components of their systems. Pre-existing semantic artefacts in the field cover domains of knowledge as diverse as human emotional responses [11] and neuroscience [12]. This is not our aim. Neither is it our aim to review the history of ideas from philosophy; but by necessity, for working on and with a mid-level ontology of cognitive processes, some philosophical position needs to be taken *as a design choice*. The suggestion to develop metadata standards for cognitive processes by analysing them as constituted by Peircean semioses of "signs in action" [13], based on the Peircean triad $s - o - s'$ of the sign s, the object o, and the interpretant s', goes back to a series of works by Sowa [14, 15]. It was more recently successfully taken up by Ghedini and the team of developers of the Elementary Multiperspective Material Ontology (EMMO) [16, 17]. Motivated by the technical requirement to facilitate alignment of knowledge bases with the EMMO, for the present task we will follow its paradigm: That is mereosemiotics, *i.e.*, Peircean semiotics in combination with nominalism and four-dimensionalist physicalism [9, 16, 17, 18]. We conceive of cognition as a distributed, collective process as advanced by the proponents of the 4E paradigm [13].

The EMMO is an element of the interoperability architecture developed by a number of Horizon projects and supported by the European Materials Modelling Council (EMMC ASBL). The PIMS-II ontology is developed to become a mid-level component of that architecture, albeit not without neglecting the need for crosswalks with semantic artefacts from other paradigms [19]. That design choice does not come without its own philosophical challenges [18], *e.g.*, because Peirce rejected nominalism [20, 21]. For the time being, it is a completely open question to what extent the creative solutions that must and will be developed to resolve these contradictions can overall be advantageous to knowledge representation in the engineering sciences. It is at least an effort worth pursuing, one to which we intend to contribute. At the domain level, this will support interoperability with digital infrastructures that operate on the basis of domain ontologies that are aligned with the EMMO, such as BattINFO and BVCO [17] as well as the system of ontologies from the Virtual Materials Marketplace (VIMMP) project [22, 23].

2.2. Categories of epistemic metadata

Epistemic metadata documentation requirements are presently being procured within a case study evaluating representative research outcomes from molecular thermodynamics [24]. Therein, four categories of epistemic metadata are considered [10, 24]:

1. *Knowledge claims*, *i.e.*, metadata describing the main propositional content attributed by an interpreter, who may or may not be the original researcher(s), to a data item or a collection of data items, expressing their understanding of the data.
2. *Validity claims*, covering both formal verification (proofs of correctness) and any elements from the sequence of empirical validation (during the study), testing (concluding the study), and reproduction/replication [25] (subsequent to the study).
3. *Provenance metadata*, describing the cognitive process from which the data and the knowledge claims were obtained.
4. *Epistemic grounding*, which may be based on some validity or provenance documentation and the typical nature of which will depend strongly on established disciplinary practice. These are metadata from which a user with the appropriate domain background is expected to see *why the knowledge claims should be accepted.*

Knowledge claims are obtained as the *outcome of a research process.* Provenance documentation *describes the research process.* A reproducibility check relies on performing a new instance of the research process that *agrees with the original process* at least in a certain, specified way (in our terminology, agreement with the provenance orthodata [10, 26]). It is successful if it *agrees with the original outcome*, at least in a certain, specified way, namely, if it agrees with the knowledge claim orthodata [10, 26].

3. Ontology design

3.1. Mereosemiotics

Functioning within an environment of EMMO-interoperable digital platforms does not require a perfect agreement with all the minor details of the EMMO's philosophy which is somewhat idiosyncratic. Instead, we aim here at reaching a very rough agreement on basic concepts and features. The two most salient design features of the EMMO are its 4D mereotopology and Peircean semiotics. Implementing this *mereosemiotics* paradigm will be helpful for constructing alignments, *e.g.*, using SKOS relations such as `broader`, `narrower`, and `closeMatch` between concepts (and, in some cases and with some reservations, also between relations). Efforts at aligning the VIMMP ontologies [22] with the EMMO have shown that *between relations*, it is often impossible to come up with a good 1:1 alignment because domain-level relations tend to be too technical: They often denote what at a higher level would appear as the cumulative presence of multiple more generic relations. In these very frequent cases, it is therefore most convenient to subsume the domain-level relation not directly under a top-level relation, but under a chain of such relations [23, Section 5.4]. Targeting an EMMO alignment, *mereosemiotic chain relations* were found to be particularly productive; the elementary components of these chains are the proper parthood relation $\dot{\mathsf{P}}$ (and its inverse, $\dot{\mathsf{P}}^{-}$) from mereology and the representation relation R (and its inverse, R^{-}) from semiotics [23, Section 5.2]. In VIMMP, it was found to be adequate to define the mereosemiotic chain relations at the mid level of the ontology architecture [22]: Below the EMMO, in which they do not occur, but above the domain ontologies which all need to reference and use them.

With this experience as a background, PIMS-II also includes a system of mereosemiotic chain relations [9, Section 4.1]. The first-order logic axiomatization of its mereotopology [9, Axioms 1–12] was designed to make this as straightforward for end users as possible by supporting a strong and simple algebraic behaviour for proper parthood. Thereby, proper parthood is idempotent ($\dot{\mathsf{P}}^2 \equiv \dot{\mathsf{P}}$), and spatiotemporal overlap, which is needed frequently (*e.g.*, to express participation in a process), is given by $\dot{\mathsf{P}}^{-}\dot{\mathsf{P}}$ and can thus occur as a factor in mereosemiotic chain relations. The relation $\dot{\mathsf{P}}\dot{\mathsf{P}}^{-}$, on the other hand, holds between any pair of objects ($\dot{\mathsf{P}}\dot{\mathsf{P}}^{-} \equiv \top$) and can thus by eliminated as a factor from the system of chain relations completely, simplifying it substantially. This system of axioms, published in 2021, is so far the only[1]) first-order logic axiomatization of mereosemiotics.

3.2. Epistemic grounding

In the PIMS-II first-order logic release [9], Axioms (13) to (24) establish the mereosemiotic relationship between representational elements, *i.e.*, representamina (signs) and their referents (objects), and the triadic cognitive steps in which they occur, depending on the type of the cognitive step: Perception, interpretation, or metonymization. These differences concern whether the triad has the structure $s - o - s'$ (perception, interpretation) or $o - s - o'$ (metonymization), and

[1]The EMMO itself has no published first-order logic axiomatization of mereotopology yet, and to simplify a future alignment, we would recommend using the existing one from PIMS-II for the EMMO as well. To date, there has not even been a stable OWL release of the EMMO – only beta versions.

whether the object needs to be physically present (only for a perception) or whether the Peircean *real causal connection* [20, p. 142] needs to be preserved (only applicable to metonymization). Axioms (25) to (30) formalize how multiple cognitive steps are combined into an overarching cognitive process, and how this connection itself can be understood to automatically yield an epistemic grounding [9]. Accordingly, each step has a *ground*, an explanatory description or annotation; this goes back to Peirce's discussion of the semiotic triad [27, p. 99]. The duality between a *cognitive chain* and the associated *grounding chain* is illustrated in Fig. 1.

We may adapt Marr's notion of Type-1 and Type-2 theories, in Marr [28] referring to AI-system outcomes, and apply it to the epistemic grounding of knowledge claims obtained from any sort of cognition: Human or artificial, individual or collective. We then find that the schema described above and in Fig. 1 can be applied in two ways, *i.e.*, that there are two types of epistemic grounding:

1. The cognitive process on which the epistemic-grounding metadata are directly based is *not* the process by which the knowledge claims were generated; instead, it is part of the outcome: It is a process by which the claims are presented and/or reproduced in a substantially different way from their original provenance. Type-1 grounding is characteristic of mathematics, where the research process leading to a novel result is unsubstantial, what counts is the usual disciplinary choice of epistemic grounding, namely, a mathematical proof. A similar distinction between the mode of presentation and research practice applies to philosophical reasoning.[2]
2. The epistemic grounding documentation is directly given by an analysis of the provenance metadata describing the research workflow. In other words, the cognitive process itself is employed as an explanation of its own outcome. Type-2 grounding is characteristic of both experiment and simulation; it applies to all knowledge grounded in process reliabilism, *e.g.*, computational reliabilism [30, 31].

Grounding is *twice removed* from the object of research: Its immediate referent is the research outcome. The analogy here is that a knowledge claim with Type-1 grounding is like a system with a Type-1 theory following Marr [28], and the same for Type 2.

3.3. Semiotic collectives

Meurer and de Figueiredo [21] have recently raised the issue whether Peircean semiotics, given its philosophically idealist outlook, can really be suitable for formulating a materialist conceptualization of cognitive processes. On the other hand, in previous work, we commented on the contradiction between Peirce's Platonist views as opposed to the radical 4D-physicalism and nominialism put forward by Ghedini and the EMMO: It is hard to bring them together [18]. It might be said that the EMMO and its paradigm of mereosemiotics, which the PIMS-II mid-level ontology shares, is not completely Peircean. At most, it is *inspired by Peirce* insofar as it reuses the semiotic triad as a design pattern.

While this is a lucky choice for the present purpose, since mereosemiotics and its implementation will withstand Meurer's and de Figueiredo's objection, it does create challenges. Specifically, this concerns data items or symbolic entities of all sorts, which can be *the same*

[2]"Allerdings muß sich die Darstellungsweise formell von der Forschungsweise unterscheiden" [29].

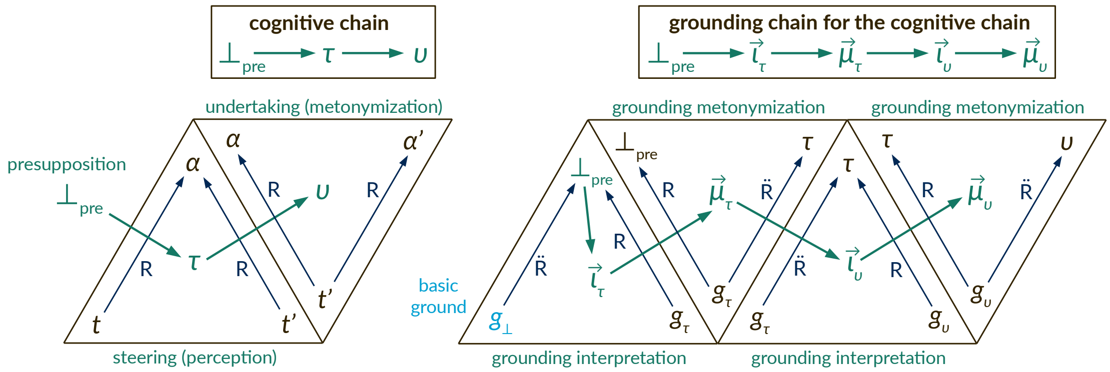

Figure 1: Left: Cognitive chain, consisting of two steps, from the PIMS-II documentation [32, 33] for one of the benchmark cases by Borgo *et al.* [34]. Right: Grounding chain corresponding to the cognitive chain on the left; g_τ is the ground of step τ, and g_υ is the ground of step υ. R stands for `isRepresentamenFor`, $\ddot{\mathrm{R}}$ stands for `isGroundFor`, and green arrows stand for `directlyGrounds`.

data item or symbol despite being expressed in different ways materially, and despite being at different positions in spacetime [35]. It is necessary to define some sort of *collective* entity for this purpose, one that groups the disconnected regions of spacetime together so that they can act as one representational element jointly. In PIMS-II, these are the four categories of semiotic collectives: Pluralities, structures, articulations, and propositions [9, Section 4.3]. Knowledge and validity claims can be subsumed under the concept of a proposition, *cf.* Fig. 2.

4. Discussion

Dark data and consequences of the darkness of data have been discussed in numerous ways [2, 36, 37, 38]. In principle, these are data whose epistemic status is unclear [10]. Darkness can be a form of appearance of data (*e.g.*, files might be missing or formatted incorrectly), but can also refer to the content (information is missing or not intelligible). Unavailable or hidden data, e.g., are dark *by form*, since they are stored in tape silos, on USB sticks, or uncurated on unsearchable file systems, irrespective of their content [2]. Darkness of data manifests itself in the *content*, e.g., when the data are statistically biased, erroneous, or epistemically unclear/opaque.

The latter manifestation, *deficiency in epistemic characterization*, is what we are mainly concerned with in the present work. When unmitigated, it can lead to massive problems in the reusability and reproducibility of scientific data. An example for the impact of content-wise dark data is the Challenger catastrophe 1986, where the space shuttle exploded just 73 seconds after the launch of the STS-51-L mission, and all crew members died. The cause of this were failing seal rings in one of the side solid rockets due to missing data in the launch preparation. In the prelaunch conference, the complete statistical data set was cut to a limited subset, leaving out the launches that did not lead to seal ring distress [37]. When a knowledge claim was formulated on this basis, namely that the launch of the rocket at the given temperature would be safe, it was thus based on data items each of which was accurate; their epistemic status, however, was not documented or retained correctly, making the data dark by content. This example can

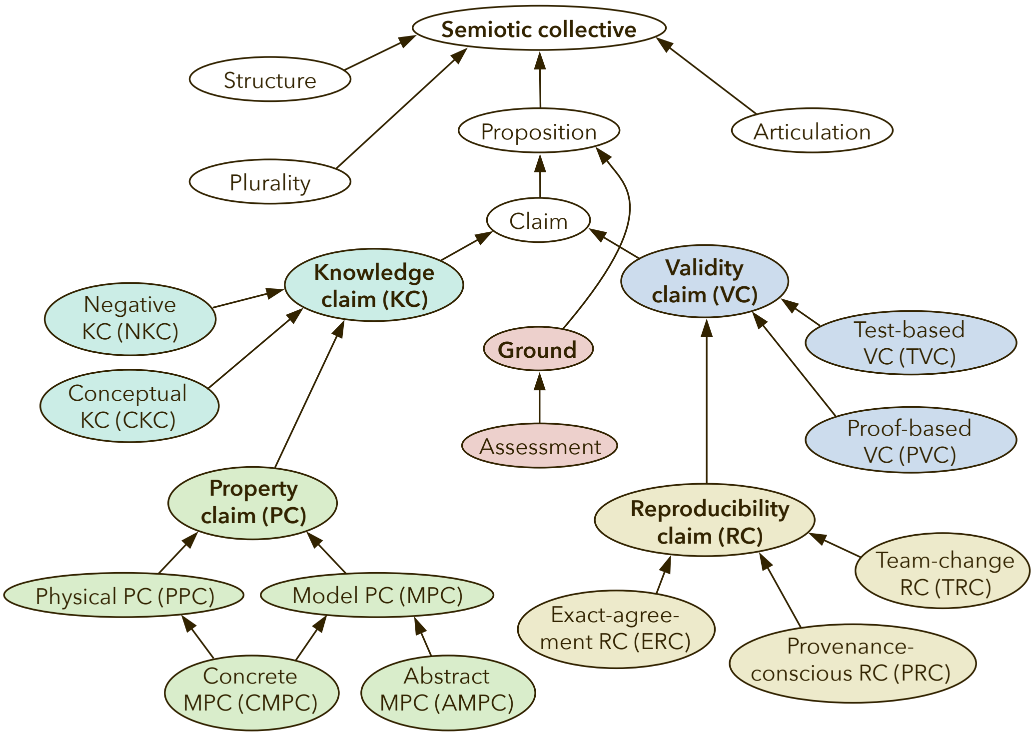

Figure 2: PIMS-II taxonomy fragment: Classes of propositions relevant to epistemic metadata considered in the ongoing case study [24]. Arrows represent subsumption $\sqsubseteq$, *i.e.*, `rdfs:subClassOf`.

be broken down into categories from Section 2.2: The knowledge claim was "the launch of the rocket at the given temperature is safe." Regarding the provenance, these data were obtained by experiments (which the data users remained aware of) and a subsequent data processing step (which they were probably unaware of). The experimental data may have been validated, but the overall research process was never validated holistically [39]. In Section 2.2, we wrote that epistemic grounding *may be* based on some provenance and validity documentation, and that it should show us why a certain knowledge claim should be accepted. In the case of the Challenger catastrophe, it was neither documented nor explained why exactly some data points were omitted. An epistemic justification for the decisive knowledge claim was formulated in darkness, and it was therefore mistaken, which eventually led to the catastrophe. De Vivo *et al.* [40] observe a similar situation in volcanological hazard assessment practices by which, they claim, epistemically ill-grounded statistics and probability analyses may obscure the actual risk of a rare catastrophic event. Obversely, documenting and retaining epistemic metadata should be expected to make a difference for the better. The use of a metadata standard for epistemic metadata as such is of course not enough to avert a catastrophe by itself. But it can encourage good practice in dealing with knowledge, by which human error as the cause of a catastrophe will at least become less likely.

The European strategy for digitalization in materials modelling, driven within H2020 from the NMBP (nanotechnologies, materials/manufacturing, biotechnology, processing) line, is now going through a critical stage related to the transition from H2020 to Horizon Europe. Platforms developed from H2020 NMBP projects are defining and implementing their concepts for economic viability and long-term preservation [41]. The community is managing the transition away from its first generation of metadata standards (MODA [7], CHADA [42], *etc.*), which were heavy in tabular input and ill-adjusted not only to the semantic web paradigm as such, but also to the openness and open-endedness that characterizes scientific practice at large [43]. This transition leads to the next generation of semantic technology in materials modelling, to a more innovative paradigm. It is already recognizable that the now emerging system of digital platforms will be based on the EMMO and physical topologies [16, 44] to some extent, but exhibit semantic heterogeneity, *e.g.*, through the OntoCommons Ecosystem [5] and, as the DORIC principles suggest [45], rely on a creative combination of diverse technologies instead of one single coherent and monolithic architecture. It will encourage the uptake of recommendations from the RDA and from EOSC, and be subordinate to the Industry 5.0 strategy of the European Commission, *e.g.*, through an Industry Commons Ecosystem (ICE) [5]. The ICE will facilitate the creation of enterprise architectures for "data interoperability" and "cross-domain data exchange" [5]. In the German research landscape, a need for disciplinary research data infrastructures was recognized, and the *national research data infrastructure* (NFDI) programme was launched [46, 47]. The NFDI is organized on a discipline-specific basis to enable a bottom-up approach to development. Metadata4Ing (m4i), which originates from The engineering-sciences consortium NFDI4Ing [48], has developed and released metadata4Ing (m4i), a mid-level ontology [49]; m4i is a process-centric ontology, oriented toward provenance documentation and the description of research processes in the engineering sciences and related disciplines. Accordingly, there are close connections between m4i and PIMS-II, including references and alignments; m4i is also partly aligned with the EMMO, mostly through PIMS-II. Finer differentiation is up to subdisciplines. This is done in collaboration with other NFDI consortia, such as the Mathematical Research Data Initiative (MaRDI) [50], establishing the epistemic foundations, *e.g.*, for applied mathematics in engineering workflows.

The same tension is at work in both cases, that of the European and that of the German programme: The tension between *proper librarianship* (digitization) and *data interoperability* (digitalization). This is not a recent development. From the 1980s, librarians started to consider machine-readable artefacts as entities subject to cataloging, requiring a dedicated treatment [4]. In the US medical sciences, from the 1982 Matheson report [51], it was found that a "library is the most logical site for information management in academic health sciences centers, but it is not yet prepared to assume a leadership role and to take advantage of new technologies; it is still semi-automated and plays a passive role in information transfer" [52]. In 2003, Marshall and Shipman [53] contrasted the "Library of Alexandria" view of the semantic web, standing for an universal, but not interoperable way of storing knowledge, on the one hand against the "Federated Knowledge Base" or "Knowledge Navigator perspective" on the other, being a metaphor for interoperable, machine-readable universal knowledge exchange. Naturally, librarianship primarily aims at cataloging and annotating its library assets. It will also *digitize* content, and very successfully: All academic publications are digitized today, and substantial metadata are made accessible digitally as well. In this way, a human reader can easily find a paper, read it, and understand what its authors found, *e.g.*, a parameterization of a molecular model that yields a certain, improved predictive accuracy when applied to some thermophysical property of some molecular compound. But this information, the main research outcome from the paper, has not been *digitalized*: It is impossible for a digital infrastructure to tell that this and exactly this is the main *knowledge claim* formulated by the authors in their publication, and hence, no data interoperability or cross-domain data exchange is facilitated. Digitalization in the engineering sciences would be served best by supporting the machine-actionable communication of the epistemic metadata and, among these, the *knowledge claims* first and foremost.

Well-known metadata models and ontologies that are widely used in the sciences, such as DataCite [54], PROV [55], METS [56], MODS [57], or PREMIS [58], focus on proper librarianship as their purpose is the description of data provenance and technical properties for catalog-based management of data, *i.e.*, sole data preservation in digital libraries. MODA, CHADA, and m4i have the same bias toward data provenance as opposed to the other categories of epistemic metadata [7, 42, 49]. While data in repositories must of course be well-cataloged, it must also be a goal to include validity claims and epistemic foundations in the data documentation as well. Otherwise, there can never be a comprehensive annotation of the data, directed toward scientific reproducibility and scientific good practice. Today, the "Library of Alexandria" worldview has become a limiting factor in advancing toward the industry 5.0.

5. Conclusion

In knowledge representation, naturally, knowledge must play a central role. All semantic artefacts serve the purpose of representing knowledge, or at least assertions the knowledge status of which needs to be clarified. Not every RDF triple is an epistemic metadata item, however. Epistemic metadata are those that help *establish the knowledge status of other data*. This is still a fairly generic category of metadata, the standardization of which at the metaontological or foundational level might in the future flow together with the development of techniques for making assertions about assertions, notably, RDF-star [59]. Provenance documentation alone is not effective to avoid dark data. Specifically, the amount of data that are dark by form can be reduced by including knowledge claims as annotations. We project that the tendency of proceeding from ontology to epistemology, even at a technical level (*e.g.*, using 4QL [60, 61]), will further increase as disciplinary research data infrastructures are being designed in view of managing huge, heterogeneous bodies of data [62, 63] that must be interpreted as holding a logically inconsistent propositional content [64]. Our present effort was driven from the bottom up, primarily addressing requirements for encoding epistemic metadata from molecular thermodynamics. By articulating the proposed solution through a mid-level ontology, we expect that some of its outcome can be generalized beyond this domain.

Acknowledgments

Co-author B.S. acknowledges funding from the NFDI programme within DFG project no. 460135501 (MaRDI). Discussions within Inprodat and EMMC are acknowledged.

References

[1] P. Humphreys, in: M. Carrier, A. Nordmann (Eds.), Science in the Context of Application, Springer (ISBN 978-90-481-9050-8), Dordrecht, 2011, pp. 131–142. doi:10.1007/978-90-481-9051-5_9.

[2] B. Schembera, J. M. Durán, Philos. Technol. 33 (2020) 93–115. doi:10.1007/s13347-019-00346-x.

[3] S. Samuel, A. Algergawy, B. König-Ries, in: Proc. JOWO 2021, CEUR-WS, 2021, p. 64. URL: http://ceur-ws.org/Vol-2969/paper64-OntoCom.pdf.

[4] S. A. Dodd, Cataloging Machine-Readable Data Files: An Interpretive Manual, ALA (ISBN 978-0-8389-0365-0), Chicago, 1982.

[5] M. Magas, D. Kiritsis, Int. J. Product. Res. 60 (2022) 479–492. doi:10.1080/00207543.2021.1989514.

[6] M. T. Horsch, techn. rep. 2022–A, Inprodat, 2022. doi:10.5281/zenodo.6552271.

[7] CEN-CENELEC Management Centre, Materials modelling: Terminology, classification and metadata, workshop agreement CWA 17284:2018 E, CEN, 2018.

[8] A. F. de Baas (Ed.), What Makes a Material Function? Let Me Compute the Ways, EU Publications (ISBN 978-92-79-63185-6), Luxembourg, 2017. doi:10.2777/417118.

[9] M. T. Horsch, in: Proc. JOWO 2021, CEUR-WS, 2021, p. 3. URL: http://ceur-ws.org/Vol-2969/paper3-FOUST.pdf.

[10] B. Schembera, M. T. Horsch, 2022. Submitted.

[11] M. R. Elkobaisi, F. Al Machot, H. C. Mayr, SN Comp. Sci. 3 (2022) 282. doi:10.1007/s42979-022-01116-x.

[12] A. Janssen, C. Klein, M. Slors, Philos. Explorat. 20 (2017) 123–128. doi:10.1080/13869795.2017.1312496.

[13] P. Atã, J. Queiroz, Tr/Fo/Aç 44 (2021) 177–204. doi:10.1590/0101-3173.2021.v44n3.15.p177.

[14] J. F. Sowa, in: B. Ganter, G. W. Mineau (Eds.), Conceptual Structures: Logical, Linguistic, and Computational Issues, Springer (ISBN 978-3-540-67859-5), Heidelberg, 2000, pp. 55–81. doi:10.1007/10722280_5.

[15] J. F. Sowa, Appl. Ontol. 10 (2015) 273–284. doi:10.3233/AO-150159.

[16] H. A. Preisig, T. F. Hagelien, J. Friis, P. Klein, N. Konchakova, in: Proc. WCCM-ECCOMAS 2020, Scipedia, 2021, p. 262. doi:10.23967/wccm-eccomas.2020.262.

[17] S. Clark, F. L. Bleken, S. Stier, E. Flores, et al., Adv. Energy Mater. 12 (2022) 2102702. doi:10.1002/aenm.202102702.

[18] M. T. Horsch, S. Chiacchiera, B. Schembera, M. A. Seaton, I. T. Todorov, in: Proc. WCCM-ECCOMAS 2020, Scipedia, 2021, p. 297. doi:10.23967/wccm-eccomas.2020.297.

[19] O. Corcho, M. Eriksson, K. Kurowski, M. Ojsteršek, et al., techn. rep., EOSC, 2021. doi:10.2777/620649.

[20] J. Hoopes (Ed.), Peirce on Signs, Univ. North Carolina Press (ISBN 978-0-80784342-0), Chapel Hill, 1991.

[21] C. F. Meurer, N. M. de Figueiredo, Tr/Fo/Aç 44 (2021) 205–210. doi:10.1590/0101-3173.2021.v44n3.16.p205.

[22] M. T. Horsch, S. Chiacchiera, M. A. Seaton, I. T. Todorov, et al., Künstl. Intell. 34 (2020) 423–428. doi:10.1007/s13218-020-00648-9.

[23] M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, B. Schembera, Data Technology in Materials Modelling, Springer (ISBN 978-3-03068596-6), Cham, 2021. doi:10.1007/978-3-030-68597-3.
[24] M. T. Horsch, B. Schembera, Preliminary case study (10 cases), 2022. In preparation.
[25] H. E. Plesser, Frontiers Neuroinform. 11 (2018) 76. doi:10.3389/fninf.2017.00076.
[26] M. T. Horsch, in: I. Huvila, et al. (Eds.), Perspectives on Paradata, 2023. In preparation.
[27] C. S. Peirce, in: J. Buchler (Ed.), Philosophical Writings of Peirce, Dover (ISBN 978-0-48620217-4), New York, 1955, pp. 98–119.
[28] D. Marr 9 (1977) 37–48. doi:10.1016/0004-3702(77)90013-3.
[29] K. Marx, MEW 23, 27, 1873.
[30] J. M. Durán, N. Formanek, Minds Machin. 28 (2018) 645–666. doi:10.1007/s11023-018-9481-6.
[31] J. M. Durán, K. R. Jongsma, J. Medic. Ethics 47 (2021) 329–335. doi:10.1136/medethics-2020-106820.
[32] M. T. Horsch, techn. rep. 2021–B, Inprodat, 2021. doi:10.5281/zenodo.4679522.
[33] http://www.molmod.info/semantics/pims-ii-examples/benchmark/case-4.ttl, ????
[34] S. Borgo, A. Galton, O. Kutz, Appl. Ontol. 17 (2022) 1–16. doi:10.3233/AO-220265.
[35] E. M. Sanfilippo, Appl. Ontol. 16 (2021) 111–135. doi:10.3233/AO-210246.
[36] P. B. Heidorn, Libr. Trends 57 (2008) 280–299. doi:10.1353/lib.0.0036.
[37] D. J. Hand, Dark Data: Why What You Don't Know Matters, Princeton Univ. Press (ISBN 978-0-6911-8237-7), Princeton, 2020.
[38] B. Schembera, J. Supercomput. 77 (2021) 8946–8966. doi:10.1007/s11227-020-03602-6.
[39] J. Lenhard, Philos. Sci. 85 (2018) 832–844. doi:10.1086/699675.
[40] B. De Vivo, E. Laor, G. Roland, in: G. F. Panza, et al. (Eds.), Earthquakes and Sustainable Infrastructure, Elsevier (ISBN 978-0-12-823503-4), Amsterdam, 2022, pp. 297–324. doi:10.1016/B978-0-12-823503-4.00001-4.
[41] G. Goldbeck, A. Simperler, G. Mogni, white paper, MarketPlace and VIMMP, 2022. doi:10.5281/zenodo.6329759.
[42] N. Romanos, M. Kalogerini, E. P. Koumoulos, K. Morozinis, et al., Mater. Today Comm. 20 (2019) 100541. doi:10.1016/j.mtcomm.2019.100541.
[43] K. Tulatz, Epistemologie als Reflexion wissenschaftlicher Praxen, transcript (ISBN 978-3-8376-4212-4), Bielefeld, 2018.
[44] H. A. Preisig, Processes 9 (2021) 592. doi:10.3390/pr9040592.
[45] M. T. Horsch, J. Francisco Morgado, G. Goldbeck, D. Iglezakis, et al., in: Proc. DORIC-MM 2021, UKRI STFC, 2021, pp. 12–27.
[46] S. Kraft, A. Schmale, H. Seitz-Moskaliuk, Y. Sure-Vetter, et al., Baust. FDM 2021 (2021) 1–9. doi:10.17192/bfdm.2021.2.8332.
[47] S. Kraft, H. Seitz-Moskaliuk, Y. Sure-Vetter, E. Wössner, et al., in: Proc. E-Science-Tage 2021, heiBOOKS (ISBN 978-3-948083-55-7), Heidelberg, 2022, pp. 374–383. doi:10.11588/heibooks.979.c13752.
[48] T. Bronger, E. Demandt, I. Heine, A. Kraft, et al., Baust. FDM 2021 (2021) 110–123. doi:10.17192/bfdm.2021.2.8329.
[49] A. Karmacharya, B. Farnbacher, C. Wiljes, D. Iglezakis, et al., Metadata4Ing documentation, NFDI4Ing, 2022.

[50] C. Görgen, R. Sinn, Mitteil. DMV 29 (2021) 122–123. doi:10.1515/dmvm-2021-0049.
[51] N. W. Matheson, J. A. Cooper, Acad. Med. 57 (1982) 1–93.
[52] H. S. Buchanan, N. Fazzone, Bull. MLA 73 (1985) 47–54.
[53] C. C. Marshall, F. M. Shipman, in: Proc. HT '03, ACM, 2003, pp. 57–66. doi:10.1145/900051.900063.
[54] J. Neumann, J. Brase, J. Comp.-Aid. Mol. Des. 28 (2014) 1035–1041. doi:10.1007/s10822-014-9776-5.
[55] L. Moreau, P. Groth, J. Cheney, T. Lebo, S. Moles, J. Web Semant. 35 (2015) 235–257. doi:10.1016/j.websem.2015.04.001.
[56] C. Chapman, S. Parker, S. Parsons, W. B. Seales, in: Proc. MTSR 2020, Springer (ISBN 978-3-030-71902-9), 2021, pp. 143–154. doi:10.1007/978-3-030-71903-6_15.
[57] J. Park, S. Maszaros, Knowl. Org. 36 (2009) 46–59. doi:10.5771/0943-7444-2009-1-46.
[58] P. Caplan, PREMIS documentation, Network Development and MARC Standards Office, 2017.
[59] T. Delva, J. Arenas Guerrero, A. Iglesias Molina, O. Corcho, et al., in: Proc. ISWC 2021 Posters, Demos, Industry, CEUR-WS, 2021, p. 374. URL: http://ceur-ws.org/Vol-2980/paper374.pdf.
[60] Ł. Białek, B. Dunin-Keplicz, A. Szałas, in: Proc. IDC 2017, Springer (ISBN 978-3-319-66378-4), 2018, pp. 49–60. doi:10.1007/978-3-319-66379-1_5.
[61] B. Dunin-Keplicz, A. Szałas, Procedia Comp. Sci. 176 (2020) 3019–3028. doi:10.1016/j.procs.2020.09.201.
[62] G. V. Belov, A. O. Erkimbaev, V. Y. Zitserman, G. A. Kobzev, I. V. Morozov, High Temp. 58 (2020) 566–582. doi:10.1134/S0018151X2004001X.
[63] S. Schimmler, T. Bönisch, M. T. Horsch, T. Petrenko, et al., in: Proc. E-Science-Tage 2021, heiBOOKS (ISBN 978-3-948083-55-7), Heidelberg, 2022, pp. 277–284. doi:10.11588/heibooks.979.c13738.
[64] K. Mu, in: Proc. KSEM 2021, Springer (ISBN 978-3-030-82146-3), 2021, pp. 319–321. doi:10.1007/978-3-030-82147-0_26.