Exploration of core concepts required for mid- and domain-level ontology development to facilitate explainable-AI-readiness of data and models

Martin T. Horsch[1,2,*], Silvia Chiacchiera[2], Ilian T. Todorov[2], Ana Teresa Correia[3], Aditya Dey[1], Natalia A. Konchakova[4], Sebastian Scholze[3], Simon Stephan[5], Kristin Tøndel[1], Arkopaul Sarkar[6], M. Hedi Karray[6], Fadi Al Machot[1,*] and Björn Schembera[7]

[1]Norwegian University of Life Sciences, Department of Data Science, Postboks 5003, 1432 Ås, Norway
[2]UKRI Science and Technology Facilities Council, Scientific Computing Department, Daresbury WA4 4AD, UK
[3]ATB Institut für Angewandte Systemtechnik Bremen GmbH, Wiener Str. 1, 28359 Bremen, Germany
[4]Helmholtz-Zentrum Hereon, Institute of Surface Science, Max-Planck-Str. 1, 21502 Geesthacht, Germany
[5]RPTU Kaiserslautern, Laboratory of Engineering Thermodynamics, 67663 Kaiserslautern, Germany
[6]University of Technology Tarbes Occitanie Pyrénées, Production Engineering Lab, 47, av. d'Azereix, Tarbes, France
[7]University of Stuttgart, Institute of Applied Analysis and Numerical Simulation, 70569 Stuttgart, Germany

Abstract

This position paper reports on the initial discussions within the Knowledge Graph Alliance's working group on explainable-AI-ready data and metadata principles, which was created in March 2024. At present, we are taking initial steps toward capturing *core concepts* related to explanation, grounding, reliance, and trust; the scope also extends to potential dual notions such as explainability, verifiability/reproducibility, reliability, and trustworthiness. These initial steps consist in reviewing core concepts as they are discussed in the literature and exploring what could be practically useful definitions of these most central concepts. One of the conclusions is that the metadata standards will need to be suitable for documenting three kinds of grounding: Grounding of knowledge, grounding of reliance, and grounding of trust. Pre-existing metadata standards at the mid and domain level are presently undergoing a redesign in order to become more modular, computationally tractable, intelligible to humans, and adjustable, which will be needed as we continue our work toward actionable recommendations. The development of this system of *lite* (OWL 2 EL) ontologies, called *MSO-EM: Ontologies for modelling, simulation, optimization (MSO) and epistemic metadata (EM)*, is carried out on a public repository.

Keywords

Applied ontology, epistemic metadata, explainable-AI-ready (XAIR), ontology redesign, reliability, reproducibility

1. Introduction

This work is part of ongoing discussions within the Knowledge Graph Alliance's working group *Explainable-AI-ready data and metadata principles* (XAIR principles).[1] This working group will develop recommendations (principles) that, if followed, contribute to making data and models explainable-AI-ready (XAIR). This will be supported by metadata standards, specifically, for epistemic metadata, *i.e.*, metadata related to the knowledge status of data and models. To our understanding, data and models are XAIR to the degree that they are semantically enriched in such a way as to facilitate making best use of explainable learning techniques, broadly understood. That is, first, this does include XAI as it is commonly understood in a narrow sense, such as techniques based on Shapley values or local

DAO-XAI 2024: 4th International Workshop on Data meets Applied Ontologies in Explainable AI, October 19–20, 2024, Santiago de Compostela, Spain
*Corresponding authors: Martin Thomas Horsch and Fadi Al Machot.
martin.thomas.horsch@nmbu.no (M. T. Horsch); fadi.al.machot@nmbu.no (F. Al Machot)
0000-0002-9464-6739 (M. T. Horsch); 0000-0003-0422-7870 (S. Chiacchiera); 0000-0001-7275-1784 (I. T. Todorov); 0000-0002-2469-6546 (A. T. Correia); 0000-0002-3093-7596 (N. A. Konchakova); 0000-0002-1193-770X (S. Scholze); 0000-0002-4578-3569 (S. Stephan); 0000-0002-8967-7813 (A. Sarkar); 0000-0002-9652-5164 (M. H. Karray); 0000-0002-1239-9261 (F. Al Machot); 0000-0003-2860-6621 (B. Schembera)

[1]https://www.kg-alliance.org/kga-wg-xai-24-4/

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

interpretable model-agnostic explanations [1]. But it also extends to deductive reasoning, since a logical or mathematical proof can be considered an explanation of that which is being proven; there, our group is particularly interested in providing a XAIR documentation to answer set programming as well as reasoning applied to (fragments of) OWL description logic. Third, our notion of explainable learning also includes simulation based on models that have some structural agreement with the phenomenon that is being modelled; there, most saliently, physics-based modelling such as molecular simulation. Eventually, it is the aim of the working group to advance good practices in data and metadata documentation that help feeding models and data obtained in all sorts of ways into frameworks where they become reusable to all these modalities of explainable learning. This means that not only the output from an explainable learning technique has to become XAIR, but also the input – such as the datasets used for training, validation, and testing in machine learning, or experimental data that are used to parameterize and validate a molecular model. It is as of now not yet characterized what exact prerequisites need to be met to achieve «XAIR status»; instead, this is supposed to be a result from the working group's community consultations. We expect it to be centered on model and data documentation through appropriate metadata, just as it is the case for the FAIR principles [2]. Accordingly, the recommendations for explainable-AI-readiness will build on FAIR, and then go beyond it.

It is a notorious issue in XAI that core concepts such as explainability, interpretability, *etc.*, are most often used without any clearly communicated conceptualization of what they would entail [3]; but in a technical context, clear definitions of technical terms are a necessity, all the more for a task like ours: Supporting data documentation for XAI-readiness. To our work, central questions at this stage include:

1. What properties do objects o and p need to have as a prerequisite for o relying on p?
2. How about o trusting p?
3. What characterizes the difference between trust and reliance?
4. What characterizes the difference between explanations and grounding?
5. What semantic artefacts need to be put in place so that the above can be documented?

The XAIR principles WG has been formed in March 2024, and its work plan for the time being extends over 40 months, until June 2027. At the present stage, we are engaged with meta-reviewing what perspectives from the literature need to be taken into account, regarding questions such as those mentioned above, but also to identify the core concepts to XAI-readiness and explore what definitions they have been given. The outcome of this exploration will feed into discussions on actionable definitions of the key concepts that the WG itself will accept; these will then be used for a revision of the previously developed ontologization of cognitive processes and epistemic metadata [4]. First, however, that pre-existing ontology, PIMS-II, needs to be redesigned into a simpler framework to become more amenable to the planned work. It will for that purpose be split into multiple *lite* mid- and domain-level OWL 2 EL ontologies, improving tractability of reasoning tasks as well as intelligibility to developers. Eventually, as an output from the group's work, reference ontologies will be provided to users who want to use them to document models and data in a XAIR way. Nonetheless, «XAIR status» is to be ontology-agnostic, and users will in no way be required to use these ontologies or give preference to them over others that support achieving compliance with the recommendations for explainable-AI-readiness equally well.

2. Literature on the core concepts for explainable-AI-readiness

A list of *XAIR core concepts* has been issued as part of a recent announcement [5]. It comprises: «Explainability and explanation; reproducibility, reliability, and reliance; opacity and transparency, interpretability and interpretation; DIKW: Data, information, knowledge, and wisdom; responsibility, trust, trustworthiness, and [. . .] motivations for trusting; model design, parameterization, and optimization; holistic validation and unit testing (of models and simulation codes); theoretical virtues (of models); epistemic agents, vices, and virtues; the four elements of FAIR [. . .]; simulation; applying and evaluating models; context awareness, subject matter, and logical subtraction» [5]. At present, we are conducting a preliminary literature analysis on a subset of the above. We call this a meta-review, since the resources that we are evaluating are themselves already reviews (or other aggregated material). The aim is not to obtain a complete understanding of how all these matters have been discussed in the literature from all possible angles, but to sample a variety of perspectives as an initial guidance to discussions about recommendations for good practice in explainable-AI-ready data and model documentation.

2.1. Wisdom hierarchy as reviewed by Rowley

A hierarchy according to which data (D) can be analysed/improved qualitatively first to information (I), then to knowledge (K), and finally to wisdom (W) has a long tradition in data management and is often visualized as a *DIKW pyramid*. This «hierarchy» is not a taxonomy, *i.e.*, wisdom is not understood to be a subclass of data, or vice versa; instead, D $\rightarrow$ I $\rightarrow$ K $\rightarrow$ W is a kind of value chain or dialectical sequence. For the present exploration, the review by Rowley [6] is our starting point. The review is from 2007, but measured by the time scale of progress within epistemology, that means it is recent.

2.2. Agency as reviewed by Conte

Within the present framework, we want to keep the concept of agency at a very general level, so that *(almost) anything can be analysed as an agent*, similar to the use that is made of the concept of a system in thermodynamics. Here, if a system has something that can be understood as sensors and actuators or is by means of some other mechanism capable of *perceiving* and *doing*, it can be analysed as an agent. This makes it all the more important to include a taxonomy of agents as part of an associated mid-level ontology, since more specific subclasses will be needed to talk about kinds of agents. Our point of departure in this field is given by the review and the definitions proposed by Conte [7].

2.3. Reproducibility as reviewed by Plesser

Reproducibility is a key component of industry standards, as seen in Quality Management Systems like ISO 9001, which emphasize reproducibility through consistent quality management practices, requiring detailed documentation and standardized procedures to ensure reliable process repetition [8]. Journals are now encouraging or requiring the submission of data and code alongside published articles to enable the replication of results, while platforms like the Open Science Framework promote open access to research materials, enhancing reproducibility by providing researchers with the tools to share their data, code, and protocols openly [9]. However, the drive for novel results obtained from complicated methods often leads to practices that hinder reproducibility [10]. A balance is needed to ensure that research outputs are both innovative and reproducible [11]. Plesser's concise review [12] is the starting point for our exploration of frameworks suitable for documenting reproducibility claims and the validation of such claims; see also Section 3.3, as well as previous work where this was discussed more in depth [13].

2.4. Trust as reviewed by Ohlhorst and the ONTrust reference ontology

Ohlhorst [14], beside making his own philosophical argument on the nature of trust, provides an review on concepts that are central to his theory, in particular, *hinge*, *entitlement*, and *virtue*. Ohlhorst's work and the state of the art summarized therein need to be taken into account for our working group's discussions on the relation between explanations and trust in AI systems (Section 3). Attempts at ontologization can build on the ONTrust reference ontology by Baratella *et al.* [15].

2.5. Reviews from the OntoCommons and CHEIKHMAT projects

The *Review of Domain Interoperability* [16] (RoDI) is a deliverable from the OntoCommons project (H2020 GA no. 958371) that preceded the formation of the Knowledge Graph Alliance. The purpose of the RoDI document is to provide a foundation to discussions and technical innovation addressing domain-level interoperability now that OntoCommons is completed and in its place, the KGA has been created as a self-sustaining organization. As such, RoDI both reviews and provides recommendations for designing semantic artefacts. These fit into the OntoCommons ecosystem [17] (OCES) that permits a plurality of foundational ontologies (DOLCE [18], EMMO [19], BFO [20]), and strategies for semantic heterogeneity and alignment, such as *bridge concepts*. This approach forms part of the background to the ongoing ontology redesign (*cf.* Section 4) within the KGA's working group on XAIR principles.

In addition, the *interoperability requirements matrix* from RoDI [16, Section 5.1] develops an understanding of key concepts that can be related to the wisdom hierarchy, at the level of successfully communicating knowledge that, in the DIKW pyramid, is referred to as wisdom [6]. There, RoDI distinguishes interoperation of data, humans, software, and organizations, and three interoperability levels: Syntactic, terminological/semantic, and pragmatic [16]. In addition to this output from OntoCommons, the CHEIKHMAT project has recently produced a review that very closely connects to our ongoing survey of key concepts, with a focus specifically on ontology-based and semantics-based XAI [21].

The list of sources and topics included in this first effort is far from sufficient. Given the task of the working group, it urgently has to be expanded by catalogues or reviews of explainability metrics [1], since this is a key part of what will need to be documented in practice. As an output by month 10 of the group's work plan (December 2024), a *Synopsis of XAIR Core Concepts* will be delivered, which will take the form of proceedings from a workshop organized by the projects AI4Work and BatCAT [5].

3. Exploration of actionable key concepts

3.1. Reliance and reliability

It does not make sense to speak of an agent A as relying upon something unless it is done in view of some objective τ. In other words, there must be an adverse consequence for the agent if the relationship fails; namely, that an expected outcome is not achieved. We do not for now require goal-directedness, *i.e.*, A is not required to have an awareness or internal representation of τ as the agent's intention; but it is necessary for A to be goal-oriented, *i.e.*, have an objective (conscious or not). The entity acting as the subject of a reliance relationship (that which does the relying) must therefore be a goal-oriented agent.

The reliance relation can be formalized as:

- The ternary predicate «A relies on B about q», where B is an interlocutor, *i.e.*, another agent (this can also *e.g.* be a queriable software running on a machine), and q is the subject matter that the relationship is about. This realizes itself by A relying upon the following rule: «The antecedent "agent B asserts c, where c is about subject matter q" necessarily implies the consequent c.»
- A more general ternary predicate «A relies upon rule r for τ», where $r : \Box_\ell(a \to c)$ can be any rule according to which antecedent a necessarily (to the degree ℓ) implies consequent c. This is to be interpreted loosely, *e.g.*, a and c will usually share free variables, with universal quantification applied to these variables. For the reliance relationship to work in practice, it is necessary for A to also rely upon a: At some point in time, A may know (or not know, but act as if) a holds, so the rule can be applied. Otherwise, reliance or non-reliance upon r remain indistinguishable.
  The index ℓ of the necessity operator represents the level of necessity at which the rule r is understood to apply; *e.g.*, in some cases, exceptions may be tolerated, and in some cases, there may be a margin within which quantitative deviations can be accepted. Descriptors for this can become complicated, see the discussion of *what-you-see-is-what-you-get guarantees* in Grote *et al.* [22]. For the predicate «relies on», the level $\ell(q)$ is understood to be contained in the subject matter q. Here, the notion of aboutness or subject matter can be taken to be that from Yablo [23].
- The even more general ternary predicate «A relies upon φ for τ», where φ can be any proposition, not restricted to a rule. The reliance relationship only makes sense if φ somehow causally contributes to τ: It must deliver some contribution to the agent A reaching the goal τ.

Rule $r : \Box_\ell(a \to c)$ is reliable whenever it is true, *i.e.*, whenever a implies consequent c to the level ℓ.

In action, the relationship of reliance is not constituted by the agent A believing or knowing that the rule holds; *i.e.*, that under all the described conditions, c is actually true. That may or may not be the case. Instead, the agent A practices reliance upon a rule r by *behaving as if A knew that r was reliable*, or in terms that will be clarified further below (Section 3.2), in A emulating an agent A' such that r is grounded to A. The emulated agent A' is identical to A in all respects other than *about* acceptance of r as knowledge. The relation of relying on someone can accordingly be expanded as

$$
\begin{aligned}
&A \text{ relies on } B \text{ about } q \\
&\quad\equiv\quad A \text{ relies upon rule } r(B,q):\ \Box_{\ell(q)}\,\forall c\,((c \text{ is about } q\ \wedge\ B \text{ asserts } c) \to c) \qquad (1)\\
&\qquad\equiv\quad \text{about } q(r), A \text{ emulates } A', \text{who is like } A, \text{except that } A' \text{ knows } r(B,q). \qquad (2)
\end{aligned}
$$

Therein, $q(r)$ refers to the scope of reliance upon r. The above does not exclude the case where A really knows r to be true. In that case, it simply reduces to $A = A'$; *i.e.*, grounded reliance does not require A to emulate another hypothetical agent who accepts r as knowledge, since A already is that agent.

Based on the above, we can characterize agents' reliability: Interlocutor B is reliable about q whenever the rule $r(B,q)$ from the right-hand side of Proposition (1) is reliable, *i.e.*, whenever it is true.

3.2. Grounding and explanations

A proposition/rule and the associated reliance relationship is grounded to some agent C if that agent knows it to be true. The most relevant case of this is applying this definition with $A = C$, *i.e.*, asking whether the agent who does the relying has sufficient knowledge to ground that reliance relationship (if that is not the case, we can infer that the agent *trusts* in the relationship, instead of knowing). However, we can also insert ourselves as C; in this case, the question becomes whether we know the rule that is relied upon to be true, *i.e.*, whether A's reliance relationship is grounded to us (only then, we would say that the rule is reliable). Following this approach, *reliance is grounded in knowledge.* Accordingly, first, some information would need to be established as having the status of knowledge – grounding of reliance presupposes grounding of knowledge. The question of what exactly can ground knowledge, and how that is to be documented, is at the core of the problem of standardizing epistemic metadata [4].

Reliance does not require any grounding, *i.e.*, we can rely on someone or upon something without *knowing* that this will be successful and lead to the intended outcome. In that case, we say that we *trust.*

Any relationship of trust presupposes that there is a relationship of reliance that is ungrounded to the trustor A, *i.e.*, the relying and trusting agent: Where knowledge would be needed to ground the reliance relationship, that knowledge is absent. This can be either because the rule that the agent relies upon is in fact unreliable, or because while it is in fact reliable, that fact is not known to A. With Proposition (2) we can summarize that according to our tentative formalization proposed here, if A trusts in (and, therefore, does not know) some φ, that is realized by A emulating an agent A' who is like A, except that A' knows φ (and, therefore, does not trust in φ). This emulation is limited to a scope $q(\varphi)$, namely, the scenarios that the proposition is intended to be applied to, *i.e.*, those that it is about [23].

Similar to reliance, trust does not require any grounding: An agent *can* trust blindly in someone or something without any good reason at all. However, trust *can* also be grounded [24], *i.e.*, an entitlement (epistemic warrant) can be provided in many different ways [14], all of which can be called *explanations.* Societal norms and community practices can impose limits on what behaviour around trust is acceptable: The agent can be held responsible by the community to trust only in appropriate agents or propositions. In science, naturally, blind trust is not tolerated, while as Koskinen [25] points out, *some* trust is necessary. The space between the necessary and the forbidden trust in data-driven modelling would need to be characterized by a social epistemology that, following Koskinen, still needs to be developed [25].

For guidelines that advance XAIR models and data, it is therefore a requirement to develop metadata standards (*cf.* Section 4) for documenting three kinds of grounding: First, epistemic grounding, or *grounding of knowledge*, in line with practices that are accepted as scientific within a scientific community; second, *grounding of reliance* by knowledge; and third, *grounding of trust* by explanations.

3.3. Reproducibility and transparency

Reproducibility ensures the capacity to consistently replicate results under identical conditions, providing reliability and verification, *cf.* Section 2.3. Key elements include documentation, standardization, and data availability; where these are absent, we speak of dark data [26, 27, 28].

Good practices in research data management avoid dark data and facilitate reproducibility, *e.g.*, by complying with suitable documentation and transparency requirements. Transparency involves providing clear, accessible, and understandable information about processes, enabling scrutiny and understanding. The key components include clarity, accessibility, accountability, and traceability.

We remark that the notion of an «implication necessary to level ℓ» proposed in Section 3.1 for rules upon which agents rely is similar to, and can probably be unified with, the «conditional necessity of φ, given κ» proposed in previous work [13] to formalize the semantics of a reproducibility claim.

4. Mid- and domain-level ontology redesign

Once the discussions on the core concepts are further progressed, the synopsis has been copiled, and own definitions of core concepts are sufficiently mature, our working group will provide recommended metadata standards for annotating models and data in a way that improves their XAI-readiness. There is a major element to this from previous work that will require a redesign: The PIMS-II ontology [4] was developed for cognitive processes (with a focus on research workflows as cognitive processes) and epistemic metadata (knowledge claims, provenance, and validity claims). This is, however, too broad for a single ontology to be easily accessible to users; it is even less suitable for a focused discussion of key concepts, since it will be hard to adjust such a monolithic conceptual scheme to novel perspectives that might be agreed upon as an outcome from group discussions. In addition, PIMS-II has developed substantial philosophical overhead (with a FOL axiomatization of its core [29] that distracts from its purpose and is little accessible to its potential user base), created from an attempt [30] to connect to the metaphysics underlying the EMMO foundational ontology [19]. However, with the development of OCES [17, 31], it is no longer needed or even advisable for mid- and domain-level ontologies to attach themselves to the EMMO so closely, just to interoperate with other EMMO-related ontologies. OCES and *bridge concepts* [31] can just as well be used for this purpose [16, 17]. Another element that might be reused, contingent upon redesign, are (some of) the VIMMP ontologies [32]; the most relevant ontologies from VIMMP would be OSMO [33, 34], the ontology version of CWA 17284 MODA [35], and MMTO [36], the ontology version of the EMMC CSA Translation Case Template [37]. The VIMMP ontologies, specially OSMO, suffer from the same problem as well: Substantial work went into them, and now, a single ontology addresses too many issues at once, reducing its usefulness for targeted development tasks; beside, like PIMS-II, the VIMMP ontologies too became complicated through attempting to conform with preliminary development versions of the EMMO and its metaontological overhead (and, in the case of OSMO, additionally with MODA). Both PIMS-II and the VIMMP ontologies also make heavy use of OWL 2 Full semantics and do not prioritize the tractability of reasoning tasks.

Generally speaking, mid- and domain-level ontology redesign should focus on refining and reorganizing ontological structures to significantly improve clarity, usability, and alignment with domain-specific requirements, ensuring the ontology accurately represents the required knowledge and relationships within a domain. This process facilitates better interoperability, data integration, and reasoning capabilities. Essential considerations in this redesign include achieving conceptual clarity by defining all concepts and relationships without ambiguity, removing redundancies, and resolving inconsistencies to enhance the ontology's functionality and usability [38]. Organizing concepts into a clear hierarchical structure is critical; this involves not only identifying and establishing natural or logical parent-child relationships but also ensuring that subcategories are appropriately nested to reflect the true complexity of the domain [39]. The design should be modular, allowing for easy updates and extensions, which supports ongoing improvements and adaptability. The ontologies should also be aligned with a foundational ontology and widely used, or otherwise important, metadata standards from the domain [38]; in our case, the latter include the reference ontology of trust [15] (*cf.* Section 2.4), the German NFDI4Ing project's Metadata4Ing [40] (m4i), and further domain ontologies as a function of the addressed use case. Within BatCAT, *e.g.*, EMMO-related domain ontologies such as the BattINFO ontology [41] and the battery value chain ontology (BVCO). In some domains, BFO-related ontologies would predominate [38]. Here, DOLCE [18] is used as *the* foundational ontology with which a *strong alignment* is made, using DOLCE Lite [42]. DOLCE has been stable for over twenty years and can therefore be relied on as a building block that will not change in the foreseeable future. In relation to other work, we can rely on weak alignment based on the SKOS relation skos:closeMatch and additionally benefit from OCES for connecting to any ontology development that has been done under the BFO and the EMMO [16, 17].

The redesigned mid- and domain-level ontologies, called *MSO-EM: Ontologies for modelling, simulation, optimization (MSO) and epistemic metadata (EM)*, are/will be licensed under CC BY-ND 4.0, and their development is openly accessible through a public github repository.[2] The design principles include:

1. The expressivity is restricted to OWL 2 EL, corresponding to $\mathcal{EL}$++ description logic.
2. Each ontology from MSO-EM has at most three taxonomy levels and three concepts at its highest level. (This does not mean that there will only be three hierarchy levels within MSO-EM as a whole, but that the tree depth for the concepts defined in a single module TTL file is limited to three.) Only include concepts that we expect to be directly required in technical practice. Remove unnecessary concepts in the middle region of the tree, if it is the leaves of the tree that developers would rather refer to instead.
3. Define few relations, prefer abstract (high-level) relations where possible, and be specific in terms of the concepts rather than the relations. There is no inherent logical reason to prefer this, but it is in line with human communication – we tend to use few, comparably generic verbs most of the time. On the other hand, the nouns we use in speech or writing easily become very specific.
4. Define relations one-way only; where applicable, use the direction that can be used to define more meaningful restrictions on concepts. Normally, this means that the greater (or whole) entity should be the subject and the smaller (or part) entity should be the object, *i.e.*, relations of the type «has part» are preferred over those of the type «is part». This is because OWL EL does not allow owl:inverseOf, and the greater-to-smaller direction is more frequently used in restrictions.
5. Use strong alignment (rdfs:subClassOf and rdfs:subPropertyOf) only in reference to DOLCE Lite, not in reference to other ontologies outside the MSO-EM system itself. All concepts from MSO-EM are subsumed under a concept from DOLCE Lite. Where it can be done, use skos:exactMatch for strong alignment with DOLCE Lite, m4i, and PIMS-II, but not with any other ontologies.
6. Use weak alignment (skos:closeMatch) where possible for D-SI, EMMO, GPO [43], schema.org, and any of the legacy ontologies other than PIMS-II; skos:broader and skos:narrower are not used.

At submission time, one of these ontology modules is ready as a demonstrator:

The MSO-EM agency ontology (agency.ttl in the repository[2]), which implements Conte's taxonomy [7] and relations and concepts connected to it; compared to the Agent branch of PIMS-II,[3] the taxonomy was reduced to three hierarchy levels, eliminating *e.g.* the concept IntelligentAgent for which, there is little use in concrete practice, since almost always, more specific subclasses would be used instead.

Due to the restrictions imposed on each single ontology, the number of ontologies in this system will be comparably large, but each of those is expected to be easy to understand. Simplicity in terms of interconnections and a comparably flat taxonomy will make it possible to adjust definitions to the outcome of community and group discussion processes; at least, this will be less problematic than in case of the large monolithic ontologies from previous work such as OSMO and PIMS-II.

[2]Persistent URL for the MSO-EM ontologies: https://www.purl.org/mso-em. Open-access repository for development: https://github.com/HE-BatCAT/mso-em. URL with which the system of ontologies can be loaded directly into protégé: https://batcat.info/semantics/mso-em/ (in protégé: «File» → «Open from URL» → input the URL to the left).

[3]http://molmod.info/semantics/pims-ii.ttl

5. Conclusion

While this position paper reflects the overall perspective from discussions within the KGA working group on XAIR principles, much of the uptake and further development will depend on the actual use made of these developments for semantic interoperability, data documentation, and the integration of models and data into AI-driven systems. The research projects involved in the present effort, so far, include AI4Work, BatCAT [44], DigiPass CSA, and MaRDI [45].[4] Within BatCAT, based on the project's requirements analysis [44], plans include to give ontology design principles such as those from Section 4 the status of a *meta-metamodel* (M3-model), at the top of a hierarchy in compliance with the Object Management Group's ISO standard Meta Object Facility [46] (MOF). Accordingly, the ontologies themselves become the *metamodel* (M2-model), below which knowledge graph shapes, specified through OO-LD,[5] are the *model* (M1-model). Using MOF helps establish a coherent system of semantic and technical interoperability, capable of supporting advanced data operations through integration with other OMG standards such as BPMN, which is also used in the project.

Acknowledgments

The work was done within the Knowledge Graph Alliance's working group on XAIR principles in collaboration with AI4Work, BatCAT, DigiPass CSA, and MaRDI. The co-authors A. T. Co. and S. Sc. acknowledge funding from the EU's Horizon Europe research and innovation programme under GA no. 101135990: AI4Work. The co-authors F. Al Ma., S. Ch., M. T. Ho., S. St., I. T. To., and K. Tø. acknowledge funding from Horizon Europe under GA no. 101137725: BatCAT, and the co-authors F. Al Ma., M. T. Ho., and N. A. Ko. under GA no. 101138510: DigiPass CSA. The co-author B. Sc. acknowledges funding from MaRDI, DFG project no. 460135501, within the German national research data infrastructure (NFDI). The authors acknowledge discussions with Colin W. Glass and Andreas Neumann.

References

[1] M. Pawlicki, A. Pawlicka, F. Uccello, S. Szelest, S. D'Antonio, R. Kozik, M. Choraś, Evaluating the necessity of the multiple metrics for assessing explainable AI: A critical examination, Neurocomputing 602 (2024) 128282. doi:10.1016/j.neucom.2024.128282.

[2] M. D. Wilkinson, M. Dumontier, IJsbrand Jan Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, et al., The FAIR guiding principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. doi:10.1038/sdata.2016.18.

[3] Z. C. Lipton, The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery, Queue 16 (2018) 31–57. doi:10.1145/3236386.3241340.

[4] M. T. Horsch, B. Schembera, Documentation of epistemic metadata by a mid-level ontology of cognitive processes, in: Proc. JOWO 2022, CEUR-WS (CEUR Workshop Proceedings no. 3249), 2022, p. CAOS: 2.

[5] F. Al Machot, L. E. Córdova López, A. T. Correia, M. T. Horsch, S. Scholze, S. Stephan, S. Werth, DCLXVI 2024, call for papers, AI4Work and BatCAT, 2024. doi:10.5281/zenodo.13142343.

[6] J. Rowley, The wisdom hierarchy: Representations of the DIKW hierarchy, Journal of Information Systems 33 (2007) 163–180. doi:10.1177/0165551506070706.

[7] R. Conte, Rational, goal-oriented agents, in: R. A. Meyers (Ed.), Computational Complexity, Springer (ISBN 978-1-4614-1799-6), New York, 2012, pp. 2578–2593. doi:10.1007/978-1-4614-1800-9_158.

[8] ISO/TC 176/SC 2, Quality management systems: Requirements, ISO 9001:2015, ISO, 2015.

[4] AI4Work (HEur GA 101135990): https://ai4work.eu/; BatCAT (HEur GA 101137725): https://batcat.info/; DigiPass CSA (HEur GA 101138510): https://ms.hereon.de/digipass/; MaRDI (NFDI – DFG ID 460135501): https://www.mardi4nfdi.de.

[5] https://github.com/OO-LD

[9] M. R. Munafò, B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. Percie du Sert, U. Simonsohn, E.-J. Wagenmakers, J. J. Ware, J. Ioannidis, A manifesto for reproducible science, Nature Human Behaviour 1 (2017) 0021. doi:10.1038/s41562-016-0021.

[10] M. Schappals, A. Mecklenfeld, L. Kröger, V. Botan, A. Köster, S. Stephan, E. J. García, G. Rutkai, G. Raabe, P. Klein, K. Leonhard, C. W. Glass, et al., Round robin study: Molecular simulation of thermodynamic properties from models with internal degrees of freedom, Journal of Chemical Theory and Computation 13 (2017) 4270–4280. doi:10.1021/acs.jctc.7b00489.

[11] R. D. Peng, Reproducible research in computational science, Science 334 (2011) 1226–1227. doi:10.1126/science.1213847.

[12] H. E. Plesser, Reproducibility vs. replicability: A brief history of a confused terminology, Frontiers in Neuroinformatics 11 (2018) 76. doi:10.3389/fninf.2017.00076.

[13] M. T. Horsch, S. Chiacchiera, G. Guevara Carrión, M. Kohns, E. A. Müller, D. Šarić, S. Stephan, I. T. Todorov, J. Vrabec, B. Schembera, Epistemic metadata for computational engineering information systems, in: Proc. FOIS 2023, IOS (ISBN 978-1-64368-468-0), 2023, pp. 302–317. doi:10.3233/faia231136.

[14] J. Ohlhorst, Trust Responsibly, Routledge (ISBN 978-1-032-46098-7), Abingdon, 2024.

[15] R. Baratella, G. Amaral, T. P. Sales, R. Guizzardi, G. Guizzardi, The many facets of trust, in: Proc. FOIS 2023, IOS (ISBN 978-1-64368-468-0), 2023, pp. 17–31. doi:10.3233/faia231115.

[16] S. Chiacchiera, A. T. Correia, D. Stokic, E. Sanfilippo, S. Borgo, C. Masolo, A. Sarkar, Y. Le Franc, J. Haack, C. Weck, J. F. Morgado, G. Goldbeck, et al., Report on the finalized review of domain interoperability, Deliverable 3.8, OntoCommons CSA, 2023. doi:10.5281/zenodo.10891825.

[17] M. Magas, D. Kiritsis, Industry Commons: An ecosystem approach to horizontal enablers for sustainable cross-domain industrial innovation, International Journal of Production Research 60 (2022) 479–492. doi:10.1080/00207543.2021.1989514.

[18] S. Borgo, C. Masolo, Ontological foundations of DOLCE, in: R. Poli, M. Healy, A. Kameas (Eds.), Theory and Applications of Ontology: Computer Applications, Springer (ISBN 978-90-481-8846-8), Dordrecht, 2010, pp. 279–295. doi:10.1007/978-90-481-8847-5_13.

[19] F. A. Zaccarino, C. Masolo, E. Ghedini, S. Borgo, From causation (and parthood) to time: The case of EMMO, in: Proc. FOIS 2023, IOS (ISBN 978-1-64368-468-0), 2023, pp. 92–106. doi:10.3233/faia231120.

[20] R. Arp, B. Smith, A. D. Spear, Building Ontologies with Basic Formal Ontology, MIT Press (ISBN 978-0-262-52781-1), Cambridge, Massachusetts, 2015.

[21] M. R. Naqvi, L. Elmhadhbi, A. Sarkar, B. Archimede, M. H. Karray, Survey on ontology-based explainable AI in manufacturing, Journal of Intelligent Manufacturing (2024). doi:10.1007/s10845-023-02304-z.

[22] T. Grote, K. Genin, E. Sullivan, Reliability in machine learning, Philosophy Compass 19 (2024) e12974. doi:10.1111/phc3.12974.

[23] S. Yablo, Aboutness, Princeton Univ. Press (ISBN 978-0-691-14495-5), 2014.

[24] J. M. Durán, N. Formanek, Grounds for trust: Essential epistemic opacity and computational reliabilism, Minds and Machines 28 (2018) 645–666. doi:10.1007/s11023-018-9481-6.

[25] I. Koskinen, We have no satisfactory social epistemology of AI-based science, Social Epistemology (2024). doi:10.1080/02691728.2023.2286253.

[26] P. B. Heidorn, Shedding light on the dark data in the long tail of science, Library Trends 57 (2008) 280–299. doi:10.1353/lib.0.0036.

[27] B. Schembera, J. M. Durán, Dark data as the new challenge for big data science and the introduction of the scientific data officer, Philosophy & Technology 33 (2020) 93–115. doi:10.1007/s13347-019-00346-x.

[28] B. Schembera, Like a rainbow in the dark: Metadata annotation for HPC applications in the age of dark data, Journal of Supercomputing 77 (2021) 8946–8966. doi:10.1007/s11227-020-03602-6.

[29] M. T. Horsch, Mereosemiotics: Parts and signs, in: Proc. JOWO 2021, CEUR-WS (CEUR Workshop Proceedings no. 2969), 2021, p. FOUST: 3.

[30] M. T. Horsch, S. Chiacchiera, B. Schembera, M. A. Seaton, I. T. Todorov, Semantic interoperability

based on the European Materials and Modelling Ontology and its ontological paradigm: Mereosemiotics, in: Proc. WCCM-ECCOMAS 2020, Scipedia, 2021, p. 297. doi:10.23967/wccm-eccomas.2020.297.

[31] A. de Baas, P. Del Nostro, J. Friis, E. Ghedini, G. Goldbeck, I. M. Paponetti, A. Pozzi, A. Sarkar, L. Yang, F. A. Zaccarini, D. Toti, Review and alignment of domain-level ontologies for materials science, IEEE Access 11 (2023) 120372–120401. doi:10.1109/access.2023.3327725.

[32] M. T. Horsch, S. Chiacchiera, M. A. Seaton, I. T. Todorov, K. Šindelka, M. Lísal, B. Andreon, E. Bayro Kaiser, G. Mogni, G. Goldbeck, R. Kunze, G. Summer, et al., Ontologies for the Virtual Materials Marketplace, Künstliche Intelligenz 34 (2020) 423–428. doi:10.1007/s13218-020-00648-9.

[33] M. T. Horsch, C. Niethammer, G. Boccardo, P. Carbone, S. Chiacchiera, M. Chiricotto, J. D. Elliott, V. Lobaskin, P. Neumann, P. Schiffels, M. A. Seaton, I. T. Todorov, et al., Semantic interoperability and characterization of data provenance in computational molecular engineering, Journal of Chemical & Engineering Data 65 (2020) 1313–1329. doi:10.1021/acs.jced.9b00739.

[34] M. T. Horsch, D. Toti, S. Chiacchiera, M. A. Seaton, G. Goldbeck, I. T. Todorov, OSMO: Ontology for simulation, modelling, and optimization, in: Proc. JOWO 2021, CEUR-WS (CEUR Workshop Proceedings no. 2969), 2021, p. FOIS Showcase: 47.

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

[36] M. T. Horsch, S. Chiacchiera, M. A. Seaton, I. T. Todorov, B. Schembera, P. Klein, N. A. Konchakova, Pragmatic interoperability and translation of industrial engineering problems into modelling and simulation solutions, in: Proc. DAMDID 2020, Springer (ISBN 978-3-030-81199-0), 2021, pp. 45–59. doi:10.1007/978-3-030-81200-3_4.

[37] M. Pezzotta, J. Friis, G. J. Schmitz, N. Konchakova, D. Höche, D. Hristova-Bogaerds, P. Asinari, L. Bergamasco, G. Goldbeck, Report on translation case studies describing the gained experience, Technical report, EMMC CSA, 2021. doi:10.5281/zenodo.4457849.

[38] R. Rudnicki, B. Smith, T. Malyuta, W. Mandrick, Best practices of ontology development, White paper, CUBRC, 2016.

[39] M. Amir, M. Baruah, M. Eslamialishah, S. Ehsani, A. Bahramali, S. Naddaf-Sh, Zarandioon, Truveta Mapper: A zero-shot ontology alignment framework, in: Proc. OM 2023, CEUR-WS (CEUR Workshop Proceedings no. 3591), 2023, pp. 1–12.

[40] D. Iglezakis, D. Terzijska, S. Arndt, S. Leimer, J. Hickmann, M. Fuhrmans, G. Lanza, Modelling scientific processes with the m4i ontology, in: Proc. CoRDI 2023, NFDI, 2023. doi:10.52825/cordi.v1i.271.

[41] S. Clark, F. L. Bleken, S. Stier, E. Flores, C. W. Andersen, M. Marcinek, A. Szczesna-Chrzan, M. Gaberšček, M. R. Palacin, M. Uhrin, J. Friis, Toward a unified description of battery data, Advanced Energy Materials 12 (2022) 2102702. doi:10.1002/aenm.202102702.

[42] H. Paulheim, A. Gangemi, Serving DBpedia with DOLCE: More than just adding a cherry on top, in: Proc. ISWC 2015, part I, Springer (ISBN 978-3-319-25006-9), Cham, 2015, pp. 180–196. doi:10.1007/978-3-319-25007-6_11.

[43] S. Stier, X. Xu, L. Gold, M. Möckel, Ontology-based battery production dataspace and its interweaving with artificial intelligence-empowered data analytics, Energy Technology 12 (2024) 2301305. doi:10.1002/ente.202301305.

[44] M. T. Horsch, D. Romanov, E. Valseth, S. Belouettar, L. E. Córdova López, J. Glutting, M. A. Janssen, P. Klein, A. Linhart, M. A. Seaton, E. D. Sødahl, N. Vizcaino, et al., Battery manufacturing knowledge infrastructure requirements for multicriteria optimization based decision support in design of simulation, in: Proc. SeMatS 2024, CEUR-WS, 2024.

[45] M. Reidelbach, B. Schembera, M. Weber, Towards a FAIR documentation of workflows and models in applied mathematics, in: Proc. ICMS 2024, Springer (ISBN 978-3-031-64528-0), 2024, pp. 254–262. doi:10.1007/978-3-031-64529-7_27.

[46] ISO/IEC JTC 1/SC 32, Information technology: Meta object facility, ISO 19502:2005, ISO, 2005.