

An Ontology for Machine Learning Interatomic Potentials

Daniel Hernández^{1,*}, Jong Hyun Jung², Yuji Ikeda², Yongliang Ou², Pranav Kumar²,
Tom Schächtel², Wenchuan Liu², Xin Li², Xi Zhang², Xiang Xu², Lifang Zhu²,
Fritz Körmann^{3,4}, Steffen Staab¹ and Blazej Grabowski²

¹*Institute for Artificial Intelligence, University of Stuttgart, Germany*

²*Institute for Materials Science, University of Stuttgart, Germany*

³*Interdisciplinary Centre for Advanced Materials Simulation (ICAMS), Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany*

⁴*Department for Computational Materials Design, Max-Planck-Institut for Sustainable Materials, Max-Planck-Str. 1, 40237 Düsseldorf, Germany*

Abstract

Machine learning interatomic potentials (MLIPs) approximate quantum-mechanical energies and forces—conventionally computed by density functional theory (DFT) or wave-function methods—at a fraction of the cost, and the field has grown into an ecosystem of algorithms, training datasets, hyperparameters, and target materials. Yet the metadata needed to compare, reproduce, and build upon MLIP studies remains scattered across papers, scripts, and ad-hoc file formats: selecting a method for a material or ranking published accuracies requires a manual literature survey. We ask what metadata a semantic layer for MLIP studies must express—captured in nine competency questions elicited from materials scientists and MLIP developers—and whether existing semantic resources can express it. A term-by-term gap analysis shows that they cannot: neither the machine-learning ontologies (ML-Schema, DMOP, MEX, CMO) nor the materials-science ontologies (EMMO, MDO, PMDco, CM-SO/ASMO), nor even their hypothetical union, answers most of the competency questions. We fill this gap with the MLIPs Ontology, an OWL 2 DL ontology in three modules—Method, Training Data, and Benchmark—that connects the existing ontologies rather than displacing them; 27 formal axioms enforce data completeness and consistency. We demonstrate the ontology on a running example (Moment Tensor Potentials for TiCr₂-H) and validate it extensionally: all nine competency questions run as SPARQL queries over a 20-paper seeded knowledge graph, backed by reasoner and pitfall checks.

Keywords

Ontology, Machine learning interatomic potentials, Materials science, OWL, FAIR data

1. Introduction

Machine learning interatomic potentials (MLIPs) [1, 2] predict atomic energies and forces at near the accuracy of their quantum-mechanical reference method—conventionally density functional theory (DFT) [3], increasingly also wave-function methods—at a fraction of its cost, enabling atomistic simulations—molecular dynamics, Monte Carlo, phonon calculations—at scales DFT cannot reach. The community around them is sizable and growing: the DFT code VASP, the molecular-dynamics engine LAMMPS, and the MLIP framework DeepMD-kit drew 13,200, 4,190, and 1,510 citations in 2025 [4], and MLIP-related publications rose from ~50/year in 2015 to over 1,500 in 2024 [5, 6]. MLIP families have proliferated over the past decade—high-dimensional neural network potentials (HDNNP) [1], Gaussian approximation potentials (GAP) [2], moment tensor potentials (MTP) [7], the atomic cluster expansion

SeMatS 2026: The 3rd International Workshop on Semantic Materials Science co-located with the 25th International Semantic Web Conference (ISWC 2026), October 25th, 2026, Bari, Italy

*Corresponding author.

✉ daniel.hernandez@ki.uni-stuttgart.de (D. Hernández); jong-hyun.jung@imw.uni-stuttgart.de (J. H. Jung);
yuji.ikeda@imw.uni-stuttgart.de (Y. Ikeda); yongliang.ou@imw.uni-stuttgart.de (Y. Ou); pranav.kumar@imw.uni-stuttgart.de
(P. Kumar); xi.zhang@imw.uni-stuttgart.de (X. Zhang); lifang.zhu@imw.uni-stuttgart.de (L. Zhu); fritz.koermann@rub.de
(F. Körmann); Steffen.Staab@ki.uni-stuttgart.de (S. Staab); blazej.grabowski@imw.uni-stuttgart.de (B. Grabowski)

ORCID 0000-0002-7896-0875 (D. Hernández); 0000-0002-2409-975X (J. H. Jung); 0000-0001-9176-3270 (Y. Ikeda);
0000-0003-1486-1197 (Y. Ou); 0000-0002-3661-5870 (P. Kumar); 0009-0001-4894-0195 (T. Schächtel); 0009-0009-9043-2491
(W. Liu); 0009-0003-3860-8388 (X. Li); 0000-0003-0319-8203 (X. Zhang); 0000-0002-4399-0580 (X. Xu); 0000-0003-2740-8172
(L. Zhu); 0000-0003-3050-6291 (F. Körmann); 0000-0002-0780-4154 (S. Staab); 0000-0003-4281-5665 (B. Grabowski)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

(ACE) [8], equivariant message-passing networks such as MACE [9] and NequIP [10], and universal graph-network potentials such as M3GNet [11]—each with its own architecture, hyperparameter space, and application domain.

This diversity makes routine tasks—selecting an algorithm for a given material, building a training set, comparing results across papers—hard [12]. Study metadata (algorithm, training set, material, DFT settings, accuracy) is scattered across papers, supplements, scripts, and library-specific formats. This raises two questions: *what* metadata must a semantic layer for MLIP studies express—we capture the answer in nine competency questions elicited from materials scientists and MLIP developers (§4)—and can existing semantic resources express it? Our gap analysis (§6.3) answers the second question negatively: ML-Schema [13], DMOP [14], MEX [15], and CMO [16] model generic ML workflows; EMMO [17], MDO [18], PMDco [19], and CMSO/ASMO [20] model the physical domain (all introduced in §3); none of them—nor even their hypothetical union—captures the MLIP-specific concepts: physics-aware hyperparameters, training-dataset construction strategies, algorithm-aligned DFT reference settings, and benchmark accuracy across method/material pairs.

To fill this gap we present the *MLIPs ontology*, an OWL 2 DL ontology [21] organized into three modules (*Method*, *Training Data*, *Benchmark*) that connects existing schemas rather than displacing them, reusing established vocabularies wherever an existing class or property already names the concept at hand: 22 formal alignment axioms plus 14 `rdfs:subClassOf/subPropertyOf` edges thread its terms through ML-Schema, PROV-O, MDO, CMSO/ASMO, schema.org, and QUDT (§5.2); 27 additional axioms enforce the MLIP-specific structure that no single existing ontology supplies, and SHACL shapes validate instance data. Its conceptual contributions are three MLIP-specific design decisions: hyperparameters are first-class type-level entities with units, defaults, and ranges (not literal values on runs); a DFT / wave-function reference-calculation hierarchy uses `rdfs:subPropertyOf` so legacy DFT data validates against family-agnostic super-properties; and an `mlips:metaSort` annotation per class carries the OntoClean [22] classification. The schema is exercised on a 20-paper expert-validated knowledge graph (all nine competency questions return non-empty rows) and is served at <https://w3id.org/mlips> with content negotiation. For consumers, cross-paper questions—which method reaches a target accuracy on a material class, which training data exists with compatible DFT settings—reduce to single SPARQL queries instead of a manual literature survey, for human researchers and autonomous agents alike; for producers, an encoded study is one that those selection queries can find, and the encoding protocol doubles as a reporting checklist that surfaces unreported metadata before reviewers do. The ontology thereby contributes a semantic layer for MLIP metadata to the FAIRification efforts under way in materials science. A concrete usage plan sustains the resource beyond this paper (§7): the META-LEARN project (2026–2030) commits the two Stuttgart institutes to maintenance and feeds the knowledge graph from its own MLIP studies, practitioners extend the controlled vocabulary without ontology-engineering skills, and we are federating with the maintainers of complementary materials-science vocabularies.

2. Running Example: Moment Tensor Potentials for TiCr₂ Laves Phases

We introduce the MLIPs ontology (presented in Section 5) through a concrete published study: training a Moment Tensor Potential (MTP) [7, 23] for hydrogen absorption in the TiCr₂–H system by Kumar et al. [24]. TiCr₂ is an *intermetallic compound* which absorbs hydrogen (*H*). It crystallizes in two close-packed forms (cubic and hexagonal), and Kumar et al. train one MTP per form over the hydrogen-content range $0 < x \leq 6$ for TiCr₂H_{*x*} (the stoichiometric ratio of H atoms per TiCr₂ formula unit). We use the cubic form (called *C15* in the materials literature) as the running example, following an MLIP study’s full lifecycle from the underlying physics through hyperparameters and training data to the evaluation of the trained model.

A *simulation cell* is a finite region of space with periodic boundary conditions containing *n* atoms at positions $\mathbf{r}_1, \dots, \mathbf{r}_n$ with atomic types $z_1, \dots, z_n$ ($z_i \in \{\text{Ti}, \text{Cr}, \text{H}\}$ for our TiCr₂H_{*x*} running example).

An interatomic potential predicts the total energy E and forces $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} E$ acting on each atom given only the atomic positions and types. DFT computes this energy accurately by solving the Kohn–Sham equations [3], but scales as $O(n^3)$ with the number of atoms, limiting it to systems of a few hundred atoms; MLIPs approximate the DFT energy surface at a fraction of the cost—scaling linearly in n for strictly local potentials—enabling simulations of millions of atoms. In the ontology, the simulation cells that populate a training dataset (atoms, positions, types) enter as instances of `AtomicConfiguration`. The example descends to this level of physical detail deliberately: reproducing or reusing a training dataset requires the simulation cells, reference settings, and sampling choices, not merely the name of the learning algorithm.

The MTP method. In the MTP framework [7], the total energy of an atomic configuration `cfg` is decomposed into local contributions $E^{\text{mtp}}(\text{cfg}) = \sum_{i=1}^n V(\mathcal{N}_i)$, where $\mathcal{N}_i$ is the *neighborhood* of atom i : the central atom’s type z_i and, for each neighbor j within the cutoff radius R_{cut} , its type z_j and relative position $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$. Atoms beyond R_{cut} do not contribute to $V(\mathcal{N}_i)$ —this *locality assumption* keeps the energy computation tractable. The local energy $V(\mathcal{N}_i) = \sum_{\alpha} \xi_{\alpha} B_{\alpha}(\mathcal{N}_i)$ is a linear combination of basis functions, where the B_{α} are *moment tensor descriptors*—invariant functions of the neighborhood that capture the local atomic environment—and $\xi = \{\xi_{\alpha}\}$ are the *parameters to be learned* by fitting to DFT reference data. The basis functions involve radial functions that vanish smoothly at R_{cut} and are evaluated above a minimum interatomic distance R_{min} (a numerical lower bound; physical interactions are not truncated). MTP thus has four hyperparameters: two geometric (the cutoff radius R_{cut} and minimum interatomic distance R_{min}) and two algorithmic that control basis size (the number of radial basis functions N_Q and the maximum polynomial level lev_{max} of the moment-tensor descriptors, both systematically increasable for higher accuracy).

The MLIP workflow—and the ontology—separate three levels of parameterization: *hyperparameter definitions* (a parameter’s name, data type, valid range, default; declared once in the controlled vocabulary or paper-locally with a `candidateForVocabulary` marker), *hyperparameter settings* (concrete values chosen for a specific MLIP run, e.g., $R_{\text{cut}} = 5.0 \text{ \AA}$, $\text{lev}_{\text{max}} = 16$), and the *learned parameters* ξ that define the trained model.

To enable queries such as “which physically motivated hyperparameters does this method accept?”, we classify hyperparameters into three non-disjoint subclasses of `Hyperparameter`: physical (parameters set by reference to a physical scale, such as an interaction range—strictly speaking, every hyperparameter is a parameter of the mathematical approximation rather than of the physics, and this subclass marks those whose values domain scientists choose on physical grounds), architectural (model-architecture choice), and training (training-procedure parameter). The non-disjointness is a modelling choice informed by domain expertise: R_{cut} and R_{min} are physically motivated (they tie to interaction-range and short-distance scales), the optimizer’s learning rate is purely training, but N_Q and lev_{max} are both architectural *and* physically motivated—they control the model size and also determine how finely the radial and angular expansions are resolved. The same pattern recurs in other families: HDNNP’s symmetry-function parameters fix both the network’s input dimension and the resolution at which the physical environment is described. A strict partition would misrepresent how materials scientists think about these parameters.

Encoding: run, settings, model. The example is expository, not the design method: the schema’s decisions were elicited from the competency questions of Section 4 and validated extensionally on the 20-paper corpus of Section 6, of which this study is one entry. The listings below show what the finished vocabulary looks like in use.¹ An `MLIPRun` applies a method to a training dataset under concrete hyperparameter settings—each an instance of `HyperparameterSetting` linking to a definition—and produces the trained model. A `TrainedModel` holds the fitted coefficients ξ :

¹Prefixes used in this paper: `mlips:` = <https://w3id.org/mlips/> (the ontology), `entity:` = <https://w3id.org/mlips/entity/> (instances), `mls:`, `prov:`, `qudt:`, `schema:`, `wd:` for ML-Schema, PROV-O, QUOT, schema.org, and Wikidata, and `mso:`, `cmso:`, `asmo:` for the materials-science ontologies of §3.

Listing 1: MLIP run, a hyperparameter setting, and the trained model.

```
entity:run-mtp-TiCr2H-c15 a mlips:MLIPRun ;
  mlips:appliesMethod entity:MTP ;
  mlips:runsOn entity:ds-TiCr2H-c15 ;
  mlips:produces entity:model-mtp-TiCr2H-c15 ;
  mlips:hasHyperparameterSetting entity:setting-rcut .
entity:setting-rcut a mlips:HyperparameterSetting ;
  mlips:forHyperparameter mlips:cutoffRadius ;
  mlips:settingValue "5.0"^^xsd:double ;
  mlips:hasUnit unit:ANGSTROM .
entity:model-mtp-TiCr2H-c15 a mlips:TrainedModel ;
  rdfs:label "C15-MTP for TiCr2-H (lev16, rcut5)" .
```

Hyperparameter *definitions* are attached to the method, whereas the *settings* of Listing 1 belong to the run. The method also names its three components (functional form, loss function, training algorithm), supported simulation types, and implementations (Listing 2):

Listing 2: MTP method declaration.

```
entity:MTP a mlips:MLIPMethod ;
  rdfs:label "Moment Tensor Potential" ;
  mlips:hasFunctionalForm entity:mtp-functional-form ;
  mlips:hasLossFunction entity:mtp-loss ;
  mlips:hasTrainingAlgorithm entity:bfgs ;
  mlips:hasHyperparameter mlips:cutoffRadius, mlips:minInteratomicDistance,
    entity:hp-nq, mlips:momentLevel ;
  mlips:hasImplementation entity:mlip-package-v2 ;
  mlips:supportsSimulation entity:sim-md, entity:sim-geopt .
```

The referenced individuals are declared analogously, each with a human-readable label; entity:hp-nq (the paper-local hyperparameter N_Q) additionally carries its datatype and a candidateForVocabulary marker (§7).

Training data. Ground truth comes from DFT calculations of total energy, forces, and (optionally) stresses for each configuration; configurations are drawn from MD trajectories, random perturbations, or active learning [23]. The initial low-hydrogen enumeration in Kumar et al. [24] yields 1,019 configurations on $2 \times 2 \times 2$ C15 supercells (the simulation-cell size: the unit cell repeated twice along each lattice vector), computed in VASP with PBE/PAW, a 400 eV plane-wave cutoff, and a $4 \times 4 \times 4$ Γ -centered k -point mesh—the Brillouin-zone sampling of the DFT calculation, a quantity distinct from the supercell size:

Listing 3: Training dataset and DFT settings.

```
entity:ds-TiCr2H-c15 a mlips:TrainingDataset ;
  rdfs:label "TiCr2-H DFT dataset (C15, step 1)" ;
  mlips:coversMaterial entity:mat-TiCr2 ;
  mlips:coversProperty mlips:Energy, mlips:Forces, mlips:Stresses ;
  mlips:datasetProvenance mlips:Published ;
  mlips:numConfigurations "1019"^^xsd:integer ;
  mlips:hasDFTCalculation entity:dft-TiCr2H .
entity:mat-TiCr2 a mlips:MaterialSystem ;
  mlips:chemicalFormula "TiCr2" ;
  mlips:materialClass "Laves phase, C15 cubic (Fd-3m)" .
entity:dft-TiCr2H a mlips:DFTCalculation ;
  mlips:hasDFTSettings entity:dft-settings-TiCr2H .
entity:dft-settings-TiCr2H a mlips:DFTSettings ;
  mlips:usedDFTCode mlips:VASP ;
  mlips:xcFunctional mlips:PBE ;
  mlips:pseudopotentialType mlips:PAW ;
  mlips:energyCutoff "400"^^xsd:double ;
  mlips:kPointMesh "4x4x4 Gamma-centered" .
```

Evaluation. Each evaluation produces a BenchmarkResult within a BenchmarkStudy, following the same pattern: the study’s entity:result-01 links the trained model of Listing 1 and the material system of Listing 3 to an AccuracyMetric of type RMSE on energies with value 3.17 meV/atom. The

complete benchmark encoding, alongside the full triple-level encoding of the study, is in the extended version [25].

3. Related Work

We organize the related work into three streams: ontologies for machine learning, ontologies for materials science, and cross-domain metadata templates and FAIR [26, 27]-data consortia (recently extended from data to computational workflows [28], the setting in which MLIP metadata is produced). The MLIPs Ontology is a domain-specific schema; for cross-domain scientific-workflow terminology we refer to Suter et al. [29]. Section 6.3 returns to the ontologies surveyed here: Table 2 assesses, term by term, which of them can answer the competency questions of Section 4.

Ontologies for Machine Learning. ML-Schema [13], from a W3C community group, provides top-level classes for ML algorithms, datasets, and experiments; DMOP [14] supports ontology-driven algorithm selection for data mining; the MEX vocabulary [15] records ML experiment metadata on PROV-O; the Common Metadata Ontology (CMO) [16] integrates pipeline metadata across ML platforms; and the SWeMLS-KG [30] catalogs semantic-web machine-learning systems at scale. These efforts cover generic ML workflows but miss MLIP-specific concepts: algorithm families with physics-specific hyperparameters (cutoff radii, angular/radial basis functions), training-set construction strategies, and DFT reference settings.

Ontologies for Materials Science. The Elementary Multiperspective Material Ontology (EMMO, formerly European Materials Modelling Ontology) [17, 31] provides a top-level ontology grounded in mereocausality theory. The Materials Design Ontology (MDO) [18] is an OWL 2 DL ontology with four modules (Core, Structure, Calculation, Provenance); it has been used for SPARQL-based integrated querying over the Materials Project [32] and the Open Quantum Materials Database (OQMD) [33]. The Platform MaterialDigital core ontology (PMDco) [19] provides a mid-level ontology aligned with the Basic Formal Ontology (BFO) [34]. CMSO and ASMO [20] describe atomic structures and simulation methods, and a recent survey by Norouzi et al. [35] analyzes 60 ontologies in the field. These ontologies cover the physical domain but miss MLIP-training concepts—per-algorithm hyperparameters, dataset construction and DFT settings, cross-paper benchmarks. Concurrent work by Ravari et al. [36] pursues LLM-assisted extraction of DFT workflows aligned to CMSO/ASMO via atomRDF, addressing workflow reproducibility rather than benchmark-comparison knowledge graphs.

Metadata schemas and platforms in computational materials science. Major computational-materials-science platforms define rich metadata schemas for simulation data, often in platform-specific formats but increasingly with Semantic Web surfaces: NOMAD [37]’s MetaInfo normalizes outputs from diverse simulation codes and is progressively aligning with EMMO; the Materials Project [32], AFLOW [38], and OQMD [33] expose custom JSON APIs, standardized for cross-database queries by OPTIMADE [39]; AiiDA [40] tracks full simulation provenance and disseminates through Materials Cloud [41]; and Kadi4Mat [42] manages FAIR research data with recent RDF export (e.g., the file repository in Circular Factory CRC 1574 [43]). Azocár Guzmán et al. [44] demonstrate FAIR RDF workflows using CMSO, ASMO, and PROV-O within pyiron, generalized by the same group [45] to a ~ 750 K-triple knowledge graph. These schemas, platforms, and workflow graphs focus on DFT calculations, material structures, and simulation provenance, but lack descriptions of specific MLIP algorithms and workflows—hyperparameters, training-set construction, and cross-paper benchmarks. Concretely, a user of these platforms today cannot ask: which trained models were fitted to PBE reference data with a 5 Å cutoff, and how do they rank on phonon benchmarks?—the platforms store MLIP runs as opaque files, so selection and comparison queries over method, training-data, and benchmark metadata have nothing to bind to. The MLIPs ontology offers these platforms an MLIP-specific shared vocabulary they can adopt to expose their metadata with explicit, machine-interpretable semantics.

Cross-domain metadata templates and FAIR-data consortia. *Datasheets for Datasets* [46] and *Model Cards* [47] are narrative templates for ML datasets and models; *Croissant* [48] formalizes the dataset side as a JSON-LD schema used by the Hugging Face Hub, Kaggle, and OpenML, with Hugging Face’s model-card metadata [49] as the trained-model counterpart. The MLIPs ontology serves the algorithm- and benchmark-side counterpart of this stack: where *Croissant* and *Datasheets* describe *what is in a dataset*, our training-data module (§5.1) captures *how the dataset was generated* via DFT or wave-function reference calculations, and our benchmark module (§5.1) structures the cross-paper accuracy comparisons that *Model Cards* leave to free-text. At cross-sector level, an MLIP-related catalog entry in the European data-portal standard DCAT-AP [50] is a `Dataset` whose distribution carries our `mlips` predicates in addition to DCAT-AP’s generic ones. Among EMMO-aligned domain ontologies, *CHAMEO* [51] (characterization methods) and *BattINFO* [52] (battery interfaces) model the underlying chemistry; our schema models the ML-driven *interpolation* of that chemistry within the same EMMO-rooted ecosystem (§6.3). At consortium scale, the German *NFDI-MatWerk* [53] consortium operates an ontology registry that materials-science vocabularies (CMSO, ASMO, BattINFO, MDO) are progressively adopting.

4. Requirements

Requirements were gathered through iterative weekly discussions within the author team—twelve materials scientists and MLIP developers and two ontology engineers—in the META-LEARN project [12]. We followed the LOT (Linked Open Terms) methodology [54], whose requirements phase yielded the following competency questions (CQs). The CQs make no claim to exhaustiveness: they delimit the scope of the v0.1.0 release (methods, training data, benchmarks). Several deliberately couple machine-learning aspects with materials aspects (e.g., CQ3 joins training datasets to DFT settings): MLIP metadata inherently couples the two—physics-aware hyperparameters and algorithm-aligned reference settings are precisely the gap the ontology addresses.

- CQ1. Which MLIP algorithms exist, and what hyperparameters does each accept (with types, ranges, and default values)?
- CQ2. Which libraries implement a given algorithm, and what versions are available?
- CQ3. For a given material system (a chemical composition together with its phase or crystal structure), which training datasets are available, and what DFT code and settings were used to generate them?
- CQ4. What is the provenance of a training dataset (published source, in-house calculation, augmented from existing data)?
- CQ5. How many atomic configurations does a training dataset contain, which properties does it cover (energies, forces, stresses), and by which sampling strategy were the configurations drawn?
- CQ6. Which published studies have benchmarked a given algorithm on a given material property, and what accuracy metrics were reported?
- CQ7. For a given material and target property, how do different algorithm and hyperparameter combinations rank by accuracy?
- CQ8. Which simulation types (e.g., molecular dynamics, phonon calculations) have been performed with a given MLIP?
- CQ9. For a given target accuracy, which algorithms and trained models are most efficient—in asymptotic complexity and in measured training and inference cost?

We extended LOT’s intensional CQ-agreement loop with an extensional validation cycle: the seeded knowledge graph (§6.1), encoded from twenty MLIP papers following a 12-question extraction protocol with agentic-AI assistance and a per-paper round-trip check (extended version [25]), was returned to domain experts for review. Reviewing encoded data rather than the schema in isolation surfaced gaps and ambiguities most efficiently; round-trip validation per paper catches schema drift across iterations.

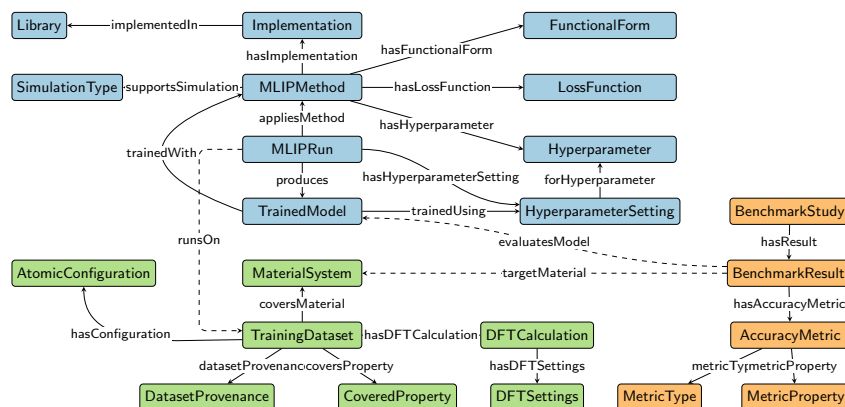


Figure 1: Overview of the MLIPs ontology: three modules and their core classes and relationships. Node fill color encodes module membership (blue: Method; green: Training Data; orange: Benchmark); cross-module relationships are drawn as dashed arrows.

5. The MLIPs Ontology

The MLIPs ontology is an OWL 2 DL ontology organized into three modules that address the competency questions of Section 4; Figure 1 provides an overview. Beyond answering the competency questions, the schema reflects four engineering choices of our own—adopted as Semantic Web good practice rather than elicited from the domain experts: alignment with existing ontologies (§5.2), SHACL validation (§5.2, §7), Linked-Data integration via Wikidata (§7), and extensibility for new algorithm families and hyperparameters (candidate-vocabulary workflow, §7). The modules are a conceptual organization: the v0.1.0 release ships a single OWL file under one namespace, with module membership documented per term; a distribution as three owl:imports-linked files is planned for v0.2.0. The ontology uses the namespace <https://w3id.org/mlips/> (prefix `mlips:`; the external prefixes are listed in footnote 1 on page 3).

5.1. Ontology Modules

Method module. MLIPMethod is the named recipe (e.g., MACE, ACE, HDNNP) for turning an atomic configuration into an energy. It bundles a parametrized functional form (FunctionalForm, via `hasFunctionalForm`), a loss function (LossFunction, via `hasLossFunction`), a training algorithm (an `mls:Algorithm` via `hasTrainingAlgorithm`, e.g., Adam, L-BFGS—the sole alignment point with ML-Schema), and an atomic environment descriptor (AtomicEnvironmentDescriptor) capturing the local-neighborhood representation each method uses (e.g., moment tensors for MTP, SOAP for GAP). Each method declares the Hyperparameter instances it accepts via `hasHyperparameter`, names its Implementations in specific Libraries and versions via `hasImplementation`, and marks the kinds of SimulationType it supports via `supportsSimulation`. An MLIPRun appliesMethod a method, runsOn a TrainingDataset, and produces a TrainedModel, with an optional `hasTrainingRun` to an underlying `mls:Run`. An MLIPRun denotes one concrete execution: training is generally non-deterministic, so repeated runs with identical settings are distinct individuals that may produce distinct TrainedModels. TrainedModel carries shortcut roles trainedWith and trainedOn (property chains through the run). Hyperparameter values are recorded as domain-specific datatype properties (e.g., cutoffRadius, numLayers) with QUDT-annotated units [55]. Efficiency metadata is split: prior-knowledge complexity (e.g., trainingComplexity, inferenceComplexity, supportsGPU) on MLIPMethod; measured cost (e.g., wall-clock, peak memory) on MLIPRun; measured inference cost on the corresponding BenchmarkResult. The module thus addresses CQ1 (methods and their hyperparameters), CQ2 (implementations and versions), CQ8 (supported simulation types), and CQ9 (efficiency).

Training Data module. A `TrainingDataset` holds `AtomicConfiguration` instances over one or more `MaterialSystem` instances. The v0.1.0 scope is deliberately restricted to crystalline systems (`MaterialSystem` $\sqsubseteq$ `cmso:CrystallineMaterial`); extending the ontology to amorphous materials will relax this axiom. `numConfigurations` records the dataset size, and `coversProperty` marks the physical properties included (energy, forces, stresses, virials). Reference data comes from a `ReferenceCalculation`, with two concrete subclasses (`DFTCalculation` and `WaveFunctionCalculation`²) each governed by its own settings type (`DFTSettings`, `WaveFunctionSettings`); method-specific data properties (`xcFunctional`, `kPointMesh` for DFT; `wfMethod`, `basisSet` for wave-function methods) live on each subtype. The DFT-specific predicates `hasDFTCalculation`, `hasDFTSettings`, and `usedDFTCode` are declared `rdfs:subPropertyOf` their family-agnostic counterparts `hasReferenceCalculation`, `hasReferenceSettings`, and `usedReferenceCode`, so legacy data validates and supertype queries pick up DFT-only data via sub-property entailment. `datasetProvenance` classifies the training data as `Published`, `InHouse`, or `Augmented` (with PROV-O’s `prov:wasDerivedFrom` tracking source datasets in the latter case); the optional `samplingStrategy` records how configurations were drawn (e.g., chemical, vibrational, active-learning). The module thus addresses CQ3 (datasets and their reference calculations), CQ4 (provenance), and CQ5 (size, covered properties, and sampling strategy).

Benchmark module. A `BenchmarkStudy` contains one or more `BenchmarkResult` instances, each capturing a single evaluation of a `TrainedModel` on a material system, linked via `evaluatesModel` and `targetMaterial`. Since a trained model already encodes the method, training data, and hyperparameter settings, a result need only reference the model and the test conditions. Results report `AccuracyMetric` instances via `hasAccuracyMetric`, each characterized by a `MetricType` (RMSE, MAE, R^2 , via `metricType`), a `MetricProperty` (energy, force, stress, via `metricProperty`), and a numeric value with unit (via `metricValue`); `reportedIn` links a result to its source publication. The module thus addresses CQ6 and CQ7; the “target property” of CQ7 corresponds to `MetricProperty`, and is distinct from the dataset-side `CoveredProperty` class: the former describes what a benchmark *measures*, the latter what a dataset *contains*. For example, a dataset covering forces (`coversProperty` `mlips:Forces`) may back a model whose benchmark result reports a force RMSE (`metricProperty` `mlips:ForceProperty`). Applied target properties (e.g., hydrogen-adsorption energy) are operationalized through energy or force metrics on specific configurations, not as a first-class concept. Evaluation-protocol details (train/test splits, n -fold cross-validation) are not yet first-class and are a planned extension of this module.

5.2. Axioms, alignment, and meta-categorization

The schema’s formal axioms fall into three categories: *alignment axioms* (22, labelled (L1)–(L22)) connect `mlips:` terms to ML-Schema, PROV-O, MDO, CMSO, schema.org, and QUDT; *meta-categorization axioms* are existential constraints inferred from each class’s `OntoClean` classification; *other axioms*—existential, cardinality, inverse-existence, property chains, domain restrictions—enforce data completeness and consistency at the schema level. The full catalog is published alongside the ontology at <https://w3id.org/mlips/>.

Alignment. The MLIPs ontology reuses established vocabularies wherever an existing class or property already names the concept at hand: ML-Schema for the ML workflow, PROV-O for activity-based provenance, MDO and CMSO for materials science, schema.org for bibliographic metadata, and QUDT for units; the 22 alignment axioms (L1)–(L22) are cataloged, with a graph visualization, in the extended version [25]. The single salient design decision—the disjointness in (L6) between `MLIPMethod` and `mls:Algorithm`, with `hasTrainingAlgorithm` (L17) as the sole connecting role—keeps the domain-level recipe distinct from the generic ML training algorithm it delegates to; the decision is discussed in detail in the extended version [25]. Beyond the 22 cataloged axioms, the schema carries 14 additional

²We include wave-function methods (CCSD(T), MP2, CASPT2) because some MLIPs are now trained on CCSD(T) reference data.

`rdfs:subClassOf / rdfs:subPropertyOf` edges from `mlips:` terms to their closest external parents (e.g., `mlips:Library` $\sqsubseteq$ `schema:SoftwareApplication`), making the alignment queryable rather than narrative-only; enumerative and bespoke constructs (e.g., `XCFunctional`, `LossFunction`) lack a clear external parent and remain unaligned. The alignment axioms were constructed manually and pin the comparator versions they target; the axiom `AtomicConfiguration` $\sqsubseteq$ `cmso:AtomicStructure` in particular is being settled in coordination with the CMSO maintainers, since an atomic configuration bundles a structure with per-configuration properties. A handful of datatype properties still encode default-unit conventions as literals; migrating them to QUDT unit individuals is planned for v0.2.0.

Meta-categorization. We classify every `mlips:` class by its OntoClean meta-properties [22] into one of four named categories (*Sortal*, *Subordinate Sortal*, *Role*, *Reified Relation*) attached as an `mlips:metaSort` annotation. The four values are defined locally (OntoClean’s OWL serialization does not resolve and lacks an entry for reified n -ary relations, the most common non-sortal kind here); they are scoped to this ontology, not offered as a general OntoClean release. The meta-sort layer is also our answer to the top-level-ontology question: rather than committing the whole schema to a foundational ontology such as BFO [34] or EMMO, the three modules share this lightweight OntoClean-derived layer, keeping the ontology cheap for tool emitters to adopt, while the alignment axioms preserve a path to foundational grounding through comparators that do commit to BFO or EMMO (PMDco, CHAMEO). One consequence is formal: each reified-relation class carries an existential axiom requiring its carrier (e.g., `HyperparameterSetting` $\sqsubseteq \exists \text{isSettingOf.MLIPRun}); the full per-class classification and CONSTRUCT rules are in the extended version [25].$

6. Evaluation

We evaluate the MLIPs ontology on a seeded knowledge graph of published MLIP studies: competency-question execution, reasoning and pitfall checks, and a gap analysis against existing ontologies.

6.1. Seeded Knowledge Graph

To exercise the ontology against real data, we encoded 20 published MLIP studies into a seeded knowledge graph. Each paper was encoded manually (with agentic-AI assistance) following a 12-question extraction protocol; every paper-graph passes a SPARQL-based round-trip check against the source, SHACL validation, and reasoner consistency (§6.2), and the encoded data was reviewed by the domain experts among the authors. The full catalog (one subsection per paper, organized around the extraction protocol) and the complete triple-level encoding of the running example (Kumar et al. [24]), of which Section 2 shows the core listings, are in the extended version [25]; both are also archived on DaRUS [56]. The graph contains 3,561 triples across the 20 paper-graph instances and 376 more in the controlled-vocabulary file `mlips-vocab.ttl`; together with the 1,098-triple T-Box, this is a 5,027-triple graph.

Diversity. The corpus deliberately covers multiple axes of variation: method families span the chronological arc of the field (HDNNP from 2007 through MACE-MP-0 in 2024) and the architectural spectrum (descriptor + regressor, deep networks, equivariant message passing, universal/foundation models); reference data is predominantly DFT across six exchange–correlation functionals and four pseudopotential/basis conventions, with one CCSD(T) transfer-learning study (ANI-1ccx) exercising the `WaveFunctionCalculation/WaveFunctionSettings` branch added in this work; materials range from elemental crystals and intermetallics through Laves phases, magnetic oxides, and high-entropy alloys to liquids, organic molecules, and universal coverage of the Materials Project (a per-axis breakdown is in the extended version [25]). Controlled-vocabulary individuals (e.g., `mlips:PBE`, `mlips:PAW`) populate the four open extension-point classes; values not yet in the vocabulary are minted paper-locally with a `candidateForVocabulary` marker and flagged for promotion review.

Table 1

Competency-question execution on the seeded corpus (5,027 triples; single machine, no reasoner; query templates in the extended version [25]).

CQ	Rows	Time (ms)	Note
CQ1	81	117	13 canonical Hyperparameter individuals; paper-local survivors carry a candidateForVocabulary marker.
CQ2	20	33	One implementation per paper.
CQ3	21	33	21 of the 22 datasets carry the full DFT or wave-function settings tuple.
CQ4	22	31	Provenance recorded for every dataset (two papers split their data).
CQ5	135	26	Property/sampling combinations across all datasets.
CQ6	46	37	46 benchmark results across the 20 studies.
CQ7	27	76	RMSE-ranked cross-paper accuracy comparison.
CQ8	38	74	Curated from prose statements (17 studies, 6 simulation types); see text.
CQ9	3	41	Only 5 of 20 papers report inferenceTimePerAtom; the threshold $RMSE < 2$ further trims the result set.

Reportable-vs-reported gaps. A first cross-corpus pass exposes the metadata gap motivating the ontology: 18/20 studies report energy errors and 15/20 report force errors, but only 9/20 report any training- or inference-cost metadata, and 0/20 report peak memory. The ontology can express each of these per paper; the literature reports them only intermittently. Section 6.2 formalizes this through competency-question SPARQL queries.

6.2. Competency Question Execution

All nine competency questions (Section 4) are expressible as SPARQL queries over the ontology, with several spanning all three modules (e.g. joining a benchmark result through its trained model to the underlying method and training dataset). The full query templates are in the extended version [25]. All nine queries return non-empty results. CQ8 initially returned zero rows, reflecting the source literature’s habit of stating supported simulation types only in prose—a metadata gap in the field, not in the schema; curating those prose statements into triples validated the supportsSimulation branch against real data.

Consistency and pitfall checks. HermiT [57] via ROBOT [58], cross-checked with ELK, confirms that the ontology—including the 27 formal axioms and the 22 alignment axioms—is consistent, every class is satisfiable, and the inferred hierarchy contains no unexpected subsumptions. The OOPS! [59] pitfall scanner reports only two *minor* findings, both tracing to external alignment-target classes referenced as `rdfs:subClassOf` parents, which we deliberately do not import; the detailed reasoning and pitfall reports are archived with the ontology distribution on DaRUS [56].

6.3. Comparison with Existing Ontologies

The headline finding of this evaluation is negative and motivates the whole resource: *no existing ontology fully answers a single one of the nine MLIP competency questions, and even the hypothetical union of all of them answers only two* (CQ2 and CQ3). Table 2 details the comparison with existing ontologies from the ML and materials-science domains (the *comparators*). For each comparator we walked the TBox and identified the terms a SPARQL query would need to answer each of the nine competency questions of Section 4: a cell is *answerable* (●) when the comparator declares all required terms, *partially answerable* (◐) when superclass reasoning approximates the question (e.g., a generic Calculation class subsumes our DFTCalculation without DFT-specific data properties), and *not answerable* (—) otherwise. The classification is TBox-level, independent of deployed instance data, and is a *gap analysis* relative to the MLIP metadata needs of Section 4—not an evaluation of the comparators against their own requirements. The per-cell mapping is archived with the ontology distribution. ML-Schema partially covers CQ1–CQ2

Table 2

Competency question coverage across comparator vocabularies (MLS = ML-Schema, Cr = Croissant). • = answerable, ◦ = partially, — = not answerable.

	MLS	MDO	CMSO ASMO	EMMO	PMDco	Cr	MLIPs
CQ1: Algo. hyperparams	◦	—	—	—	—	—	•
CQ2: Implementations	◦	—	—	—	—	—	•
CQ3: Training data + DFT	—	◦	◦	◦	◦	◦	•
CQ4: Dataset provenance	—	—	—	—	—	◦	•
CQ5: Dataset size + coverage	—	—	—	—	—	◦	•
CQ6: Published benchmarks	—	—	—	—	—	—	•
CQ7: Accuracy ranking	—	—	—	—	—	—	•
CQ8: Simulation types	—	—	◦	◦	—	—	•
CQ9: Efficiency metadata	—	—	—	—	—	—	•

through mls:Algorithm and mls:Implementation but lacks physics-specific hyperparameters and units; MDO and CMSO/ASMO partially cover CQ3 but not MLIP-specific provenance or DFT settings at the required granularity; CQ4–CQ7—the benchmark–method–dataset triangle central to the MLIPs ontology—are covered by none. The complementary scope is captured by linking rather than duplication: ML-Schema’s mls:Run remains reachable via hasTrainingRun, and CMSO/ASMO, MDO, EMMO, and PMDco supply crystal-structure detail via MaterialSystem.

Croissant. Croissant [48], the dataset-side counterpart positioned in §3, exhibits the mirror-image profile: it partially covers the dataset-centric CQ3–CQ5 (dataset description, provenance, field-level property coverage) but declares none of the method-, benchmark-, or efficiency-specific terms. Croissant describes the training set as a distributable artifact; the MLIPs ontology describes how that set was produced and how the resulting model performs—confirming the complementarity stated in §3.

Union of comparators. The *union* above—the hypothetical merged ontology of all five comparators, well-defined through their shared PROV-O/schema.org/QUDT alignments—fully answers CQ2 and CQ3; for CQ1 and CQ4–CQ9 it remains ◦ or — because the MLIP-specific specializations (physics-aware hyperparameter individuals, provenance categories, sampling strategies, the benchmark chain, efficiency metadata) are absent from every comparator. The contribution concentrates in those seven CQs.

7. Sustainability and Maintenance

The MLIPs Ontology is developed within the META-LEARN project [12], an ERC Advanced Grant (2026–2030); it is an authoring artefact of the Stuttgart Institute for Artificial Intelligence and the Institute for Materials Science, which jointly commit to maintaining it beyond the project funding period. The project’s own MLIP studies publish their metadata into the MLIPs knowledge graph, so the graph grows with the group’s research output independently of external adoption.

Governance and editorship. The Institute for Artificial Intelligence contributes Semantic-Web expertise and maintains the persistent-IRI and CI infrastructure; the Institute for Materials Science provides the domain expertise that sustains the schema’s MLIP-specific concepts. Three editors have special dedication to the project (Hernández, Jung, Grabowski); issues are triaged in weekly editor meetings and changes land via reviewed pull requests. A FAIRness assessment with FOOPS! [60] accompanies the LOV registration.

Versioning, hosting, and continuous validation. The ontology follows Semantic Versioning [61] (current release 0.1.0), each release depositing an archival snapshot to DaRUS [62]. The persistent IRI <https://w3id.org/mlips>, registered with the W3C Permanent Identifier Community Group, serves the ontology with HTTP content negotiation (XHTML+RDFa, Turtle, RDF/XML); instance IRIs under <https://w3id.org/mlips/entity/> are reserved for dereferencing. Turtle parsing, the round-trip check (bit-equivalence against the source, catching schema drift), OWL reasoning, and the full competency-question battery (§6.2) are automated as build targets and re-run before every release.

Community engagement and adoption. Two pathways drive adoption. *Producers:* MLIP frameworks (e.g., MACE, NequIP, DeePMD-kit) already serialize the run configurations and trained-model metadata that the schema captures, so a parser layer—analogue to NOMAD’s [37] normalizers for DFT codes—can emit MLIPs-Ontology triples at training time, making the ontology self-populating; the 12-question extraction protocol covers *legacy* studies that predate such emitters. The motivation to populate is direct: encoded studies are the ones selection queries return, and the protocol doubles as a completeness checklist (§6.1). *Consumers:* cross-paper questions reduce to single SPARQL queries against the merged graph (e.g., “which MLIPs support phonon calculations with RMSE < 5 meV/atom?”). The schema’s `mlips:sameAsWikidata` links already cover the major DFT and wave-function codes and 18 of 21 encoded `MaterialSystem` instances; we will contribute the schema to the NFDI-MatWerk ontology registry and file interoperability proposals with NOMAD and Materials Cloud [41]. The contribution infrastructure is in place: the development repository is public with an issue tracker, contribution guidelines, and a canonical citation file, and the ontology has been submitted to the LOV registry. Beyond registries, we are in contact with the maintainers of complementary materials-science vocabularies about federating semantic services for materials science, with our service contributing the MLIP metadata.

8. Conclusion

We have presented the MLIPs ontology, an OWL 2 DL ontology whose three modules—Method, Training Data, and Benchmark—fill a gap left by existing ML and materials-science ontologies: the metadata that connects an MLIP method to its training set, DFT reference settings, trained model, and benchmark results. All competency questions reduce to SPARQL queries over the seeded knowledge graph, and the ontology is consistent under HermiT and ELK.

Outlook: AI-driven research. A machine-readable substrate of methods, training data, and benchmarks is a prerequisite for AI-driven research in the MLIP field. The agentic extraction protocol that seeded the knowledge graph already demonstrates machine-in-the-loop population; the same structure lets autonomous agents retrieve prior training runs, identify suitable training sets and hyperparameters, and reproduce or extend a study from its recorded settings. We see the ontology as a step toward MLIP studies whose results ship as queryable data alongside the paper.

Resource Availability Statement. The MLIPs ontology (OWL file, SHACL shapes, controlled vocabulary, and full axiom catalog) is available at <https://w3id.org/mlips> under CC BY 4.0. The seeded knowledge graph is archived on DaRUS (DOI 10.18419/DARUS-5948 [56]); its 12-question extraction protocol, the encoded-paper catalogue, and the fully worked encoding of the running example are appendices of the extended version [25] (arXiv:2607.23219).

Acknowledgements. Funded by the European Union (ERC, META-LEARN, grant agreement No. 101200433).

Declaration on Generative AI

During the preparation of this work, the authors used Claude, ChatGPT, Grammarly, and LanguageTool in order to: drafting content, paraphrase and reword, improve writing style, grammar and spelling check. After using these tools and services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Physical Review Letters* 98 (2007) 146401. doi:10.1103/PhysRevLett.98.146401.
- [2] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, *Physical Review Letters* 104 (2010) 136403. doi:10.1103/PhysRevLett.104.136403.
- [3] W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects, *Physical Review* 140 (1965) A1133–A1138. doi:10.1103/PhysRev.140.A1133.
- [4] L. Talirz, L. M. Ghiringhelli, B. Smit, Trends in atomistic simulation software usage [article v1.0], *Living Journal of Computational Molecular Science* 3 (2021) 1483. doi:10.33011/livecoms.3.1.1483.
- [5] E. Kocer, T. W. Ko, J. Behler, Neural network potentials: A concise overview of methods, *Annual Review of Physical Chemistry* 73 (2022) 163–186. doi:10.1146/annurev-physchem-082720-034254.
- [6] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, G. Csányi, Gaussian process regression for materials and molecules, *Chemical Reviews* 121 (2021) 10073–10141. doi:10.1021/acs.chemrev.1c00022.
- [7] A. V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, *Multiscale Modeling & Simulation* 14 (2016) 1153–1173. doi:10.1137/15M1054183.
- [8] R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, *Physical Review B* 99 (2019) 014104. doi:10.1103/PhysRevB.99.014104.
- [9] I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner, G. Csányi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, in: *Advances in Neural Information Processing Systems*, volume 35, 2022, pp. 11423–11436.
- [10] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, B. Kozinsky, E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, *Nature Communications* 13 (2022) 2453. doi:10.1038/s41467-022-29939-5.
- [11] C. Chen, S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, *Nature Computational Science* 2 (2022) 718–728. doi:10.1038/s43588-022-00349-3.
- [12] B. Grabowski, META-LEARN: Meta-learned machine-learning interatomic potentials for ab initio engineering of chemical and microstructural complexity, ERC Advanced Grant, project ID 101200433. <https://cordis.europa.eu/project/id/101200433>, 2024. Grant agreement No. 101200433.
- [13] G. C. Publio, D. Esteves, A. Lawrynowicz, P. Panov, et al., ML-Schema: Exposing the semantics of machine learning with schemas and ontologies (2018). [arXiv:1807.05351](https://arxiv.org/abs/1807.05351).
- [14] C. M. Keet, A. Lawrynowicz, C. d'Amato, A. Kalousis, P. Nguyen, R. Palma, R. Stevens, M. Hilario, The data mining Optimization ontology, *Web Semantics: Science, Services and Agents on the World Wide Web* 32 (2015) 43–53. doi:10.1016/j.websem.2015.01.001.
- [15] D. Esteves, D. Moussallem, C. B. Neto, T. Soru, et al., MEX vocabulary: A lightweight interchange format for machine learning experiments, in: *Proceedings of the 11th International Conference on Semantic Systems (SEMANTiCS)*, ACM, 2015, pp. 169–176. doi:10.1145/2814864.2814883.
- [16] M. Foltin, et al., Common metadata framework and common metadata ontology for AI pipeline metadata, *Frontiers in Big Data* (2024).

- [17] P. Del Nostro, J. Friis, E. Ghedini, G. Goldbeck, O. Holtz, O. M. Roscioni, F. A. Zaccarini, D. Toti, Elementary multiperspective material ontology: Leveraging perspectives via a showcase of EMMO-based domain and application ontologies, in: Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) – Volume 2: KEOD, SciTePress, 2024, pp. 135–142. doi:10.5220/0012910200003838.
- [18] P. Lambrix, R. Armiento, H. Li, O. Hartig, M. Abd Nikooie Pour, Y. Li, The materials design ontology, Semantic Web 15 (2024) 481–515. doi:10.3233/SW-233340.
- [19] B. Bayerlein, M. Schilling, H. Birkholz, M. Jung, J. Waitelonis, L. Mädler, H. Sack, A perspective on digital knowledge representation in materials science and engineering, Materials & Design 237 (2024) 112603. doi:10.1016/j.matdes.2023.112603.
- [20] S. Menon, atomRDF: A framework for ontology-based atomistic structure generation and querying, 2024. AtomRDF: <https://github.com/pyscal/atomRDF>; CMSO – Computational Material Sample Ontology: <https://github.com/OCDO/cmso>; ASMO – Atomistic Simulation Methods Ontology: <https://github.com/OCDO/asm>.
- [21] W3C OWL Working Group, OWL 2 web ontology language document overview, W3C Recommendation, 2012. URL: <https://www.w3.org/TR/owl2-overview/>.
- [22] N. Guarino, C. A. Welty, An overview of OntoClean, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, International Handbooks on Information Systems, 2nd ed., Springer, 2009, pp. 201–220. doi:10.1007/978-3-540-92673-3_9.
- [23] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, A. V. Shapeev, The MLIP package: moment tensor potentials with MPI and active learning, Machine Learning: Science and Technology 2 (2021) 025002. doi:10.1088/2632-2153/abc9fe.
- [24] P. Kumar, F. Körmann, B. Grabowski, Y. Ikeda, Machine learning potentials for hydrogen absorption in TiCr₂ Laves phases, Acta Materialia 297 (2025) 121319. doi:10.1016/j.actamat.2025.121319.
- [25] D. Hernández, J. H. Jung, Y. Ikeda, Y. Ou, P. Kumar, T. Schächtel, W. Liu, X. Li, X. Zhang, X. Xu, L. Zhu, F. Körmann, S. Staab, B. Grabowski, An ontology for machine learning interatomic potentials (2026). arXiv:2607.23219, extended version, with the extraction protocol, the encoded-paper catalogue, and the worked encoding of the running example in appendices.
- [26] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., The FAIR guiding principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. doi:10.1038/sdata.2016.18.
- [27] M. D. Wilkinson, M. Dumontier, I. Jan Aalbersberg, et al., Addendum: The FAIR guiding principles for scientific data management and stewardship, Scientific Data 6 (2019) 6. doi:10.1038/s41597-019-0009-6.
- [28] S. R. Wilkinson, M. Alogalaa, K. Belhajjame, et al., Applying the FAIR principles to computational workflows, Scientific Data 12 (2025) 328. doi:10.1038/s41597-025-04451-9.
- [29] F. Suter, T. Coleman, I. Altıntaş, R. M. Badia, B. Balis, K. Chard, I. Colonnelli, E. Deelman, P. Di Tommaso, T. Fahringer, C. A. Goble, S. Jha, D. S. Katz, J. Köster, U. Leser, K. Mehta, H. Oliver, J. L. Peterson, G. Pizzi, L. Pottier, R. Sirvent, E. Suchyta, D. Thain, S. R. Wilkinson, J. M. Wozniak, R. Ferreira da Silva, A terminology for scientific workflow systems, Future Generation Computer Systems 174 (2026) 107974. doi:10.1016/j.future.2025.107974.
- [30] F. J. Ekaputra, M. Llugiqi, M. Sabou, A. Ekelhart, H. Paulheim, A. Breit, A. Revenko, L. Waltersdorfer, K. E. Farfar, S. Auer, Describing and organizing semantic web and machine learning systems in the SWeMLS-KG, in: The Semantic Web – 20th International Conference, ESWC 2023, volume 13870 of *Lecture Notes in Computer Science*, Springer, 2023, pp. 372–389. doi:10.1007/978-3-031-33455-9_22.
- [31] M. T. Horsch, S. Chiacchiera, B. Schembera, M. A. Seaton, I. T. Todorov, Reliable and interoperable computational molecular engineering: 2. semantic interoperability based on the European Materials and Modelling Ontology (2020). arXiv:2001.04175.
- [32] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, Commentary: The Materials Project: A materials genome approach to

- accelerating materials innovation, *APL Materials* 1 (2013) 011002. doi:10.1063/1.4812323.
- [33] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials design and discovery with high-throughput density functional theory: The Open Quantum Materials Database (OQMD), *JOM* 65 (2013) 1501–1509. doi:10.1007/s11837-013-0755-4.
 - [34] R. Arp, B. Smith, A. D. Spear, *Building Ontologies with Basic Formal Ontology*, MIT Press, Cambridge, MA, 2015.
 - [35] E. Norouzi, J. Waitelonis, H. Sack, The landscape of ontologies in materials science and engineering: A survey and evaluation (2024). [arXiv:2408.06034](https://arxiv.org/abs/2408.06034), [arXiv:2408.06034](https://arxiv.org/abs/2408.06034).
 - [36] S. Baghaee Ravari, A. Azocár Guzmán, S. Menon, S. Sandfeld, T. Hickel, M. Stricker, Ontology-aligned structuring and reuse of multimodal materials data and workflows towards automatic reproduction (2026). [arXiv:2601.12582](https://arxiv.org/abs/2601.12582).
 - [37] L. M. Ghiringhelli, C. Baldauf, T. Bereau, et al., Shared metadata for data-centric materials science, *Scientific Data* 10 (2023) 626. doi:10.1038/s41597-023-02501-8.
 - [38] S. Curtarolo, W. Setyawan, G. L. W. Hart, et al., AFLOW: An automatic framework for high-throughput materials discovery, *Computational Materials Science* 58 (2012) 218–226. doi:10.1016/j.commatsci.2012.02.005.
 - [39] C. W. Andersen, R. Armiento, E. Blokhin, et al., OPTIMADE, an API for exchanging materials data, *Scientific Data* 8 (2021) 217. doi:10.1038/s41597-021-00974-z.
 - [40] S. P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, et al., AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance, *Scientific Data* 7 (2020) 300. doi:10.1038/s41597-020-00638-4.
 - [41] L. Talirz, S. Kumbhar, E. Passaro, et al., Materials cloud, a platform for open computational science, *Scientific Data* 7 (2020) 299. doi:10.1038/s41597-020-00637-5.
 - [42] N. Brandt, L. Griem, C. Herrmann, E. Schoof, G. Tosato, Y. Zhao, P. Zschumme, M. Selzer, Kadi4Mat: A research data infrastructure for materials science, *Data Science Journal* 20 (2021) 8. doi:10.5334/dsj-2021-008.
 - [43] R. B. Thapa, D. Hernandez, N. Brandt, J.-F. Klein, E. Hoffmann, S. Staab, M. Selzer, G. Lanza, Data infrastructure for the Circular Factory, in: *Proceedings of the 3rd Knowledge Graphs for Sustainability Workshop (KG4S)*, volume 4002 of *CEUR Workshop Proceedings*, 2025.
 - [44] A. Azocár Guzmán, H.-T. Luu, S. Menon, T. Hickel, N. Merkert, S. Sandfeld, Towards knowledge-based atomistic simulation workflows (2026). [arXiv:2602.01358](https://arxiv.org/abs/2602.01358), [arXiv:2602.01358](https://arxiv.org/abs/2602.01358).
 - [45] A. Azocár Guzmán, S. Menon, T. Hickel, S. Sandfeld, Ontology-based knowledge graph infrastructure for interoperable atomistic simulation data (2026). [arXiv:2604.06230](https://arxiv.org/abs/2604.06230).
 - [46] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. Daumé III, K. Crawford, Datasheets for datasets, *Communications of the ACM* 64 (2021) 86–92. doi:10.1145/3458723.
 - [47] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, T. Gebru, Model cards for model reporting, in: *Proc. Conf. on Fairness, Accountability, and Transparency (FAT*)*, 2019, pp. 220–229. doi:10.1145/3287560.3287596.
 - [48] M. Akhtar, O. Benjelloun, C. Conforti, P. Gijsbers, J. Giner-Miguel, N. Jain, M. Kuchnik, Q. Lhoest, P. Marcenac, M. Maskey, P. Mattson, L. Oala, P. Ruysen, R. Shinde, E. Simperl, G. Thomas, S. Tykhonov, J. Vanschoren, S. Vogler, C.-J. Wu, Croissant: A metadata format for ML-ready datasets, in: *NeurIPS Track on Datasets and Benchmarks*, 2024.
 - [49] Hugging Face, Model cards on the Hugging Face hub, Documentation, 2024. URL: <https://huggingface.co/docs/hub/model-cards>.
 - [50] SEMIC, European Commission, DCAT application profile for data portals in Europe, Version 3.0.0, 2024. URL: <https://semiceu.github.io/DCAT-AP/>.
 - [51] EMMC ASBL and OYSTER project, CHAMEO: Characterisation methodology domain ontology, EMMO domain ontology, 2023. URL: <https://emmo-repo.github.io/domain-characterisation-methodology/>.
 - [52] BIG-MAP consortium, BattINFO: Battery interface ontology, EMMO-aligned domain ontology, 2023. URL: <https://big-map.github.io/BattINFO/>.
 - [53] NFDI-MatWerk consortium, NFDI-MatWerk: National research data infrastructure for materials

science and engineering, Consortium website, 2024. URL: <https://nfdi-matwerk.de/>.

- [54] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, R. García-Castro, LOT: An industrial oriented ontology engineering framework, *Engineering Applications of Artificial Intelligence* 111 (2022) 104755. doi:10.1016/j.engappai.2022.104755.
- [55] R. Hodgson, P. J. Keller, J. Hodges, J. Spivak, QUDT: Quantities, units, dimensions and data types ontologies, 2014. <https://www.qudt.org/>.
- [56] D. Hernández, J. H. Jung, B. Grabowski, MLIPs ontology, 2026. URL: <https://doi.org/10.18419/DARUS-5948>, stuttgart Dataverse deposit, version V1.
- [57] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, Z. Wang, HermiT: An OWL 2 reasoner, *Journal of Automated Reasoning* 53 (2014) 245–269. doi:10.1007/s10817-014-9305-1.
- [58] R. C. Jackson, J. P. Balhoff, E. Douglass, N. L. Harris, C. J. Mungall, J. A. Overton, ROBOT: A tool for automating ontology workflows, *BMC Bioinformatics* 20 (2019) 407. doi:10.1186/s12859-019-3002-3.
- [59] M. Poveda-Villalón, A. Gómez-Pérez, M. C. Suárez-Figueroa, OOPS! (OntOlogy pitfall scanner!): An on-line tool for ontology evaluation, in: *International Journal on Semantic Web and Information Systems*, volume 10, 2014, pp. 7–34. doi:10.4018/ijswis.2014040102.
- [60] D. Garijo, O. Corcho, M. Poveda-Villalón, FOOPS!: An ontology pitfall scanner for the FAIR principles, in: *Proceedings of the ISWC 2021 Posters, Demos and Industry Tracks*, volume 2980 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021.
- [61] T. Preston-Werner, Semantic versioning 2.0.0, 2013. URL: <https://semver.org/spec/v2.0.0.html>.
- [62] University of Stuttgart, DaRUS—data repository of the University of Stuttgart, 2024. doi:10.18419/darus, <https://darus.uni-stuttgart.de/>.