Norges miljø- og biovitenskapelige universitet

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

M. T. Horsch,[1,2] S. Chiacchiera,[2] I. T. Todorov,[2] A. T. Correia,[3] A. Dey,[1] N. A. Konchakova,[4] S. Scholze,[3] S. Stephan,[5] K. Tøndel,[1] A. Sarkar,[6] M. H. Karray,[6] F. Al Machot,[1] B. Schembera[7]

[1]Norwegian Univ. Life Sciences, Ås, Norway [2]UKRI STFC Daresbury Laboratory, UK
[3]ATB Institute, Bremen, Germany [4]Helmholtz-Zentrum Hereon, Geesthacht, Germany
[5]RPTU Kaiserslautern, Germany [6]Univ. Technol. Tarbes Occitanie Pyrénées, France
[7]Univ. Stuttgart, Germany

NMBU Fakultet for realfag og teknologi Forskergruppe materialteori og -informatikk

XAIR principles working group

XAI: Explainable artificial intelligence

AIR: Artificial-intelligence ready

XAIR: Explainable AI-ready

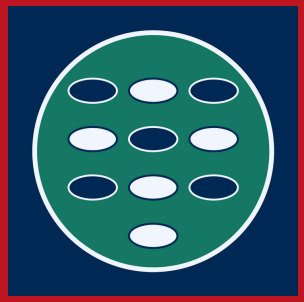
Data are **XAIR** to the degree that they are semantically enriched so that **best use** can be made of **interpretable learning** techniques. These include XAI in the narrow sense (learning by induction) and logical reasoning (deduction).

Tendency: **Making data trustworthy through explanations** will increasingly become a requirement. Data must become explainable-AI-ready (XAIR).

Slogan: "FAIR and XAIR data." (Sounds similar to the idiom "fair and square.")

Noregs miljø- og biovitskaplege universitet

Materialteori og -informatikk



Digitalisering på Ås

BatCAT Battery Cell Assembly Twin

DigiPass Know your products!

KGA

Epistemic metadata and mid-level ontology development

DAO-XAI 2024

19. oktober 2024

Materialteori og -informatikk

Opacity vs. transparency

European AI Act: "To address concerns related to **opacity** and […] fulfil their obligations under this Regulation, **transparency** should be required for high-risk AI systems before they are placed on the market […]. High-risk AI systems should […] enable deployers to understand how the AI system works […]. High-risk AI systems should be accompanied by **appropriate information**".

Epistemic opacity:

The concept was introduced by **Humphreys** in *Extending Ourselves*[1]

[1]P. Humphreys, *Extending Ourselves Computational Science, Empiricism, and Scientific Method*, **2004**.

Opacity vs. transparency

European AI Act: "To address concerns related to **opacity** and […] fulfil their obligations under this Regulation, **transparency** should be required for high-risk AI systems before they are placed on the market […]. High-risk AI systems should […] enable deployers to understand how the AI system works […]. High-risk AI systems should be accompanied by **appropriate information**".

Epistemic opacity can occur when simulation-based and data-driven methods are used. The concept was introduced by **Humphreys** in *Extending Ourselves*[1] (2004), developed further in later work,[2] and has had a substantial impact.[3]

> **Epistemic opacity** (Humphreys, 2011): A «process is **epistemically opaque** relative to a cognitive agent *X* at time *t* […if …] *X* does not know at *t* all of the **epistemically relevant elements**»[2]

[1]P. Humphreys, *Extending Ourselves Computational Science, Empiricism, and Scientific Method*, **2004**.

[2]P. Humphreys, in M. Carrier, A. Nordmann, *Science in the Context of Application*, *pp.* 131–142, Springer, **2011**.

[3]J. M. Durán, N. Formanek, *Minds and Machines* **28**(4): 645–666, doi:10.1007/s11023-018-9481-6, **2018**.

Epistemic metadata

Epistemic metadata are the information that **establishes the knowledge status** of data or digital objects.[1]

> **Questions we must answer to establish the knowledge status:**
>
> **a)** "what **knowledge claim** φ has been formulated?,"
> **b)** "where do the data and the claim come from?" (**provenance**),
> **c)** "what **validity claim** was made about φ?,"
> **d)** "why should we accept any of this?" (**grounding**).

Key epistemic metadata items are the **knowledge claims** made based on data, their **provenance**, **validation** and **reproducibility**, and **epistemic grounding**.

In *Proc. JOWO 2022*, CEUR *vol.* **3249**: *p. 2 (CAOS)*, CEUR-WS, **2022**.
In *Proc. ICAPAI 2023*, doi:10.1109/icapai58366.2023.10193944, IEEE, **2023**.
In *Proc. FOIS 2023*, *pp.* 302–319, doi:10.3233/faia231136, IOS, **2023**.

Requirements for epistemic metadata: Case study

Epistemic metadata and their documentation were explored for the domain of molecular modelling and simulation within engineering thermodynamics:

First stage report (10 cases), doi:10.5281/zenodo.7516532, **2023**.

Discussion of *five papers each* from *two research groups* (London, Berlin) without involving the papers' authors. Obtained a tentative **taxonomy for epistemic metadata**, later implemented into the PIMS-II ontology.

Second stage report (12 claims), doi:10.5281/zenodo.7608074, **2023**.

Discussion of *two claims each* from *six papers*, with two papers each from three research groups (London, Berlin, Kaiserslautern), involving the papers' authors. Discussed aspects such as the **grounding of knowledge claims** with authors.

Imperial College London

berlin

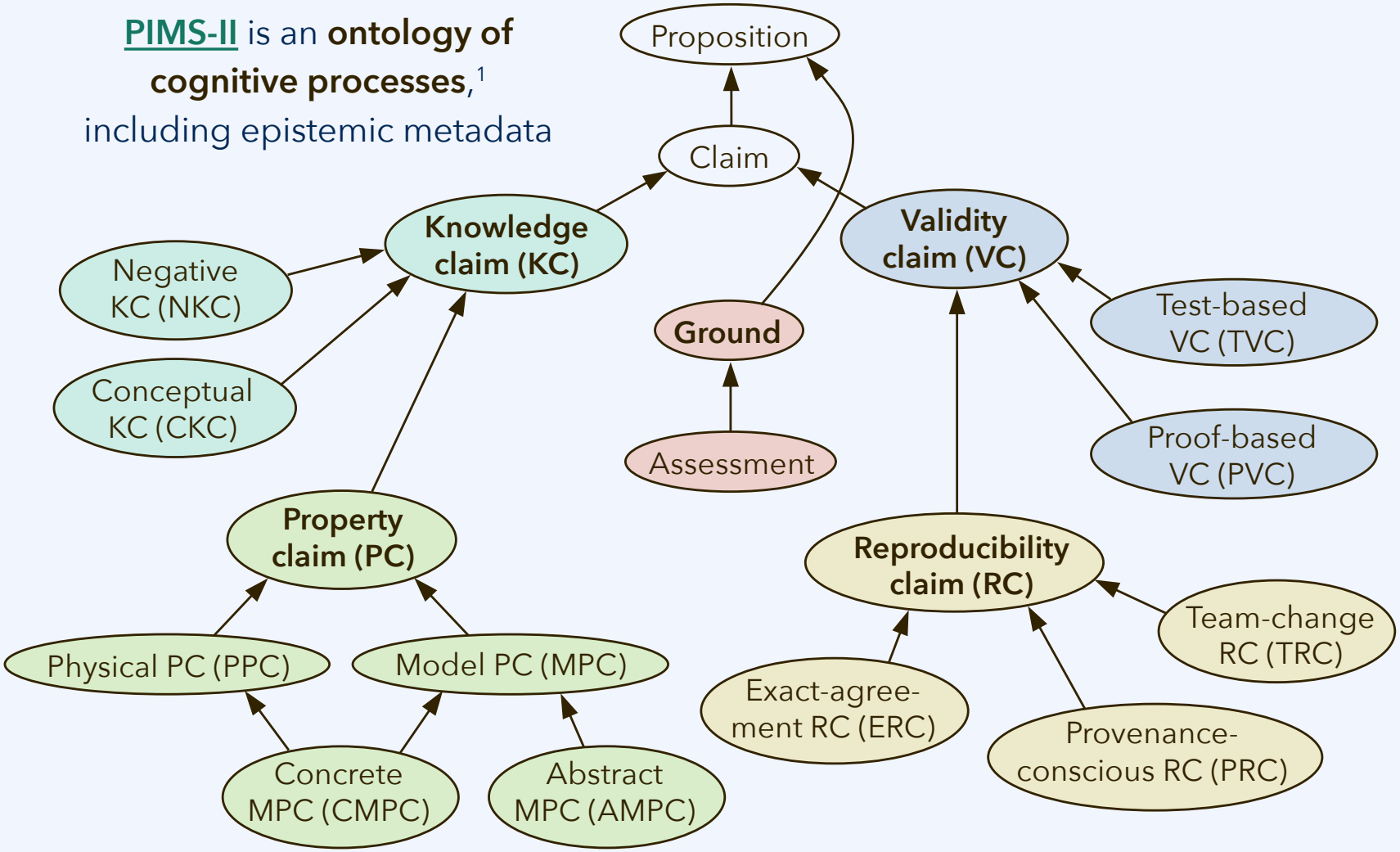
RPTU Rheinland-Pfälzische Technische Universität

NMBU

UKRI Science and Technology Facilities Council

Mid-level ontology of epistemic metadata

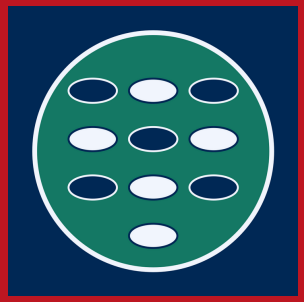
PIMS-II is an **ontology of cognitive processes**,[1] including epistemic metadata



[1]OWL implementation under http://www.molmod.info/semantics/pims-ii.ttl

Noregs miljø- og biovitskaplege universitet

Materialteori og -informatikk



Digitalisering på Ås

Ontology redesign and community work on XAIR principles

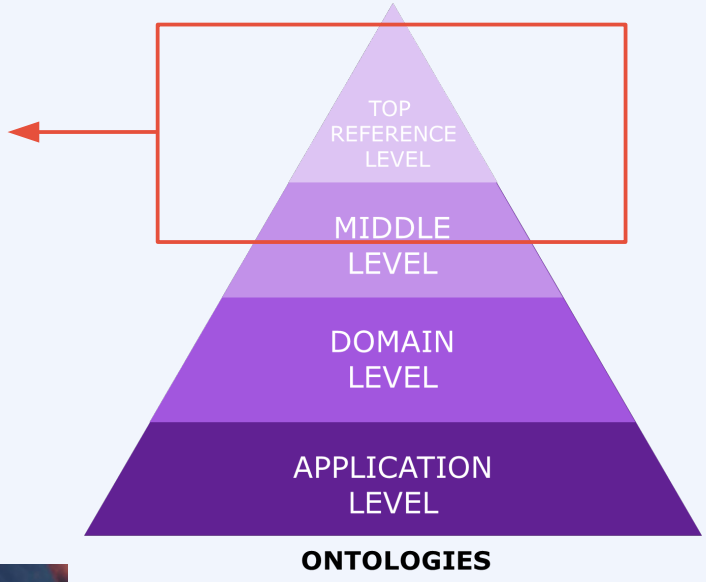
DAO-XAI 2024

19. oktober 2024

Materialteori og -informatikk

OntoCommons CSA (2020 – 2023)

- **Community development** that brought together applied ontology, industrial digitalization, and computational science and engineering.
- OntoCommons ecosystem[1] (OCES) including the three **foundational ontologies** BFO, DOLCE, and EMMO. Development of **bridge concepts**[2] as a technique for ontology alignment.



[1]M. Magas, D. Kiritsis, *Int. J. Production Res.* **60**(2): 479–492, doi:10.1080/00207543.2021.1989514, **2022**.

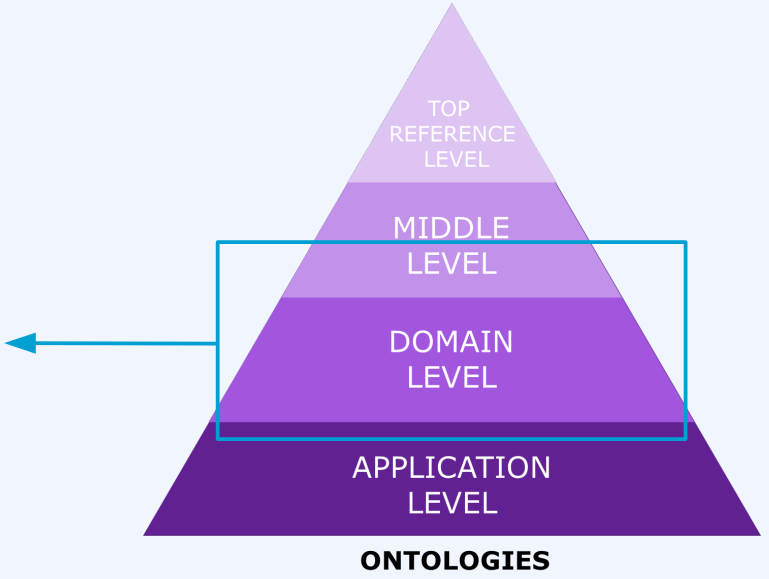
[2]A. de Baas *et al.*, *IEEE Access* **11**: 120372–120401, doi:10.1109/access.2023.3327725, **2023**.

Review of Domain Interoperability[1]

OntoCommons CSA collected and supported the design and alignment of **domain-level inter-operability standards**.

The overall analysis of modes of interoperability, relevant tools and components, and recommen-dations was delivered in the form of **RoDI: The Review of Domain Interoperability**.[1]

In particular, there are **syntactic**, **semantic**, and **pragmatic** modes of interoperability.[1]



	data	human	organization	software
data (d)	$d \leftrightarrow d$	$d \leftrightarrow h$	$d \leftrightarrow o$	$d \leftrightarrow s$
human (h)	—	$h \leftrightarrow h$	$h \leftrightarrow o$	$h \leftrightarrow s$
organization (o)	—	—	$o \leftrightarrow o$	$o \leftrightarrow s$
software (s)	—	—	—	$s \leftrightarrow s$

Matrix structure for interoperability requirements: Who interoperates with whom?

[1]S. Chiacchiera *et al.*, OntoCommons deliverable 3.8, "Finalized Review of Domain Interoperability," **2023**.

XAIR principles WG

March – December 2024

Months 1 to 10: **Synopsis of literature work on core concepts**

- Identify the core concepts, analyse and summarize the literature characterizing these concepts.
- M10, report, public, "**Synopsis of XAIR core concepts**."

Months 11 to 16: **Work on actionable core concepts**

- Discussion of use cases; annotation of examples; draft of mid-level and domain ontologies.
- M16, report, public, "Request for comments on actionable XAIR core concepts."

Months 17 to 22: **Work and consultation on XAIR principles**

- Community discussion toward “XAIR principles.”
- M22, report, public, "Request for comments on XAIR data and metadata […] principles […]".

Months 23 to 34: **Stable release of mid and domain ontologies**

- Ontologies for XAIR are finalized to the extent that no major changes will occur in the future.
- M34, ontologies, public, "[…] Ontologies covering the XAIR key concepts […]".

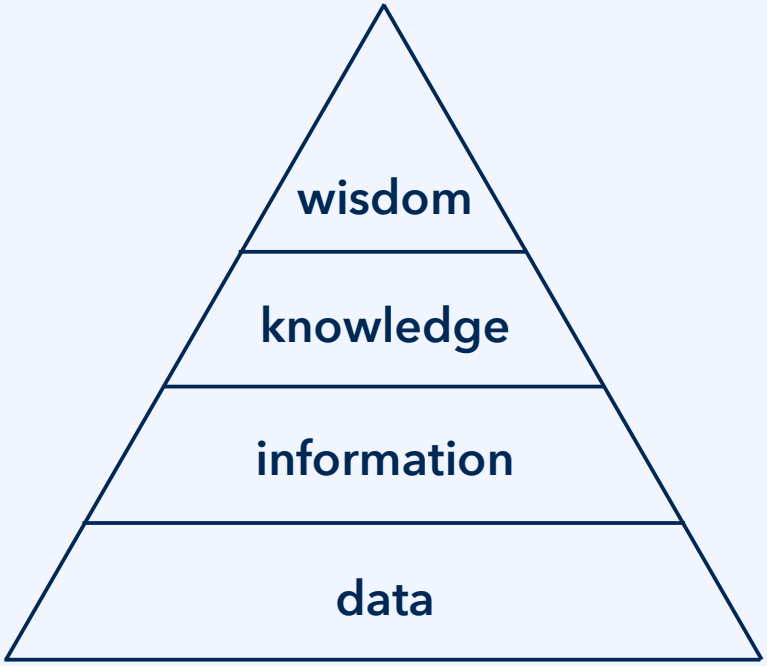
Months 35 to 40: **XAIR forward plan and ontology governance**

- Critical analysis of accomplishments and forward-looking discussion.
- M40, report, public, "XAIR forward plan and ontology governance."

XAIR principles WG: Themes of discussion

Hierarchy of data, information, knowledge, and wisdom (**DIKW pyramid**[1]).

[1]J. Rowley, *J. Inform. Sys.* **33**: 163–180, doi:10.1177/0165551506070706, **2009**.



Pragmatic competency and interoperability, including agreed good practices.

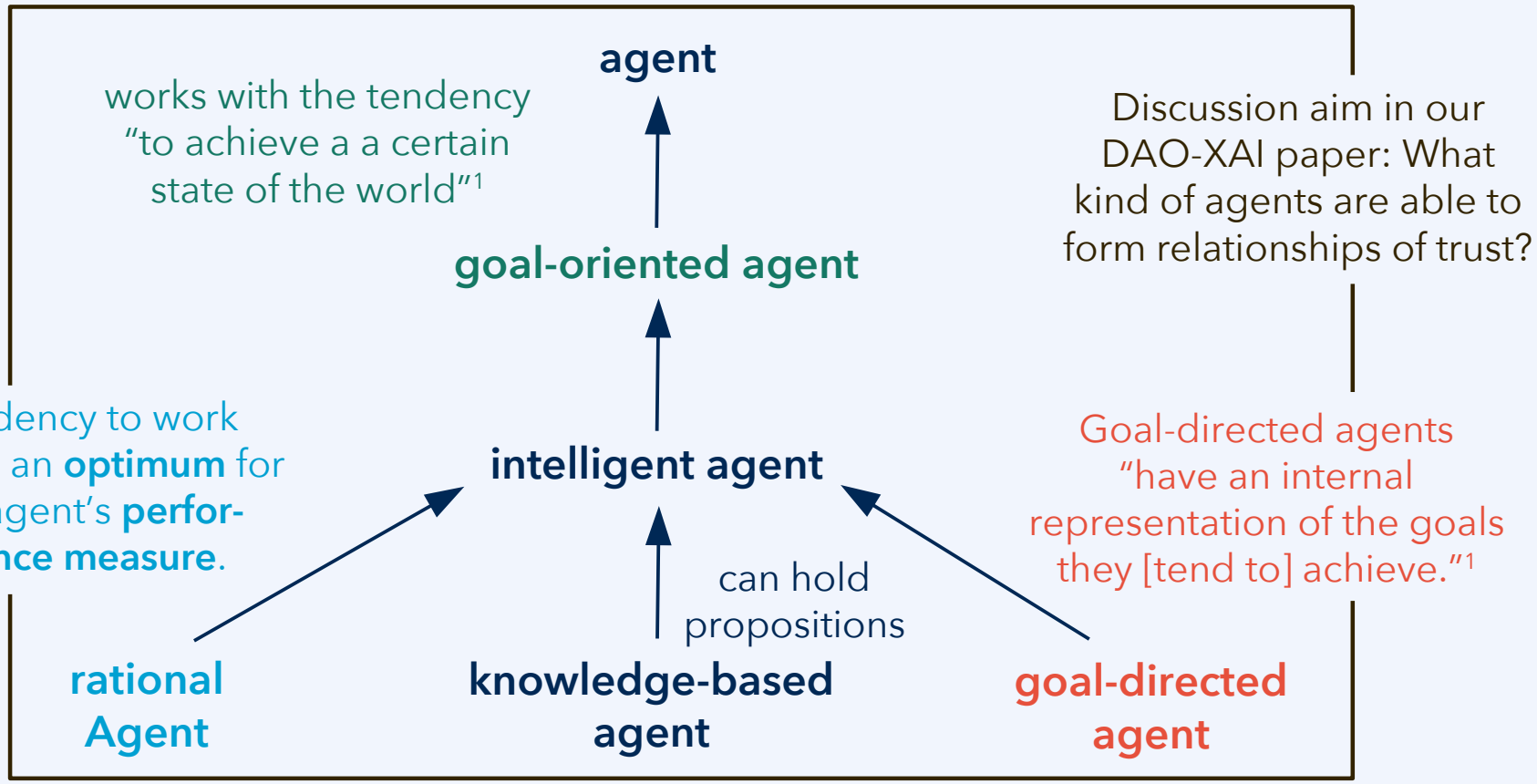
Epistemic metadata documentation: Establish the knowledge status.

Semantic interoperability: Data become information if their meaning is agreed.

Syntactic interoperability: Data exchanged in an agreed format.

XAIR principles WG: Themes of discussion

Agency: Discussed with the taxonomy by Conte[1] as a basis.



[1]R. Conte, "Rational, goal-oriented agents," doi:10.1007/978-0-387-30440-3_445, in R. A. Meyers (ed.), *Encyclopedia of Complexity and Systems Science*, Springer, **2009**.

XAIR principles WG: Themes of discussion

Reproducibility: Discussed with the review by Plesser[1] as a basis.

Consider the case where a reproducing researcher *b* contradicts findings by *a*:

1) Reseacher *a* did κ and found φ.

2) Reseacher *b* did γ and found $\zeta \neq \varphi$.

[1]H. E. Plesser, *Frontiers Neuroinform.* **11**: 76, doi:10.3389/fninf.2017.00076, **2018**.

XAIR principles WG: Themes of discussion

Reproducibility: Discussed with the review by Plesser[1] as a basis.[2]

Common formulation and schema for reproducibility claims (RCs):

«Whenever research process κ'' is carried out, it must lead to the outcome φ''.»

Consider the case where a reproducing researcher *b* contradicts findings by *a*:

1) Reseacher *a* did κ (consistent with κ'') and found φ (consistent with φ'').

   Here, *a* also made the **positive reproducibility claim $\psi = \Box(\varphi'' \mid \kappa'')$.**

2) Reseacher *b* did γ, **consistent with κ''**, and found ζ, **inconsistent with φ''**.

   Here, *b* made the **negative reproducibility claim $\Diamond(\neg\varphi'' \mid \kappa'') \equiv \neg\Box(\varphi'' \mid \kappa'') \equiv \neg\psi$.**

3) What is relevant there is the **contradiction between ψ and $\neg\psi$**.

Claim ψ is usually implicit, ascribed to *a* based on unwritten community rules.[2]

[1]H. E. Plesser, *Frontiers Neuroinform.* **11**: 76, doi:10.3389/fninf.2017.00076, **2018**.

[2]In *Proc. FOIS 2023*, *pp.* 302–317, doi:10.3233/faia231136, **2023**.

Epistemic grounding, reliance, and trust

Trust: The majority view among philosophers of trust is: **Trust is reliance + X**, with different conceptualizations of what must be there on top of reliance.[1, 2]

Applied ontology community: Baratella *et al.*[3] have been working on trust.

They argue:[3] "The trustor is necessarily an 'intentional entity', that is, a cognitive agent, an agent endowed with [*i.e.* conscious of] goals and beliefs."

"the thesis that trust is reliance plus some extra factor seems implausible."

"suppose that Tom trusts Mary for keeping his secrets, but he continues to look for evidence that she will keep them. Clearly, he is not trusting her."

[1]Faulkner, *Analysis* **73**(3): 424–429, doi:10.1093/analys/anv037, **2015**.
[2]Stout, *Int. J. Philos. Studies* **30**(3): 339–356, doi:10.1080/09672559.2022.2121892, **2022**.
[3]Baratella *et al.*, «The many facets of trust», in *Proc. FOIS 2023*, doi:10.3233/faia231115, **2024**.

Epistemic grounding, reliance, and trust

Trust: The majority view among philosophers of trust is: **Trust is reliance + X**, with different conceptualizations of what must be there on top of reliance.

Applied ontology community: Baratella *et al.*[1] have been working on trust.

They argue:[1] "The trustor is necessarily an 'intentional entity', that is, a cognitive agent, an agent endowed with [*i.e.* conscious of] goals and beliefs." (In our formalization: A goal-directed agent who is also a knowledge-based agent.)

Our DAO-XAI paper follows a paradigm where trust does not require reliance, but *enables reliance even in the absence of a complete reliabilist grounding*. In this, we believe that we are close to being in agreement with Baratella *et al.*[1]

We conclude that the trustor must be an agent able to *not know* some proposition φ, yet emulate an agent who knows (or is certain) that φ holds.

[1]Baratella *et al.*, «The many facets of trust», in *Proc. FOIS 2023*, doi:10.3233/faia231115, **2024**.

Epistemic grounding, reliance, and trust

Trust: See also the reference ontology of trust ONTrust by Baratella *et al.*[1]

	trust	reliance
Type-1 The **results** establish their own validity.[2]	*Typical:* Mathematical argument (proof) over of a conceptual framework designed around widely accepted definitions and axioms.	*Schema:* A new theory is more reliable because it is simpler, covers more phenomena, or represents underlying physics. (**theoretical virtues**)
Type-2 The **provenance** of the results tells that they are valid.[2]	*Case study example:* Chatwell and Vrabec argue: It is OK to use a cutoff radius of 5.5σ for the LJ potential, since this was done in three cited works from the literature.	*Typical:* We used a model, method, and simulation code validated in the past and – usually – very accurate. (**process reliabilism**)

[1]https://github.com/unibz-core/trust-ontology

[2]In *Proc. JOWO 2022*, *p.* 2 (CAOS), **2022**.

Mid-level ontology refactoring

PURL for the new system, MSO-EM (ontologies for **modelling, simulation, optimization**, and **epistemic metadata**), which is under construction:

https://www.purl.org/mso-em

BatCAT organizational github: https://github.com/HE-BatCAT

Design principles:

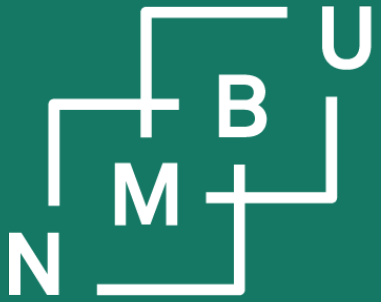
- Strong alignment with DOLCE (through DOLCE Lite)
- OWL2 EL profile expressivity level
- Ongoing development, with easy stable access to versioned releases
- Simple modules, each with maximum three taxonomy levels and maximum three top concepts
- Backwards compatibility with equivalences to the preceding mid-level ontology development (PIMS-II) to the maximum possible extent
- All modules of the ontology are directly aligned with DOLCE

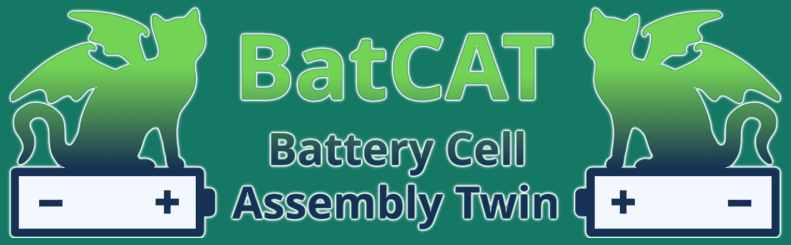
DigiPass CSA has received funding from the European Union's Horizon Europe research and innovation programme under **grant agreement no. 101138510**. Views and opinions expressed are however those of the author only and do not necessarily reflect those of the project, the European Health and Digital Executive Agency (EHDEA), or the European Union. Neither DigiPass CSA nor the EHDEA or the EU can be held responsible for them.

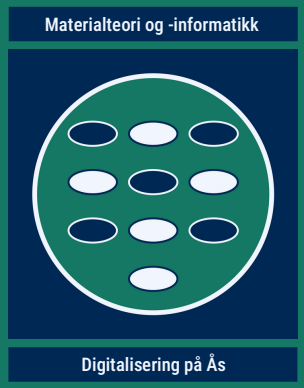
Finansiert av Den europeiske union

19th October 2024

NMBU Fakultet for realfag og teknologi Forskergruppe materialteori og -informatikk







BatCAT has received funding from the European Union's **Horizon Europe** research and innovation programme under **grant agreement no. 101137725**. Views and opinions expressed are however those of the author only and do not necessarily reflect those of the project, the European Climate, Infrastructure and Environment Executive Agency (CINEA), or the European Union. Neither BatCAT nor the CINEA or the EU can be held responsible for them.

Finansiert av Den europeiske union

19th October 2024

NMBU Fakultet for realfag og teknologi Forskergruppe materialteori og -informatikk

Norges miljø- og biovitenskapelige universitet

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

M. T. Horsch,[1,2] S. Chiacchiera,[2] I. T. Todorov,[2] A. T. Correia,[3] A. Dey,[1] N. A. Konchakova,[4] S. Scholze,[3] S. Stephan,[5] K. Tøndel,[1] A. Sarkar,[6] M. H. Karray,[6] F. Al Machot,[1] B. Schembera[7]

[1]Norwegian Univ. Life Sciences, Ås, Norway [2]UKRI STFC Daresbury Laboratory, UK
[3]ATB Institute, Bremen, Germany [4]Helmholtz-Zentrum Hereon, Geesthacht, Germany
[5]RPTU Kaiserslautern, Germany [6]Univ. Technol. Tarbes Occitanie Pyrénées, France
[7]Univ. Stuttgart, Germany

NMBU Fakultet for realfag og teknologi Forskergruppe materialteori og -informatikk