Advanced Engineering Materials

ADVANCED ENGINEERING MATERIALS
www.aem-journal.com

RESEARCH ARTICLE OPEN ACCESS

Toward Knowledge-Based Workflows: A Semantic Approach to Atomistic Simulations for Mechanical and Thermodynamic Properties

Abril Azócar Guzmán[1] | Hoang-Thien Luu[2] | Sarath Menon[3] | Tilmann Hickel[4] | Nina Merkert[2] | Stefan Sandfeld[1]

[1]Institute for Advanced Simulations – Materials Data Science and Informatics (IAS-9), Forschungszentrum Jülich GmbH, Jülich, Germany | [2]Institute of Metallurgy – Chair of Computational Material Sciences/Engineering, Technical University of Clausthal, Clausthal-Zellerfeld, Germany | [3]Interdisciplinary Centre for Advanced Materials Simulation (ICAMS), Ruhr-Universität Bochum, Bochum, Germany | [4]Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany

Correspondence: Abril Azócar Guzmán (a.azocar.guzman@fz-juelich.de)

Received: 28 January 2026 | **Revised:** 13 April 2026 | **Accepted:** 16 April 2026

Keywords: atomistic simulations | mechanical properties | molecular dynamics | simulation workflows

ABSTRACT

Mechanical and thermodynamic properties, including the influence of crystal defects, are critical for evaluating materials in engineering applications. Molecular dynamics simulations provide valuable insight into these mechanisms at the atomic scale. However, current practice often relies on fragmented scripts with inconsistent metadata and limited provenance, which hinders reproducibility, interoperability, and reuse. FAIR data principles and workflow-based approaches offer a path to address these limitations. We present reusable atomistic workflows that incorporate metadata annotation aligned with application ontologies, enabling automatic provenance capture and FAIR-compliant data outputs. The workflows cover key mechanical and thermodynamic quantities, including equation of state, elastic tensors, mechanical loading, thermal properties, defect formation energies, and nanoindentation. We demonstrate validation of structure–property relations such as the Hall–Petch effect and show that the workflows can be reused across different interatomic potentials and materials within a coherent semantic framework. The approach provides artificial intelligence (AI)-ready simulation data, supports emerging agentic AI workflows, and establishes a generalizable blueprint for knowledge-based mechanical and thermodynamic simulations.

1 | Introduction

Atomistic modeling provides a foundation for understanding mechanical properties, defect dynamics, and phase-transformation processes in materials. Like other computational studies, it was historically conducted through ad hoc scripts that combined method choices, parameters, and analysis steps in a nonstandardized manner. The shift toward systematic workflows and high-throughput infrastructures, such as the Materials Project [1, 2], AFLOW [3], and NOMAD repository [4], has shown the strong demand for scalable and reproducible computational methodologies. Workflow-based automation has also proven valuable for developing and validating machine-learning interatomic potentials (MLIPs), for example, through frameworks that support data generation, potential fitting, and large-scale validation [5]. As density functional theory (DFT), empirical potentials, and MLIPs continue to expand the scope and fidelity of simulations, the resulting data landscape has grown in both volume and complexity, creating a pressing need for structured and interoperable approaches [6].

Based on these developments, molecular dynamics (MD) techniques become increasingly powerful. Despite significant advances, however, they continue to face challenges related to nonstandardized workflows and data representations, as outlined above. Most studies rely on fragmented, script-based simulations in

which input parameters, analysis steps, and methodological assumptions are kept within monolithic input files. This lack of standardization in metadata and provenance makes it difficult to reproduce simulation protocols across research groups or even within the same project over time. Reuse of workflows is often limited, particularly when switching between modeling approaches that use different software, each having its own conventions and file formats. Moreover, simulation outputs are rarely integrated into broader research data infrastructures, which restricts their accessibility and reuse. API-based initiatives such as OPTIMADE [7] and PLUMED [8] help promote transparency and reproducibility in atomistic simulations data, but do not address the deeper semantic and interoperability challenges discussed here.

Therefore, a transition toward FAIR research data is essential for advancing computational materials science. The FAIR principles require that data be findable, accessible, interoperable, and reusable, which, in turn, demands standardized metadata and machine-readable provenance [9]. Specific to the materials community, FAIR data practices can enable new research directions and require shared, domain-aware metadata [10, 11]. Semantic interoperability is particularly important for enabling cross-study comparison, combining results from different simulation methods, and integrating computational data with experimental measurements. Materials science ontologies have proven effective for annotating metadata while capturing complex descriptions of materials, methods, and workflows in a consistent manner [12, 13]. These semantic representations provide the foundation for knowledge graphs and artificial intelligence (AI) workflows that can search, interpret, and reason over large bodies of simulation data.

Emerging AI technologies increasingly rely on structured knowledge representations to support reasoning and decision making [14]. Materials science knowledge graphs provide these capabilities by encoding relationships among materials, simulation methods, and studied properties in a machine-readable format. This structure allows downstream tasks such as complex querying, verification of structure–property relationships, automated selection of modeling strategies, and rapid exploration of high-dimensional data landscapes. As a result, workflows enriched with semantic metadata are well positioned to serve as the backbone of next-generation AI-assisted materials research.

Although several workflow engines exist for atomistic and in particular DFT simulations [15–17], most lack the semantic rigor or domain specificity needed to support interoperable and reusable workflows. Current standardization efforts largely focus on electronic-structure data and rarely extend to the calculation of mechanical properties, thermodynamic quantities, or crystal defect energetics that are essential for engineering applications. As a result, no unified framework is currently available that semantically captures both mechanical behavior and thermodynamic processes across different simulation methods.

We propose semantically annotated workflow nodes for atomistic simulations of mechanical and thermodynamic properties. In this work, we use the term knowledge-based workflows in the sense of explicit knowledge representation, where predefined simulation workflows are combined with ontology-aligned semantic annotation, provenance capture, and knowledge-graph integration to make workflow outputs machine-readable, interoperable, and queryable. This establishes a foundation for more advanced knowledge-driven capabilities in future developments. The data journey for the proposed framework is schematized in Figure 1. It is specifically intended for calculations of the equation of state, elastic tensor, compression and tensile loading, nanoindentation, thermal expansion coefficient, specific heat, Gibbs free energies, and defect formation energies. We demonstrate the reusability of the workflows across different empirical potentials, including EAM, MLIPs, and universal MLIPs. We also show targeted queries for physical insight, such as the validation of Hall–Petch strengthening law. We therefore provide a generalizable framework that combines reusable property-specific workflow nodes with semantic annotation and knowledge-graph integration, enabling FAIR and AI-ready atomistic data for systematic comparison, validation, and reuse of engineering-relevant mechanical and thermodynamic properties.

2 | Methodology

2.1 | Overall Scientific Workflow

Our methodology follows an overall scientific workflow, as shown in Figure 2. It consists of two computational stages: (i) a simulation workflow for property calculation and (ii) a semantic annotation and knowledge graph creation stage. The simulation workflow comprises the creation or import of an atomic structure, system equilibration, a production MD simulation, and postprocessing. The simulation workflow is composed of independent workflow

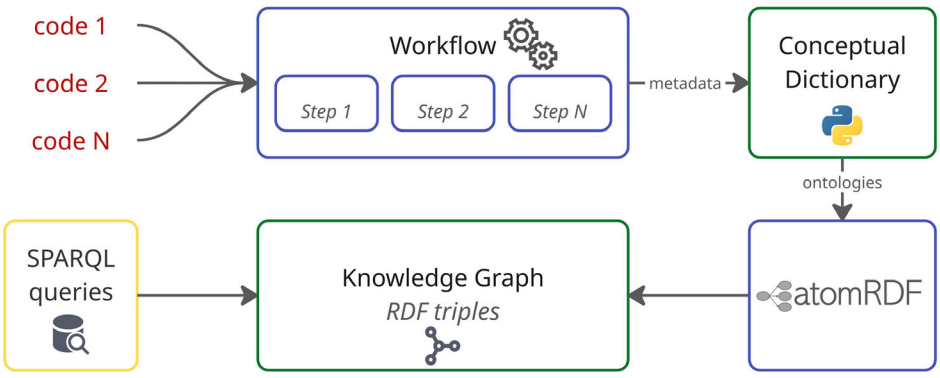

FIGURE 1 | Data journey proposed for atomistic simulations, representing data resulting from atomistic software codes into a knowledge graph.

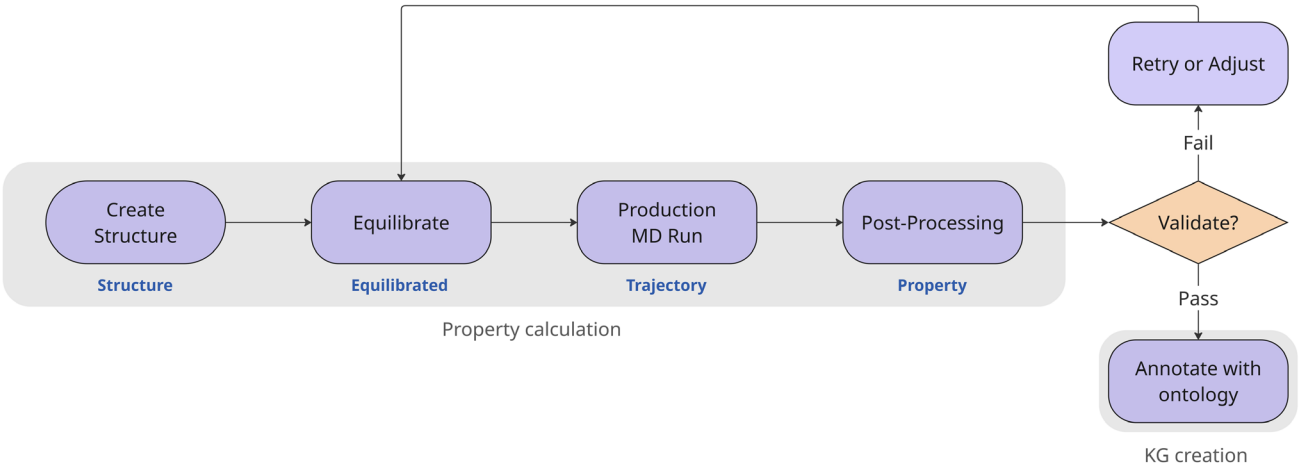

FIGURE 2 | Schematic overview diagram of the overall scientific workflow. The property calculation workflow covers structure creation, MD simulation, and postprocessing. Its outputs are then subjected to a validation step, after which validated results are semantically annotated and integrated into the knowledge graph. The "KG creation" stage is described in Section 2.3, while the "Property calculation" workflows are described in further detail in Section 2.2.

nodes with clearly defined inputs and outputs, as well as an explicit execution order. Between these two computational stages, the resulting data and metadata are subjected to a validation step, in which the user curates which results are suitable for inclusion in the knowledge graph. Validated results are then semantically annotated and integrated into the knowledge graph.

2.1.1 | Validation Step

In the present work, validation refers to a set of quality-control checks applied to the outputs of the simulation workflow before knowledge-graph ingestion. These checks ensure that only physically meaningful and consistent results are stored. Typical validation criteria include convergence of thermodynamic quantities during equilibration, physically reasonable ranges for calculated properties such as elastic constants or energies. The specific validation criteria depend on the type of workflow (i.e., the material property being calculated), the intended use of the data, and the expected fidelity of the underlying model. In the current implementation, this step is primarily performed manually by the user, who curates which results are suitable for inclusion in the knowledge graph, although it can be partially automated depending on the property and use case. A representative set of validation checklists, along with an example of validation integrated directly into the workflow, is provided in the Supporting Information (Section S1).

To ensure complete reproducibility in the spirit of Section 1, for the simulation workflows used for property calculation, we employ a workflow management system that formalizes the simulation process into discrete, traceable units. Each step of the workflow is defined at a coarse-grained level, corresponding to a distinct scientific task, and represented as an independent workflow node with clearly specified inputs and outputs. Many existing systems, such as pyiron [15], jobflow [16], or file-based frameworks like the Common Workflow Language (CWL) [18], adopt a similar structural paradigm.

We use the pyiron software [15], in a graph and node based implementation [19], to construct and execute simulation workflows for property calculation. Our choice is motivated by its ability to directly handle complex Python objects without requiring file-based serialization. Beyond defining the core Python function, the only additional requirement is the use of a decorator, which enables the composition of multiple nodes into a complete and executable workflow.

In addition to the functionality provided by pyiron, we design the individual workflow nodes to be context-aware: each node generates a Python dictionary containing comprehensive scientific metadata for its respective step. These metadata records are subsequently aggregated and, after validation, used to construct a knowledge graph in the KG creation stage, providing a structured and machine-interpretable representation of the workflow and its execution context, as described in more detail in Section 2.3.

An example is shown in Listing 1, which illustrates the complete workflow for calculating the equation of state of a given system using a specified interatomic potential. In this case, the property calculation stage includes generating the initial structure and performing MD simulations with Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) at a range of volumes, with the output of each step passed directly to the next. The metadata from each step is aggregated and then parsed into the knowledge graph in the KG creation stage.

In this work, we provide reusable and reproducible simulation workflows for the calculation of several material properties, including the equation of state, elastic constants (shear modulus, bulk modulus, and Poisson ratio), thermal expansion coefficient, and thermodynamic quantities such as specific heat and free energy, which can be used to determine phase transition temperatures. We also include workflows for mechanical tests under compression and tension. In addition, auxiliary nodes for generating atomic and polycrystalline structures are provided. The scientific methodology behind each of these nodes is briefly described in Section 2.2, along with a discussion of representative results in Section 3.1.

The current implementation uses LAMMPS as the simulation engine. Solver-dependent components (input generation, execution,

```python
from workflows.pyiron.build import bulk
from workflows.pyiron.evcurves import calculate_ev_curves
from pyiron_workflow import Workflow, as_function_node
from conceptual_dictionary import ConceptualDict

pair_style = "eam/alloy"
pair_coeff = "* * workflows/potentials/Fe_Ack.eam Fe"

#ontology-aware metadata aggregation
cd = ConceptualDict()

#property calculation workflow
wf = Workflow('ev1')
wf.structure = bulk('Fe', cubic=True, cdict=cd)
wf.ev_curves = calculate_ev_curves(wf.structure,
                                   pair_style, pair_coeff, cdict=cd)
wf.run()
cd.to_yaml("metadata.yaml")

#kg creation
from atomrdf import WorkflowParser, KnowledgeGraph

kg = KnowledgeGraph()
wf = WorkflowParser(kg)
wf.parse("metadata.yaml")

```

LISTING 1 | Complete workflow for the calculation of equation of state. In the property calculation stage, an atomic structure is created, and the equation of state is calculated. Metadata is aggregated in each step and then added to the knowledge graph.

and parsing) are encapsulated in dedicated workflow nodes, while semantic annotation and provenance capture are handled independently of the underlying engine. This modular separation enables straightforward substitution of the simulation backend; for instance, an equivalent workflow using Quantum ESPRESSO [20, 21] is provided in the Supporting Information (Section S3).

2.2 | Property Calculation Workflows

As discussed in Sections 2.1 and 2.3, we present workflows for predicting material properties from atomistic simulations that are scalable, reproducible, and both technically and semantically interoperable. In the remainder of this section, we briefly describe the scientific background behind the property calculation workflows. The complete set of workflows, including all input parameters, is available in the corresponding repository [22].

All atomistic simulations in this work are performed using the LAMMPS [23]. To ensure reproducibility and comparability across workflows, a consistent set of fundamental configuration parameters is used. We use the `metal` unit system [24], which is standard for metallic materials and provides energies in electronvolts, distances in Å, and times in picoseconds.

In general, the simulation protocols follow a standardized sequence: velocities are initialized from a Gaussian distribution at the target temperature, the system is equilibrated in the isothermal–isobaric ensemble to remove residual stresses and establish thermal equilibrium, and the production simulation is then performed, during which the relevant physical quantities are recorded. All simulations make use of the domain decomposition and parallelization capabilities provided by LAMMPS, enabling efficient execution of workflows on large systems and on high-performance computing clusters.

2.2.1 | Equation of State: Equilibrium Volume and Bulk Modulus

The equation of state, or energy–volume curve, describes the variation in the total energy of a system as a function of its volume near equilibrium. The equation of state gives access to a number of material properties, such as the equilibrium energy and volume, and the bulk modulus. The first step in this workflow is the creation of atomic structure. We first relax the structure and then calculate the energies in a given volume range. We calculate the energy of a given input atomic structure at different volumes and fit the resulting data using the Birch–Murnaghan equation of state [25]. The corresponding simulation workflow is composed of the workflow nodes `structure` (structure generation), `relaxed_structure` (structural relaxation), and `ev_curves` (energy–volume sampling and equation-of-state fitting).

2.2.2 | Elastic Properties: Elastic Constants and Derived Moduli

Elastic constants are fundamental properties of crystalline materials and characterize the relationship between stress and strain in the elastic regime. The elastic tensor of an anisotropic material has 21 components represented as a 6-by-6 matrix C_{ij}. The bulk modulus B quantifies a material's resistance to homogeneous compression, and the shear moduli G_1 and G_2 indicate its resistance to shear deformation in distinct crystallographic directions. Poisson's ratio ν defines the transverse contraction in relation to

the longitudinal extension when subjected to uniaxial stress. In the present work, for cubic crystals, the reported scalar Poisson's ratio is derived using $G_2 = (C_{11} - C_{12})/2$. These quantities can be calculated from the elastic constants following Equation (1). We first create an atomic structure. Then, we calculate these quantities at 0 K by deforming the simulation box and measuring the stress response, following the approach in LAMMPS [26]. The corresponding simulation workflow is composed of the following workflow nodes: `structure` (structure generation) and `elastic` (elastic tensor calculation and derivation of elastic moduli).

$$B = \frac{C_{11} + 2C_{12}}{3}, \qquad G_1 = C_{44}, \qquad G_2 = C' = \frac{C_{11} - C_{12}}{2}, \qquad \nu = \frac{3B - 2G_2}{2(3B + G_2)} = \frac{C_{12}}{C_{11} + C_{12}} \tag{1}$$

2.2.3 | Mechanical Loading: Stress–Strain Response

We provide workflows for two common types of mechanical response tests that can be performed in MD simulations: compression tests and tensile tests. These tests can be applied either uniaxially or hydrostatically and enable the calculation of several mechanical properties, including stiffness, yield stress, flow stress, and indicators of phase transitions or phase stability under pressure. In a first step, the atomic structure is created. This structure is then used to create a polycrystalline structure of given grain size using `atomsk` [27]. This polycrystal is first equilibrated at a high temperature and subsequently at the target, lower, temperature to remove initial stresses and achieve a realistic grain boundary structure. Such a sample is then subjected to hydrostatic compression or uniaxial tension along the z-axis at a specified strain rate. During the simulation, both the stress response and the deformation of the simulation cell are recorded. The corresponding simulation workflow is composed of the following workflow nodes: `struct` (initial structure generation), `poly` (polycrystal generation), and `tensile` (mechanical loading and stress–strain response calculation). In the postprocessing stage, the strain and the flow stress after yielding, measured at 12%–15% strain, are recorded.

2.2.4 | Thermal Properties: Thermal Expansion and Specific Heat

Here, we calculate the thermal expansion coefficient α_P and specific heat c_P, which describe the coupling between temperature and mechanical response of a system. To this end, an atomic structure is created, which is then equilibrated in the isothermal–isobaric ensemble (NPT). After the equilibration stage, an MD simulation in the same ensemble is performed, and the internal energy at each timestep is recorded. From the fluctuations in this ensemble, the specific heat at constant pressure, c_P, can be determined as [28]

$$\langle (\delta(\mathcal{U} + PV))^2 \rangle_{NPT} = k_B T^2 c_P \tag{2}$$

where $\mathcal{U}$ is the internal energy, and P and V are the instantaneous pressure and volume, respectively, at temperature T. k_B denotes the Boltzmann constant. The same fluctuation formalism also provides an estimate of the thermal expansion coefficient, α_P

$$\langle \delta V \, \delta(\mathcal{U} + PV) \rangle_{NPT} = k_B T^2 V \alpha_P \tag{3}$$

The corresponding simulation workflow is composed of the following workflow nodes: `bulk` (unit-cell generation), `structure` (supercell construction), and `thermal` (thermal expansion and specific heat calculation).

2.2.5 | Free Energy Calculations: Gibbs Free Energy and Phase Stability

Gibbs free energies govern phase stability under pressure. We provide workflows for the calculation of both Gibbs and Helmholtz free energies. The free energy calculation starts with the creation of an atomic structure, which is then equilibrated at the given temperature and pressure. The free energy of the crystal at this temperature and pressure is then calculated using nonequilibrium thermodynamic integration methods, followed by the determination of the free energy as a function of pressure. Both methods are employed as implemented in `calphy` [29]. The corresponding simulation workflow is composed of the following workflow nodes: `bulk` (unit-cell generation), `structure` (supercell construction), and `free_energy` (`calculate_free_energy`, free-energy calculation as a function of thermodynamic conditions).

2.2.6 | Nanoindentation: Hardness and Load–Displacement Response

Nanoindentation is a widely used technique for investigating hardness, elastic modulus, and conditions for dislocation bursting, and stress-induced phase transformations not just in experiments but also in MD simulations [30]. The procedure involves moving a nanoindentor to the surface of the material under test and then indenting it to evaluate the material properties. The nanoindentation simulation workflow is designed to mimic experimental measurement protocols to provide insight into the mechanical response of materials at the atomic level.

The different steps in the workflow include creation of atomic structure, preparing substrates with suitable surface geometry and vacuum layer, with subsequent thermal equilibration to remove any residual stresses, followed by a vertical spherical indentor, that indents the material surface with a constant velocity. This enables recording raw force-depth curves alongside time-stamped atomic configurations at specified intervals, which can then be analyzed to compute fitted curves for the extraction of modulus and hardness, as well as to detect pop-in spots and their magnitudes. The corresponding simulation workflow is composed of the following workflow nodes: `structure` (structure generation), `indentation` (nanoindentation simulation), and, where needed for analysis, auxiliary nodes such as `read_final_structure`, `plot_force_depth`, `plot_temperature`, and `plot_centrosymmetry`.

2.2.7 | Defect Energetics: Point-Defect Formation Energies

One important quantity related to crystallographic defects is the formation energy. Here, we present workflows for calculating the formation energy of atoms acting as point defects in a host metal matrix. The workflow begins with the creation of a pristine atomic structure and the calculation of its total energy. Subsequently, a structure with substitutional or interstitial

defects is generated, followed by a second total energy calculation. The defect formation energy is then computed using

$$E_f^X = E_{tot}^{nH+1X} - \left(E_{bulk}^{nH} + E_{ref}^{1X}\right) \quad (4)$$

where E_{tot}^{nH+1X} is the total energy of the configuration containing n host (H) atoms and a single (X) atom occupying either an interstitial or substitutional site, E_{bulk}^{nH} denotes the total energy of the corresponding bulk host metal supercell with the same number of (H) atoms, and E_{ref}^{1X} represents the reference energy of an (X) atom in a reference configuration. The corresponding simulation workflow is composed of the following workflow nodes: `bulk` (bulk structure generation), `structure` (supercell construction), `defect` (`create_vacancy`, `create_substitutional`, or `create_interstitial`, defect creation), and `relax` (relaxation of the defect structure). The final formation energy is then obtained from the relaxed bulk and defect energies using the corresponding formation-energy expression.

2.3 | Semantic Annotation and Knowledge Graph Creation

Compliance with the FAIR principles [9] requires describing data with metadata. Specifically, in terms of interoperability the metadata should use a formal, accessible, shared, and broadly applicable language for knowledge representation, for which the standard choices are RDF or OWL [31, 32]. These formal requirements are addressed at the community level through initiatives such as NFDI-MatWerk [33], whose objective is to support the development and harmonization of semantic standards for interoperable materials research data. The domain-specific semantic artifacts used in this work are developed within this context. Beyond adhering to the FAIR principles, semantic interoperability in materials science presents the benefit of expressing structure–property relationships and in this way injecting physical knowledge to the data itself.

To describe the metadata of the data produced in this article, we use the Computational Materials Sample Ontology (CMSO) [34, 35] and the Atomistic Simulation Methods Ontology (ASMO) [35, 36]. Concepts related to the computational sample, including the material specification, crystal structure, and simulation box, are contained in CMSO. However, ASMO describes the methods used for modeling, in this case MD, the interatomic potential, and the simulation algorithms. The workflow aspects are represented using the provenance ontology PROV-O, a lightweight W3C standard ontology that defines a set of concepts for describing provenance information across different application domains [37]. As shown in Figure 1, the metadata of each workflow run is stored in a conceptual dictionary, a Python dictionary with keys aligned to the necessary ontology classes. In the example shown in Figure 3a, the `ConceptualDict` object aggregates the metadata generated at each step, enabling subsequent semantic integration. The JSON or YAML representation of the dictionary is then serialized into RDF triples using the atomRDF software [35, 38]; a schematic representation is shown in Figure 3b. The main objects are the computational sample and the simulation instance.

A key design principle of the `ConceptualDict` is its independence from any specific workflow engine. It provides a lightweight abstraction for capturing semantically relevant metadata without constraining workflow definition or execution. As a result, the same metadata structure can be populated across different workflow environments, ensuring consistent description of workflow execution and results. Listing 2 shows the same workflow as Listing 1 implemented using jobflow, illustrating the environment-agnostic nature of the semantic annotation approach.

3 | Application Use Cases

As described in Section 2.3, we employ a framework that allows us to produce atomistic simulation data in reproducible workflows while simultaneously annotating the data with ontologies,

(a)

```
{
    "computational_sample": [
        {
            "id": "1PuLyh1",
            "material": {
                "crystal_structure": {
                    "spacegroup_symbol": "Im-3m"}
            },
            "simulation_cell": {
                "volume": {
                    "value": 23.271176375000003},
                "number_of_atoms": 2
            }}],
    "workflow": [
        {
            "algorithm": "EquationOfStateFit",
            "method": "MolecularStatics",
            "input_sample": [
                "1PuLyh1"],
            "output_sample": [
                "iyVpP2K"],
            "calculated_property": [
                {
                    "label": "BulkModulus",
                    "value": 177.62895636765737,
                    "unit": "GigaPA"}],
            "software": {
                "uri": "https://doi.org/10.1016/j.cpc.2021.108171",
                "label": "LAMMPS"},
            "thermodynamic_ensemble": "MicrocanonicalEnsemble"}]
}
```

(b)

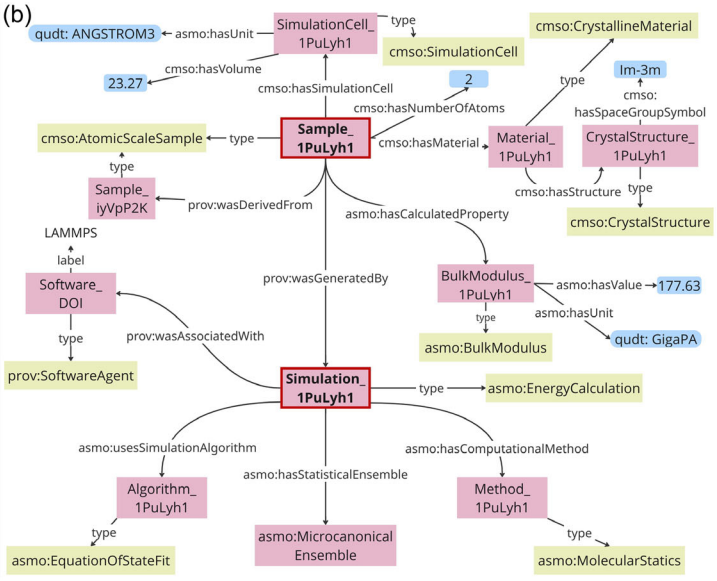

FIGURE 3 | (a) Conceptual dictionary snippet in JSON format. (b) Schematic representation of the RDF serialization of the metadata fields shown in the JSON. The identifiers are shortened for visualization purposes, DOIs or unique identifiers are assigned to all instances.

```python
from workflows.jobflow.build import bulk
from workflows.jobflow.evcurves import calculate_ev_curves
from jobflow import job, run_locally, Flow
from conceptual_dictionary import ConceptualDict

pair_style = "eam/alloy"
pair_coeff = "* * workflows/potentials/Fe_Ack.eam Fe"

#ontology-aware metadata aggregation
cd = ConceptualDict()

#property calculation workflow
structure = bulk(name ='Fe', cubic=True, cdict=cd)
ev_curves = calculate_ev_curves(structure.output,
                                pair_style, pair_coeff, cdict=cd)
flow = Flow([structure, ev_curves])
run_locally(flow)

```

LISTING 2 | Workflow for the calculation of equation of state using jobflow. Only the property calculation stage is shown as KG creation stage is identical to Listing 1.

serializing the metadata as RDF triples. In this section, we showcase the utility of this framework by presenting three use cases: (i) we employ the workflows on a prototypical material system and compare the calculated material properties with existing data; (ii) we show how the workflows can be used to compare materials properties across different models; and (iii) we validate structure–property relationships. These represent routine and important tasks in the computational materials science domain. In use cases (ii) and (iii), we demonstrate in particular how the use of ontologies with workflows can formalize knowledge about the physical meaning of the methods and material systems employed to produce the data. We aim to illustrate the added value of producing FAIR data with semantic descriptions for accelerating the research process in materials science.

3.1 | Demonstration on a Prototypical Material System

In this use case, we apply our workflows on iron as a prototypical system. Iron provides a well-established reference material with multiple stable phases and a broad set of validated interatomic potentials, making it suitable for benchmarking and demonstrating the general applicability of our workflows. The choice of an interatomic potential is critical for accurate representation of material properties. We use an embedded atom method [39] potential (referred to as EAM01 [40]) throughout this work due to its applicability in predicting the properties of iron under pressure, including pressure-induced phase transformations.

For the equation of state (Section 2.2.1), we perform calculations for the body-centered-cubic (bcc), face-centered-cubic (fcc), and hexagonal close-packed (hcp) crystal structures using the corresponding unit cells. The results of the Birch–Murnaghan equation of state, calculated at 0.5% intervals, are shown in Figure 4a. As expected, the bcc structure is identified as the ground state of Fe, while the fcc and hcp structures exhibit very similar energies. The bulk modulus obtained with this potential is 177 GPa, which agrees well with experimental calculated value of 170 GPa [41]. The bulk modulus can also be calculated from the methods described in Section 2.2.2, along with the complete elastic tensor. The tensor elements and derived moduli are summarized in Table 1.

We employ two workflows for calculating the mechanical response of a material system: stress–strain curves (Section 2.2.3) and nanoindentation (Section 2.2.6). We focus on uniaxial tensile tests (Section 2.2.3) and create polycrystalline bcc Fe structures of specified grain size, equilibrated first at a temperature of 600 K and then at the target temperature of 10 K. We then employ uniaxial tension along the z-axis at a strain rate of $5 \times 10^8\ \mathrm{s}^{-1}$. For all grain sizes, the polycrystal contains more than 10 million atoms, and the simulation is carried out for 400 ps. The resulting stress–strain curves for two grain sizes, 25 and 70 Å, are shown in Figure 4b. The curves clearly display the elastic region at low strain, followed by the yield point and the flow stress regime at higher strains. We further perform nanoindentation (Section 2.2.6) following the simulation procedures and hardness calculations adopted from Luu et al., [44]. In Figure 4f, we show the calculated force and hardness with increasing indentation depth for an Fe system of 54 000 atoms. These examples collectively demonstrate that our workflow framework supports simulations on atomistic systems spanning a wide range of sizes, from tens of thousands of atoms in the nanoindentation example to more than 10^7 atoms in the large-scale polycrystalline tensile simulations. The structured metadata captured at each step ensures that the complete simulation setup, including grain-generation parameters, applied loading conditions, and all thermodynamic details, is stored in the knowledge graph. This data can be queried and analyzed, enabling reuse of the generated data.

The specific heat and the coefficient of thermal expansion can be calculated using the workflows in Section 2.2.4. For these calculations, we use a system consisting of 6750 atoms arranged in a bcc lattice for Fe. The simulations are performed at zero pressure at a range of different temperatures. Each simulation is run for 1 ns, and the instantaneous energy and volume are recorded throughout. The calculated c_P values are shown in Figure 4c. Our results show good agreement with experimental observation of 0.45 $\mathrm{Jg}^{-1}\mathrm{K}^{-1}$ [42]. The atomic volume at different

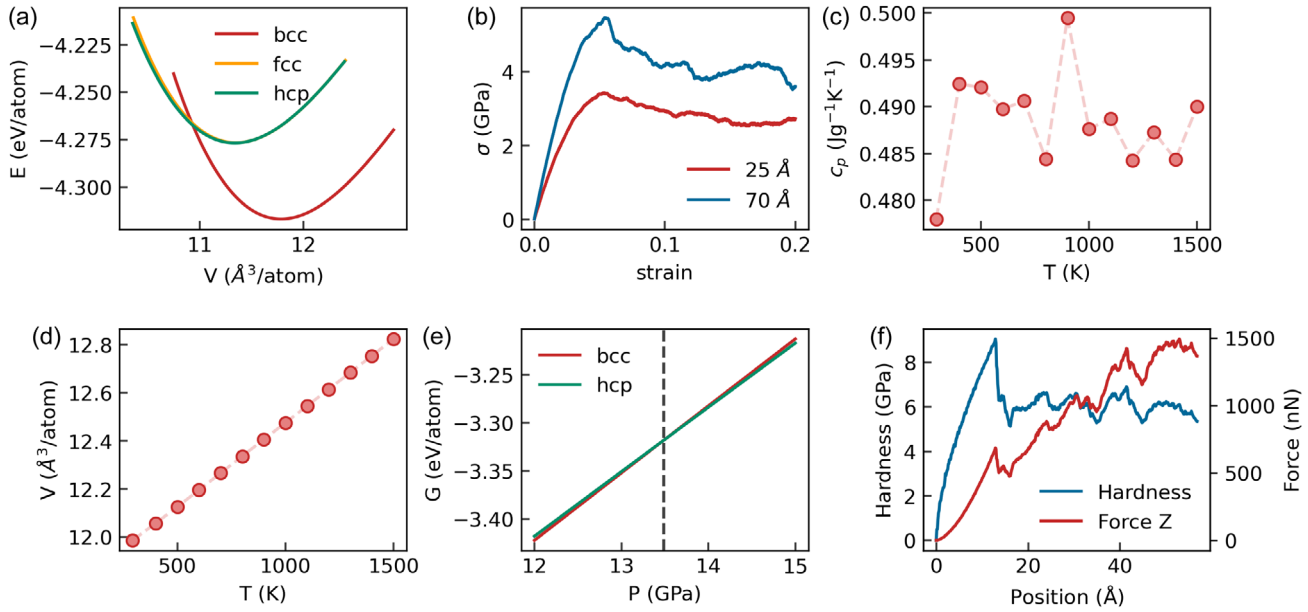

FIGURE 4 | Thermophysical properties of bcc Fe as calculated with the EAM01 potential: (a) Energy–volume curve in comparison with those of different crystal structures in Fe, (b) flow stress with increasing grain size, (c) specific heat, (d) coefficient of thermal expansion, (e) phase transition bcc to hcp as a function of pressure, and (f) hardness and forces from nanoindentation.

temperatures are also obtained, as shown in Figure 4d. The thermal expansion coefficient can be obtained from Equation (3) as well as from the slope in Figure 4d. Using these two approaches, we compute values of $53 \times 10^{-6}\,\mathrm{K}^{-1}$ and $58 \times 10^{-6}\,\mathrm{K}^{-1}$ at 293 K, respectively. Both values exceed the experimentally observed coefficient of $35 \times 10^{-6}\,\mathrm{K}^{-1}$ [43], indicating that the EAM01 potential captures the qualitative temperature dependence but exhibits increased anharmonicity in this regime.

We calculate the free energy of bcc and hcp crystal structures of Fe in the pressure range of 12–15 GPa using the workflows described in Section 2.2.5. We perform the calculations at 100 K and use a system size of approximately 2000 atoms for both crystal structure. The bcc-to-hcp transition pressure is identified from the intersection of the free-energy curves of the separate bcc and hcp phases, yielding a value of approximately 13.5 GPa. This result is consistent with previous theoretical calculations [40] and experimental observations [45].

Finally, we use the workflows described in Section 2.2.7 to investigate the energetics of a carbon impurity in iron. The carbon atom occupies octahedral sites as well as substitutional positions in bcc Fe. The calculated energies for this case are reported in Table 2, with simulation cell sizes of $5 \times 5 \times 5$, resulting in 250 Fe atoms. To calculate the defect formation energies, we use a universal MLIP, GRACE [46], that can describe both Fe and C. The reference value for the C for the calculation of the formation energy is that of diamond (−9.1 eV/atom). The calculated defect formation energies show good agreement with DFT values, with interstitial and substitutional C lying within the expected energetic range reported in previous studies, such as 0.72 eV for the octahedral C in bcc Fe [47].

TABLE 1 | Comparison of selected calculated properties with reference values from the literature.

Property	This work	Reference
Bulk modulus (B, GPa)	177–178	170 [41]
Shear modulus (G_1, GPa)	116	116 [41]
Shear modulus (G_2, GPa)	49	48 [41]
Poisson ratio (ν, –)	0.37	0.3679 [41]
Elastic constant (C_{11}, GPa)	243	239 [41]
Elastic constant (C_{12}, GPa)	145	136 [41]
Elastic constant (C_{44}, GPa)	116	116 [41]
Specific heat (c_P at 293 K, J g^{-1}K^{-1})	0.49	0.45 [42]
Thermal expansion (α at 293 K, K^{-1})	$53–58 \times 10^{-6}$	35×10^{-6} [43]
Pressure bcc → hcp (GPa)	13.5	13.75 [40]

TABLE 2 | Formation energy of C in bcc Fe with the GRACE potential [46], calculated taking C in diamond as a reference.

Case	Value, eV
C octahedral	0.665
C substitutional	2.935

3.2 | Evaluating Interatomic Potentials through Automated Workflow Execution

In MD simulations, the calculated properties often exhibit a strong dependence on the chosen model of atomic interactions, that is, the interatomic potential. Therefore, systematic validation tests are necessary to identify and select suitable potentials for a given material or property of interest.

Here, we demonstrate two key advantages of our workflow framework. First, the workflows are fully reusable and can be easily adapted to different interatomic potentials without modification. Second, because metadata are recorded at every stage of the workflow, all relevant information about the potential, including its type, source, and bibliographic reference, is automatically captured and linked to the generated data.

As a representative example, we compute the bulk modulus and elastic constants for 12 different interatomic potentials. This includes 10 EAM potentials (EAM01–EAM10) [40, 48–56], as well as an atomic cluster expansion (ACE) MLIP [57] and a universal graph atomic cluster expansion (GRACE) potential [46]. A complete list of interatomic potentials used in this work is provided in the Supporting Information (Section S2).

The resulting values are stored in the `ConceptualDict` object and subsequently parsed into the knowledge graph, enabling structured data storage and semantic querying. This approach allows the data to be queried directly using SPARQL. However, formulating SPARQL queries can be a complex task and often requires detailed knowledge of the underlying ontology [58]. Therefore, we use the `tools4RDF` library to assist in the generation of queries based on the ontologies in use [59]. We nevertheless present the SPARQL queries explicitly to improve clarity and to help understand how the ontology predicates are applied. Listing 3 shows an example SPARQL query, based on the CMSO and ASMO ontologies, that retrieves both the bulk modulus value and the corresponding potential used in its calculation. The ontology-based structure ensures that all data are stored in a semantically consistent and machine-interpretable manner.

The software infrastructure we employ, comprising the workflow nodes, the workflow execution environment, the conceptual dictionary, and the knowledge graph creation routines, ensures that identical crystal structures are used across all simulations and that each simulation entry in the knowledge graph retains complete metadata, including details of the interatomic potential and computational parameters. The bulk modulus values resulting from the SPARQL query in Listing 2 are shown in Figure 5a.

A similar SPARQL query for the elastic constant C_{11} provides the corresponding extracted values, presented in Figure 5b. This demonstrates how specific material properties can be programmatically retrieved from the knowledge graph, along with their complete provenance information. Such ontology-based data representation enhances interoperability across different simulation workflows and facilitates the reuse of both data and metadata within the broader materials science research ecosystem.

We show another example for querying defect formation energies computed using the GRACE [46] model. The corresponding SPARQL query is given in Listing 4. This example highlights functionality that becomes available when using our knowledge-based workflows. When a substitutional defect is introduced by replacing an atom, the workflow automatically annotates that the resulting sample contains a substitutional defect. As a result, SPARQL can be used to retrieve all samples that include substitutional impurities. In traditional workflows, where only atomic positions are stored, such information is easily lost and identifying defects could require manual inspection. The SPARQL query returns the defect formation energies together with the chemical species present in the system. The resulting values are shown in Figure 5c and tabulated in Table 3. The calculated substitutional defect formation energies in Fe are in good agreement with the DFT values [60], with deviations of at most 0.2 eV for the Cu, Si, Al, and Mg defects considered here. These examples collectively

```
PREFIX cmso: <http://purls.helmholtz-metadaten.de/cmso/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX asmo: <http://purls.helmholtz-metadaten.de/asmo/>
SELECT DISTINCT ?AtomicScaleSample ?BulkModulusValue ?Reference
WHERE {
    ?AtomicScaleSample asmo:hasCalculatedProperty ?BulkModulus .
    ?BulkModulus asmo:hasValue ?BulkModulusValue .

    ?BulkModulus asmo:wasCalculatedBy ?Simulation .
    ?Simulation asmo:hasInteratomicPotential ?Potential .
    ?Potential cmso:hasReference ?Reference .

   { ?AtomicScaleSample rdf:type cmso:AtomicScaleSample . }
   { ?BulkModulus rdf:type asmo:BulkModulus . }
}
```

LISTING 3 | SPARQL query that retrieves bulk modulus values together with the associated interatomic potentials and bibliographic references.

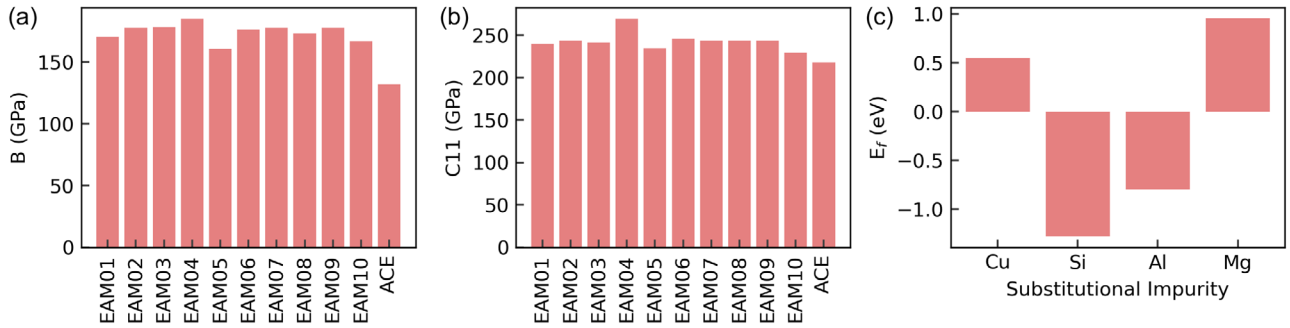

FIGURE 5 | Results of SPARQL queries from the knowledge graph: (a) bulk modulus and (b) elastic constant, c_{11}, with different EAM potentials. (c) Formation energy of substitutional impurities (X = Cu, Si, Al, Mg) in Fe computed using the GRACE [46] model.

```
PREFIX cmso: <http://purls.helmholtz-metadaten.de/cmso/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX asmo: <http://purls.helmholtz-metadaten.de/asmo/>
PREFIX podo: <http://purls.helmholtz-metadaten.de/cdos/podo/>
PREFIX cdco: <http://purls.helmholtz-metadaten.de/cdos/cdco/>
SELECT DISTINCT ?AtomicScaleSample ?value ?label ?elements
        (GROUP_CONCAT(DISTINCT ?symbol; separator=", ") AS ?elements)
WHERE {
    ?AtomicScaleSample asmo:hasCalculatedProperty ?property .
    ?AtomicScaleSample cmso:hasMaterial ?material .
    ?material cdco:hasCrystallographicDefect ?defect .
    ?AtomicScaleSample cmso:hasSpecies ?species .
    ?species cmso:hasElement ?element .
    ?element cmso:hasChemicalSymbol ?symbol .
    ?property rdfs:label ?label .
    ?property asmo:hasValue ?value .

    { ?Defect rdf:type podo:SubstitutionalImpurity . }
    FILTER (?label='DefectFormationEnergy')
}
GROUP BY ?AtomicScaleSample ?value ?label
}
```

LISTING 4 | SPARQL query that retrieves defect formation energies for samples containing substitutional impurities. The query identifies atomic-scale samples, their calculated defect formation energies, and the chemical species present.

TABLE 3 | Defect formation energies in Fe from GRACE and DFT [60].

Host	Defect	GRACE, eV	DFT, eV	ΔE, eV
Fe	Cu	0.6	0.8	0.2
Fe	Si	−1.3	−1.1	0.2
Fe	Al	−0.8	−0.7	0.1
Fe	Mg	1.0	1.0	0.0

demonstrate how our ontology-aware workflows enable systematic comparison of interatomic potentials and provide semantically enriched data suitable for large-scale analysis.

3.3 | Querying Structure–Property Relations

One of the fundamental principles of materials science is to understand structure–property relationships and employ these effectively for materials design. In that aim, the microstructure of the material plays a fundamental role; the presence of crystallographic defects directly affects the physical properties of the material. We propose that semantically annotated data can be leveraged to express such relationships and allow direct querying to validate fundamental physics principles.

Here, we present the example of the Hall–Petch effect, which gives the relation between the grain size and the strength of metallic materials and is described by

$$\sigma_y = \sigma_i + kd^{-1/2} \tag{5}$$

where σ_y is the yield stress (or flow stress at higher strains); σ_i the stress required to move dislocations through a single crystal or very large grain; k the Hall–Petch coefficient, a constant specific to the material; and d is the average grain diameter. The validation of Hall–Petch laws is highly relevant for engineering applications, and similar data-driven analyses have been demonstrated by automatically extracting grain size and yield strength data from the literature [61].

From the data generated using our uniaxial tensile test workflow, described in Section 2.2.3, for polycrystalline Fe using EAM01 potential, we query computational samples with different grain sizes and the corresponding calculated flow stress, as shown in Listing 5.

The resulting values are then plotted in Figure 6, which depicts the average flow stress as a function of the inverse square root of the average grain size, both quantities are automatically annotated at the workflow stage. We see the grain boundary strengthening mechanism according to the Hall–Petch equation. Initially, the stress increases with decreasing grain size, until the strengthening limit after which the stress is reduced due to grain boundary sliding. This softening is known as the inverse Hall–Petch effect. Furthermore, we perform a numerical fit of Equation (3.3) and obtainin a slope $k = 3.3 \pm 0.5\ \text{GPa nm}^{(1/2)}$, $\sigma_0 = 3.0 \pm 0.1$ GPa, and a goodness of fit, $R^2 = 0.836$. The strengthening limit is reached at 11 nm, in good agreement with existing studies [62–64].

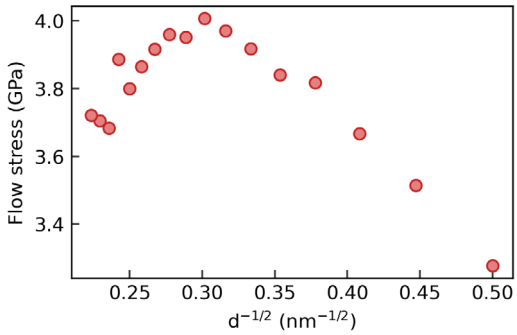

FIGURE 6 | Variation of flow stress with inverse square root of the average grain size.

4 | Discussion

This work illustrates that combining workflow orchestration with semantic metadata enables a knowledge-based approach that significantly enhances the interoperability and interpretability of atomistic simulation data. We provide workflows for mechanical and thermodynamic properties calculations and produce annotated datasets specific to the Fe and Fe–X material systems.

The use cases demonstrate that semantically annotated workflows can be used for data-driven validation of physical laws, such as the Hall–Petch effect, directly from existing simulation data. The framework presented in this work also supports systematic comparison across interatomic potentials, which is essential for uncertainty quantification in atomistic modeling. The workflows intentionally capture more complex systems and simulations such as tensile and compression tests, free energy calculations, nanoindentation, and defect formation energies, showing that semantic workflow design is scalable to mechanical behavior and phase transformation scenarios that are relevant for engineering applications. While the present work focuses on Fe-based systems, the framework is directly applicable to multicomponent alloys, including chemically complex systems such as high-entropy alloys. While the present work focuses on Fe-based systems, the framework is directly applicable to multicomponent alloys, including chemically complex systems such as high-entropy alloys. Extending the approach to more complex molecular or polymer-based systems introduces additional requirements, such as the representation of molecular topology. Addressing these systems requires extending both the workflow nodes and the underlying ontologies, while the overall modular architecture remains unchanged.

Figure 7 highlights that, in the present work, knowledge-based workflows refer to modular workflow nodes enriched with semantic annotation and integrated into a FAIR and queryable knowledge-graph framework, thereby overcoming key limitations of traditional ad hoc simulation practices.

The added value of the proposed infrastructure can be best illustrated considering the implications for interoperability and reusability:

i. Interoperability: In terms of FAIR data, the captured metadata is aligned with application- and domain-level ontologies, providing a common vocabulary to describe the material system, computational sample, simulation workflow, and calculated properties. The use of consistent semantic descriptors enables linking outputs from DFT calculations data with classical MD and MLIP simulations in the same knowledge space. This framework also supports semantic integration with external resources, such as materials databases, and has the potential to interoperate with experimental data. In terms of FAIR workflows, our nodes are agnostic to the specific workflow environment. This is demonstrated in Listings 1 and 2, where two different

```
PREFIX cmso: <http://purls.helmholtz-metadaten.de/cmso/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX asmo: <http://purls.helmholtz-metadaten.de/asmo/>
SELECT DISTINCT ?AtomicScaleSample ?FlowStress ?FlowStressValue ?Unit ?GrainSize
WHERE {
    ?AtomicScaleSample cmso:hasSimulationCell ?SimulationCell .
    ?SimulationCell cmso:hasGrainSize ?GrainSize .
    ?AtomicScaleSample asmo:hasCalculatedProperty ?FlowStress .
    ?FlowStress asmo:hasValue ?FlowStressValue .
    ?FlowStress asmo:hasUnit ?Unit .
   { ?AtomicScaleSample rdf:type cmso:AtomicScaleSample . }
   { ?FlowStress rdf:type asmo:FlowStress . }
}
```

LISTING 5 | SPARQL query to extract average grain size and flow stress from the knowledge graph.

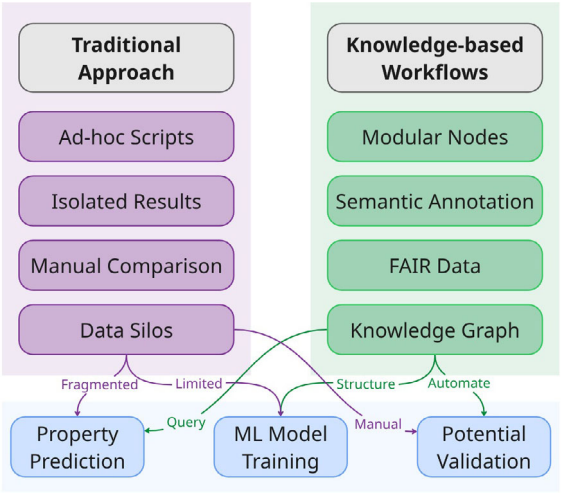

FIGURE 7 | Comparison of traditional ad hoc simulation scripts (left) and the knowledge-based workflows concept adopted in this work (right). Traditional approaches produce isolated results that require manual comparison and lead to data silos. Semantic workflows produce FAIR-compliant data and queryable knowledge graphs that improve interoperability and reuse.

workflow systems, pyiron and jobflow [16], are used. Further connections to other workflow systems are possible, for example, using the Python Workflow Definition [65]. Overall, this shows that the ontology-aware workflow nodes can be seamlessly reused across different workflow systems.

ii. Reusability: The central advantage of the proposed approach for generating simulation workflows and data is the reusability for different tasks. Workflows can be executed with different interatomic potentials without modifying the code, as demonstrated by the systematic evaluation of 10 different potentials in Section 3.2. The semantic layer captures both the physical concepts and the procedural steps of each simulation, allowing the same workflow to be reused across different material systems or thermodynamic conditions by adjusting only the relevant metadata parameters. This capability significantly increases the automation potential for generating comparable datasets for benchmarking or model validation and supports systematic studies of methodological uncertainty in atomistic simulations.

Although the proposed framework improves interoperability and reusability in the production of FAIR data and workflows, there are several limitations that must be acknowledged:

- Workflows for calculation of properties relevant for engineering applications remain strongly application-specific and require domain expertise for correct configuration and interpretation. MD is fundamentally constrained by accessible time scales, which constrains realistic loading rates and transformation kinetics. Consequently, the selection of appropriate input parameters still depends on knowledgeable users, and poor choices directly affect the quality of the results. In addition, sensitivity to boundary conditions, such as cell shape, loading mode, or thermostat and barostat algorithms, continues to affect the comparability of simulation outcomes. The semantic layer can document these choices, but it cannot yet eliminate the underlying physical dependencies.
- The generalizability of workflows to entirely new simulation scenarios is not automatic. Accurate semantic annotation requires domain expertise, and both the ontologies and the workflow templates need to be extended to new physical processes or modeling methods. Nevertheless, the general architecture is expandable to new simulation scenarios. Other approach is that of `semantikon` [66], which offers greater flexibility by allowing serialization with any ontology, although it shifts the responsibility of ensuring consistent and meaningful semantic annotation to the user. While the framework is designed to be extensible, its application to new classes of materials or simulation methods may require additional domain-specific extensions of both the workflows and the underlying ontologies.
- Finally, with respect to integration into broader semantic infrastructures, our highest-level alignment currently relies on PROV-O. While PROV-O provides lightweight and flexible provenance representation, alignment with upper-level ontologies such as the Basic Formal Ontology (BFO) [67] would further enhance interoperability. Existing mappings of PROV-O to BFO indicate that the two are not incompatible [68], but adopting a top-level-aligned design would improve integration within the wider materials science semantic ecosystem.

Future directions include deeper integration with external knowledge graphs and broader research data infrastructures to enable federated search across simulation results, experimental measurements, and literature-derived knowledge. Another opportunity is connecting the semantic workflow framework to emerging agentic AI pipelines that can build on the structured knowledge representation developed here to autonomously select workflows, choose interatomic potentials, propose validation steps, and generate new simulation data. An example of such is `LangSim`, a library that couples large language models with simulation workflows [69]. The semantic layer can enhance transparency and trustworthiness in such AI-driven approaches. In addition, expanding the underlying ontologies to capture microstructure evolution, more complex defect interactions, and advanced thermodynamic descriptors will allow the framework to support a broader range of physical phenomena and materials modeling tasks. Ultimately, semantically annotated workflows not only serve as documentation tools but as a framework to accelerate understanding and generation of insights from data, validation of models and scientific reproducibility.

5 | Conclusion and Outlook

In this work, we introduced knowledge-based workflows for atomistic simulations of mechanical and thermodynamic properties, demonstrating how metadata annotation, ontology alignment, and provenance tracking transform traditional simulation scripts into structured and interoperable research objects. By decomposing simulations into modular workflow components bound to domain and application ontologies, we enabled consistent description of materials, computational samples and methods, and calculated properties across a broad set of atomistic simulation tasks. The

workflows developed for Fe use cases illustrate how this approach improves reusability, facilitates systematic comparison across datasets, and supports transparent interpretation of material properties.

A central outcome of this work is enhanced interoperability and reusability. The semantic layer harmonizes metadata across simulation software and modeling approaches, consistently annotating data from workflows used with different materials systems, potentials, or boundary conditions. This interoperability extends naturally to institutional databases and national research data infrastructures solutions within NFDI-MatWerk, enabling integration with experimental datasets, external simulation repositories, and emerging knowledge graph ecosystems.

The structured metadata and modular workflow design also position the framework for future AI-driven materials discovery. Semantically annotated datasets support targeted query of physical concepts and validation of structure–property relationships, while agentic AI systems can be developed to leverage reusable workflow nodes to select simulation protocols, propose validation tasks, and orchestrate large-scale data generation. Such capabilities provide a pathway toward automated exploration of mechanical properties, high-throughput screening of phase stability, and informed alloy design workflows.

Looking ahead, the framework can be extended to more complex simulation scenarios, e.g., diffusion processes and defect–defect interactions, and multiscale coupling with mesoscale or continuum models. As the library of reusable workflows and materials science ontologies grows, the field can progressively shift from script-based simulations toward knowledge-driven computational materials science. We therefore encourage the community to adopt, refine, and expand the workflows and ontologies introduced here, advancing a shared semantic foundation that supports reproducible, scalable, and AI-ready computational materials science.

Acknowledgments

This work was supported by the consortium NFDI-MatWerk, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure—NFDI 38/1—project number 460247524. H.-T.L. and N.M. acknowledge funding by the DFG (Project-ID 394563137—SFB 1368 and ME 6073/20-1). A.A.G. gratefully acknowledges computing time on the supercomputer JURECA at Forschungszentrum Jülich under grant project ID `defectskg`.

Open Access funding enabled and organized by Projekt DEAL.

Funding

Deutsche Forschungsgemeinschaft (46024752, 394563137).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The workflows are provided here https://github.com/pyscal/semantic-workflows. The repository includes the Python implementations of the workflow nodes for each property calculation. An interactive Binder instance is also available, allowing the workflows to be executed directly in a browser without local installation. The data and corresponding metadata is available here https://doi.org/10.5281/zenodo.18380870.

References

1. A. Jain, S. P. Ong, G. Hautier, et al., "The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation," *APL Materials* 1, no. 1 (2013): 011002.

2. M. K. Horton, P. Huck, R. X. Yang, et al., "Accelerated Data-Driven Materials Science with the Materials Project," *Nature Materials* 24, no. 10 (2025): 1522.

3. S. Curtarolo, W. Setyawan, G. L. Hart, et al., "AFLOW: An Automatic Framework for High-Throughput Materials Discovery," *Computational Materials Science* 58 (2012): 218.

4. M. Scheidgen, L. Himanen, A. N. Ladines, et al., "NOMAD: A Distributed Web-Based Platform for Managing Materials Science Research Data," *Journal of Open Source Software* 8, no. 90 (2023): 5388.

5. S. Menon, Y. Lysogorskiy, A. L. M. Knoll, et al., "From Electrons to Phase Diagrams with Machine Learning Potentials Using Pyiron Based Automated Workflows," *npj Computational Materials* 10, no. 1 (2024): 261.

6. L. Himanen, A. Geurts, A. S. Foster, and P. Rinke, "Data-Driven Materials Science: Status, Challenges, and Perspectives," *Advanced Science* 6, 21 (2019): 1900808.

7. M. L. Evans, J. Bergsma, A. Merkys, et al., "Developments and Applications of the OPTIMADE API for Materials Discovery, Design, and Data Exchange," *Digital Discovery* 3 (2024): 1509.

8. M. Bonomi, G. Bussi, C. Camilloni, et al., "Promoting Transparency and Reproducibility in Enhanced Molecular Simulations," *Nature Methods* 16, no. 8 (2019): 670.

9. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., "The FAIR Guiding Principles for scientific Data Management and Stewardship," *Scientific Data* 3, no. 1 (2016): 160018.

10. M. Scheffler, M. Aeschlimann, M. Albrecht, et al., "FAIR Data Enabling new Horizons for Materials Research," *Nature* 604, no. 7907 (2022): 635.

11. L. M. Ghiringhelli, C. Baldauf, T. Bereau, et al., "Shared Metadata for Data-Centric Materials Science," *Scientific Data* 10, no. 1 (2023): 626.

12. B. Bayerlein, T. Hanke, T. Muth, et al., "A Perspective on Digital Knowledge Representation in Materials Science and Engineering," *Advanced Engineering Materials* 24, no. 6 (2022): 2101176.

13. B. Bayerlein, M. Schilling, P. von Hartrott, and J. Waitelonis, "Semantic Integration of Diverse Data in Materials Science: Assessing Orowan Strengthening," *Scientific Data* 11, no. 1 (2024): 434.

14. A. Ghafarollahi and M. J. Buehler, "SciAgents: Automating Scientific Discovery Through Bioinspired Multi-Agent Intelligent Graph Reasoning," *Advanced Materials* 37, no. 22 (2025): 2413523.

15. J. Janssen, S. Surendralal, Y. Lysogorskiy, et al., "Pyiron: An Integrated Development Environment for Computational Materials Science," *Computational Materials Science* 163 (2019): 24.

16. A. S. Rosen, M. Gallant, J. George, et al., "Jobflow: Computational Workflows Made Simple," *Journal of Open Source Software* 9, no. 93 (2024): 5995.

17. S. P. Huber, S. Zoupanos, M. Uhrin, et al., "AiiDA 1.0, a Scalable Computational Infrastructure for Automated Reproducible Workflows and Data Provenance," *Scientific Data* 7, no. 1 (2020): 300.

18. M. R. Crusoe, S. Abeln, A. Iosup, et al., "The CWL Community," *Communications of the ACM* 65, no. 6 (2022): 54.

19. L. Huber and J. Neugebauer, "Pyiron_Workflow," accessed February 2, 2025, https://github.com/pyiron/pyiron_workflow.

20. P. Giannozzi, S. Baroni, N. Bonini, et al., "QUANTUM ESPRESSO: A Modular and Open-Source Software Project for Quantum Simulations of Materials," *Journal of Physics: Condensed Matter* 21, no. 39 (2009): 395502.

21. P. Giannozzi, O. Andreussi, T. Brumme, et al., Advanced Capabilities for Materials Modelling with Quantum ESPRESSO," *Journal of Physics: Condensed Matter* 29, no. 46 (2017): 465901.

22. "Semantic Workflows Repository," https://github.com/pyscal/semantic-workflows.

23. A. P. Thompson, H. M. Aktulga, R. Berger, et al., "LAMMPS - A Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales," *Computer Physics Communications* 271 (2022): 108171.

24. Lammps Unit Systems, "LAMMPS Documentation," accessed February 15, 2025, https://docs.lammps.org/units.html.

25. F. Birch, "Finite Elastic Strain of Cubic Crystals," *Physical Review* 71 (1947): 809.

26. LAMMPS Developers, "LAMMPS Example: Elastic Constant Calculation," GitHub repository, 2025, accessed February 17, 2025, https://github.com/lammps/lammps/tree/develop/examples/ELASTIC.

27. P. Hirel, "Atomsk: A Tool for Manipulating and Converting Atomic Data Files," *Computer Physics Communications* 197 (2015): 212.

28. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, 2nd ed. (Oxford University Press, 2017).

29. S. Menon, Y. Lysogorskiy, J. Rogal, and R. Drautz, "Automated Free-Energy Calculation from Atomistic Simulations," *Physical Review Materials* 5 (2021): 103801.

30. S. Huang and C. Zhou, *JOM* 69, no. 11 (2017): 2256.

31. W3C RDF Working Group, "Rdf 1.1 Concepts and Abstract Syntax," 2014, https://www.w3.org/TR/rdf11-concepts/, W3C Recommendation.

32. W3C OWL Working Group, "Owl 2 Web Ontology Language Document Overview," 2012, https://www.w3.org/TR/owl2-overview/, W3C Recommendation.

33. C. Eberl, M. Niebel, E. Bitzek, et al., "Consortium Proposal NFDI-Matwerk," 2021, https://doi.org/10.5281/zenodo.5082837.

34. A. A. Guzmán and S. Sandfeld, "Computational Material Sample Ontology (CMSO)," 2024, https://doi.org/10.5281/zenodo.10805536.

35. A. A. Guzman, S. Menon, T. Hickel, and S. Sandfeld, "Ontology-Based Knowledge Graph Infrastructure for Interoperable Atomistic Simulation Data," 2026, https://arxiv.org/abs/2604.06230.

36. A. A. Guzmán and S. Sandfeld, "Atomistic Simulation Methods Ontology (ASMO)," 2024, https://doi.org/10.5281/zenodo.10805591.

37. W3C Provenance Working Group, "Prov-O: The Prov Ontology," 2013, URL https://www.w3.org/TR/prov-o/, W3C Recommendation.

38. S. Menon and A. A. Guzmán, atomRDF, "Python Tool for Ontology-Based Creation, Manipulation, and Quering of Atomic Structures," 2024, https://github.com/pyscal/atomRDF.

39. M. S. Daw and M. I. Baskes, "Embedded-Atom Method: Derivation and Application to Impurities, Surfaces, and Other Defects in Metals," *Physical Review B* 29 (1984): 6443.

40. N. Gunkelmann, E. M. Bringa, K. Kang, G. J. Ackland, C. J. Ruestes, and H. M. Urbassek, "Polycrystalline Iron under Compression: Plasticity and Phase Transitions," *Physical Review B* 86, no. 14 (2012): 144111.

41. J. J. Adams, D. S. Agosta, R. G. Leisure, and H. Ledbetter, "Elastic Constants of Monocrystal Iron from 3 to 500 K," *Journal of Applied Physics* 100, no. 11 (2006): 113530.

42. M. W. Chase, C. A. Davies, J. R. Downey, D. J. Frurip, R. A. McDonald, and A. N. Syverud, *NIST-JANAF Thermochemical Tables, NIST Standard Reference Database 13* ,Version 1.0 (National Institute of Standards and Technology, 1985).

43. Y. M. Kozlovskii and S. V. Stankus, "The Linear Thermal Expansion Coefficient of Iron in the Temperature Range of 130–1180 K," *Journal of Physics: Conference Series* 1382, no. 1 (2019): 012181.

44. H.-T. Luu, S.-L. Dang, T.-V. Hoang, and N. Gunkelmann, "Molecular Dynamics Simulation of Nanoindentation in Al and Fe: On the Influence of System Characteristics," *Applied Surface Science* 551 (2021): 149221.

45. B. J. Jensen, I. Gray, T. G., and R. S. Hixson, "Direct Measurements of the α-ε Transition Stress and Kinetics for Shocked Iron," *Journal of Applied Physics* 105, no. 10 (2009): 103502.

46. Y. Lysogorskiy, A. Bochkarev, and R. Drautz, "Graph Atomic Cluster Expansion for Foundational Machine Learning Interatomic Potentials," *npj Computational Materials* 12, no. 1 (2026): 114.

47. E. Hristova, R. Janisch, R. Drautz, and A. Hartmaier, "Solubility of Carbon in α-Iron under Volumetric Strain and Close to the Σ5(310)[001] Grain Boundary: Comparison of DFT and Empirical Potential Methods," *Computational Materials Science* 50, no. 3 (2011): 1088.

48. Y. Sun, F. Zhang, M. I. Mendelev, R. M. Wentzcovitch, and K.-M. Ho, "Two-Step Nucleation of the Earth's Inner Core," *Proceedings of the National Academy of Sciences* 119, no. 2 (2022): e2113059119.

49. S. Starikov, D. Smirnova, T. Pradhan, et al., "Angular-Dependent Interatomic Potential for Large-Scale Atomistic Simulation of Iron: Development and Comprehensive Comparison with Existing Interatomic Models," *Physical Review Materials* 5 (2021): 063607.

50. L. Proville, D. Rodney, and M.-C. Marinica, "Quantum Effect on Thermally Activated Glide of Dislocations," *Nature Materials* 11, no. 10 (2012): 845.

51. S. Chiesa, P. M. Derlet, S. L. Dudarev, and H. V. Swygenhoven, "Optimization of the Magnetic Potential for α-Fe," *Journal of Physics: Condensed Matter* 23, no. , 20 (2011): 206001.

52. L. Malerba, M. Marinica, N. Anento, et al., "Comparison of Empirical Interatomic Potentials for Iron Applied to Radiation Damage Studies," *Journal of Nuclear Materials* 406, no. 1 (2010): 19, fP6 IP PERFECT Project: Prediction of Irradiation Damage Effects in Reactor Components.

53. P. A. Olsson, "Semi-Empirical Atomistic Study of Point Defect Properties in BCC Transition Metals," *Computational Materials Science* 47, no. 1 (2009): 135.

54. H. Chamati, N. Papanicolaou, Y. Mishin, and D. Papaconstantopoulos, "Embedded-Atom Potential for Fe and its Application to Self-Diffusion on Fe(100)," *Surface Science* 600, no. 9 (2006): 1793.

55. X. W. Zhou, R. A. Johnson, and H. N. G. Wadley, "Misfit-Energy-Increasing Dislocations in Vapor-Deposited CoFe/NiFe Multilayers," *Physical Review B* 69 (2004): 144113.

56. M. I. Mendelev, S. Han, D. J. Srolovitz, G. J. Ackland, D. Y. Sun, and M. Asta, "Development of New Interatomic Potentials Appropriate for Crystalline and Liquid Iron," *Philosophical Magazine* 83, no. 35 (2003): 3977.

57. B. Bienvenu, M. Todorova, J. Neugebauer, et al., "Development of an Atomic Cluster Expansion Potential for Iron and its Oxides," *npj Computational Materials* 11, no. 1 (2025): 81.

58. A.-C. N. Ngomo, L. Bühmann, C. Unger, J. Lehmann, and D. Gerber, "Sorry, I Don't Speak SPARQL: Translating SPARQL Queries into Natural Language," In Proceedings of the 22nd International Conference on World Wide Web (2013).

59. S. Menon, A. A. Guzmán, O. Waseda, S. Sandfeld, and T. Hickel, "Tools4RDF: A Python Toolkit for Working with RDF Data," *Journal of Open Source Software* 11, no. 117 (2026): 9482.

60. B. Jelinek, S. Groh, M. F. Horstemeyer, et al., "Modified Embedded Atom Method Potential for Al, Si, Mg, Cu, and Fe Alloys," *Physical Review B* 85 (2012): 245102.

61. P. Kumar, S. Kabra, and J. M. Cole, "Auto-Generating Databases of Yield Strength and Grain Size using ChemDataExtractor," *Scientific Data* 9, no. 1 (2022): 292.

62. J. B. Jeon, B.-J. Lee, and Y. W. Chang, "Molecular Dynamics Simulation Study of the Effect of Grain Size on the Deformation Behavior of Nanocrystalline Body-Centered Cubic Iron," *Scripta Materialia* 64, no. 6 (2011): 494.

63. Q. Fang, L. Li, J. Li, and H. Wu, "Strengthening Mechanism of Gradient Nanostructured Body-Centred Cubic Iron Film: From Inverse Hall-Petch to Classic Hall-Petch," *Computational Materials Science* 152 (2018): 236.

64. J. Herman, M. Govednik, S. P. Patil, and B. Markert, "Molecular Dynamics Simulation Study of the Mechanical Properties of Nanocrystalline Body-Centered Cubic Iron," *Surfaces* 3, no. 3 (2020): 381.

65. J. Janssen, J. George, J. Geiger, et al., "A Python Workflow Definition for Computational Materials Design," *Digital Discovery* 4 (2025): 3149.

66. Osamu Waseda, "Semantikon," accessed February 14, 2025, https://github.com/pyiron/semantikon.

67. J. N. Otte, J. Beverley, and A. Ruttenberg, "BFO: Basic Formal Ontology1," *Applied Ontology* 17, no. 1 (2022): 43.

68. T. Prudhomme, G. De Colle, A. Liebers, et al., "A Semantic Approach to Mapping the Provenance Ontology to Basic Formal Ontology," *Scientific Data* 12, no. 1 (2025): 282.

69. Y. Zimmermann, A. Bazgir, A. Al-Feghali, et al., "34 Examples of LLM Applications in Materials Science and Chemistry: Towards Automation, Assistants, Agents, and Accelerated Scientific Discovery," *arXiv preprint* (2025), https://arxiv.org/abs/2505.03049.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.