迈向基于知识的工作流：一种用于研究材料力学和热力学性质的原子模拟语义方法

*Abril Az ´ocarGuzm ´an * Hoang - Thien Luu SarathMenon TilmannHickel Nina Merkert Stefan Sandfeld*Dr. A. A. Guzm ´an, Prof. Dr. S. Sandfeld 先进模拟研究所——材料数据科学与信息学（IAS-9），Forschungszen- trum J ¨ulich GmbH，52425 J ¨ulich，德国 * a.azocar.guzman@fz-juelich.de Dr. H.-T. Luu, Prof. Dr. N. Merkert 冶金研究所——计算材料科学与工程教席，克劳斯塔尔工业大学，38678 克劳斯塔尔-采勒费尔德，德国 Dr. S. Menon 先进材料模拟跨学科中心（ICAMS），波鸿鲁尔大学，44801 波鸿，德国。Dr. T. Hickel 联邦材料研究与测试研究所，12489 柏林，德国

arXiv:2602.01358v1 [cond-mat.mtrl-sci] 1 Feb 2026

关键词： 原子模拟、分子动力学、力学性质、模拟工作-流

力学性质和热力学性质（包括晶体缺陷的影响）对于评估工程应用中的材料至关重要。分子动力学模拟为深入了解原子尺度上的这些机制提供了宝贵见解。然而，当前的实践通常依赖彼此割裂的脚本，这些脚本的元数据不一致且数据溯源有限，从而阻碍了可复现性、互操作性和复用。FAIR数据原则和基于工作流的方法为解决这些局限提供了路径。我们提出了可复用的原子模拟工作流，这些工作流融合了与应用本体相一致的元数据标注，能够自动捕获数据溯源并输出符合 FAIR 要求的数据。工作流涵盖关键的力学量和热力学量，包括状态方程、弹性张量、机械加载、热学性质、缺陷形成能和纳米压痕。我们展示了对霍尔-佩奇效应等结构-性能关系的验证，并表明这些工作流能够在统一的语义框架内，跨不同的原子间势和材料复用。该方法提供了适用于人工智能的模拟数据，支持新兴的智能体人工智能工作流，并为基于知识的力学和热力学模拟确立了可推广的蓝图。

1 引言

原子建模为理解材料的力学性质、缺陷动力学和相变过程奠定了基础。与其他计算研究一样，过去这类研究通常借助临时脚本开展，将方法选择、参数和分析步骤以非标准化的方式组合起来。向系统化工作流和高通量基础设施转变的趋势（例如材料项目 [1, 2], AFLOW [3], NOMAD数据库 [4], ）表明，人们对可扩展、可复现的计算方法有着强烈需求。基于工作流的自动化也已证明有助于开发和验证机器学习原子间势（MLIP），例如，某些框架支持数据生成、势函数拟合和大规模验证 [5]。随着密度泛函理论（DFT）、经验势和机器学习原子间势不断拓展模拟的范围并提高其保真度，所产生的数据版图在体积和复杂性方面均不断扩大，因此亟需结构化且可互操作的方法 [6]。

基于这些进展，分子动力学（MD）技术变得日益强大。然而，尽管取得了重大进展，它们仍面临着上述与工作流和数据表示缺乏标准化相关的挑战。大多数研究依赖零散的、基于脚本的模拟，其中输入参数、分析步骤和方法论假设都被保存在单体式输入文件中。元数据和数据溯源缺乏标准化，使得不同研究团队之间，甚至同一项目在不同时期，都难以复现模拟方案。工作流的复用往往受到限制，尤其是在切换

到使用不同软件的建模方法时，每种软件都有其自身的约定和文件格式。此外，模拟输出很少被整合进更广泛的研究数据基础设施，从而限制了其可访问性和复用。OPTIMADE [7]和PLUMED [8] 等基于 API 的倡议有助于提升原子模拟数据的透明度和可复现性，但并未解决此处讨论的更深层次的语义和互操作性挑战。

因此，向FAIR研究数据的转变对于推动计算材料科学至关重要。FAIR 原则要求数据可查找、可访问、可互操作且可重用，这进而需要标准化元数据和机器可读的溯源信息[9]。就材料领域而言，FAIR 数据实践能够开辟新的研究方向，并需要共享的、具备领域感知的元数据 [10, 11]。语义互操作性对于实现跨研究比较、整合不同模拟方法的结果，以及将计算数据与实验测量相结合尤为重要。材料科学本体已被证明能够有效标注元数据，同时以一致的方式刻画材料、方法和工作流的复杂描述 [12, 13]。这些语义表示为知识图谱以及能够检索、解释并对大量模拟数据进行推理的人工智能（AI）工作流奠定了基础。

新兴的人工智能技术日益依赖结构化知识表示来支持推理与决策 [14]。材料科学知识图谱通过以机器可读的格式编码材料、模拟方法与所研究性能之间的关系来提供这些能力。这种结构使复杂查询、结构-性能关系的验证、建模策略的自动选择以及高维数据图景的快速探索等下游任务成为可能。因此，富含语义元数据的工作流有望成为下一代人工智能辅助材料研究的骨干。

尽管已有若干用于原子模拟、尤其是密度泛函理论模拟的工作流引擎 [15,16, 17], ，但其中大多数缺乏支持可互操作且可复用工作流所需的语义严谨性或领域特异性。当前的标准化工作主要聚焦于电子结构数据，很少延伸到对工程应用至关重要的力学性质、热力学量或晶体缺陷能的计算。因此，尚无统一框架能够在不同模拟方法之间从语义上同时刻画力学行为与热力学过程。

我们提出了用于机械和热力学性质原子模拟的语义标注工作流。图1示意了该框架所提出的数据流程。该框架专为计算状态方程、弹性张量、压缩与拉伸载荷、纳米压痕、热膨胀系数、比热、吉布斯自由能及缺陷形成能而设计。我们展示了这些工作流在不同经验势下的可复用性，其中包括嵌入原子法、机器学习原子间势和通用机器学习原子间势。我们还展示了用于获取物理洞见的定向查询，例如验证霍尔-佩奇强化定律。我们提供了一套可推广的蓝图，用于从原子模拟生成 FAIR 且人工智能就绪的数据，以便将其集成到研究数据基础设施和知识图谱中。

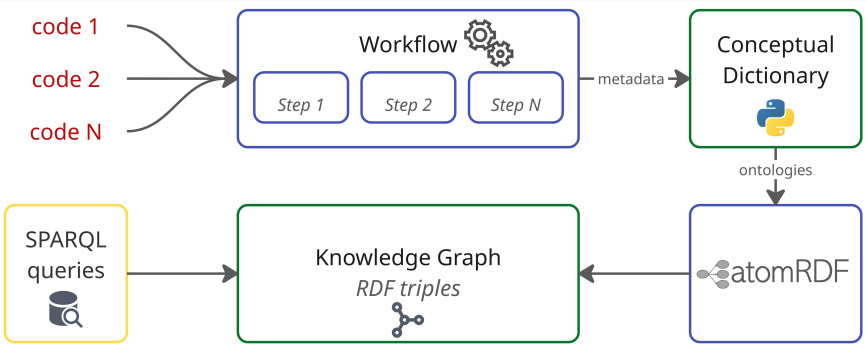

图1：为原子模拟提出的数据流程，将由原子级软件代码产生的数据表示到知识图谱中。

2 方法

2.1 工作流架构

如图2所示，我们的工作流架构遵循一套通用步骤。它包含两个阶段：（i）性质计算阶段和（ii）知识图谱（KG）创建。第一阶段包括创建或导入原子结构、系统平衡、生产分子动力学模拟以及后处理。每个步骤均作为独立的工作流节点实现，并明确规定其输入和输出以及执行顺序。随后进入知识图谱创建阶段；在验证之后，将前一阶段生成的数据和元数据进行语义标注，并整合到知识图谱中。

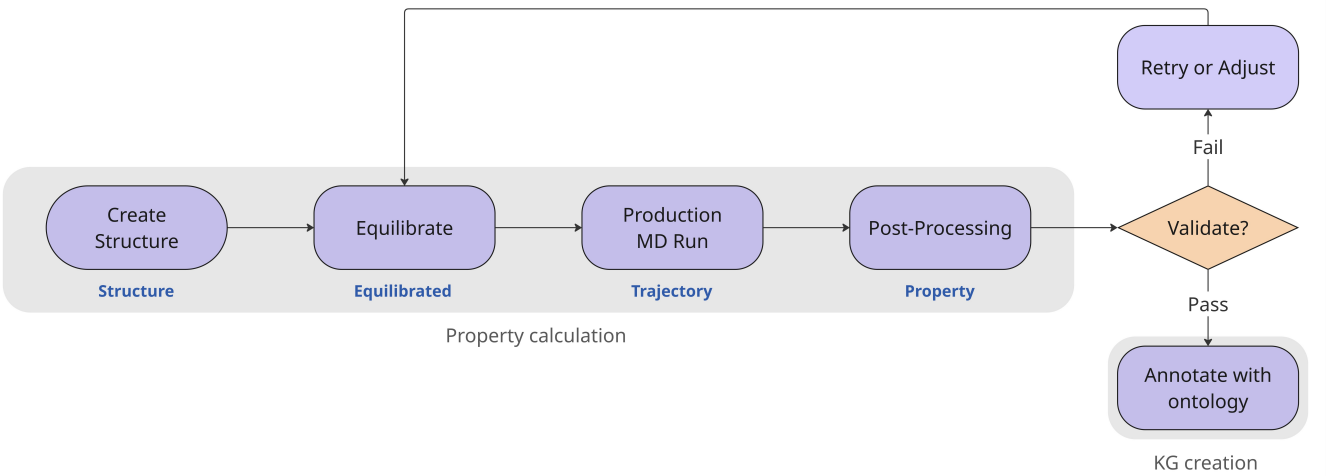

图 2：工作流示意图，展示了所有模拟中从结构创建到验证所采用的一般步骤顺序。成功的结果会用本体进行标注，并添加到知识图谱中。“知识图谱创建”工作流见第 2.2 节，“性质计算”工作流则在第 2.3 节中进一步详细说明。

为秉承第 1 节的理念，确保完全可复现性，对于性质计算工作流，我们采用工作流管理系统，将科学流程形式化为离散且可追踪的单元。工作流中的每一步都在粗粒度层面定义，对应一项独立的科学任务，并表示为一个具有明确输入和输出的独立工作流节点。许多现有系统，如 pyiron [15], jobflow [16], ，以及通用工作流语言（CWL） [18], 等基于文件的框架，都采用了类似的结构范式。

我们使用 pyiron 软件 [15], 采用基于图和节点的实现方式 [19], 来构建和执行性质计算工作流。选择该软件是因为它能够直接处理复杂的 Python 对象，而无需进行基于文件的序列化。除了定义核心 Python 函数外，唯一的额外要求是使用装饰器，以便将多个节点组合成一个完整且可执行的工作流。

除pyiron所提供的功能外，我们将各个工作流节点设计为具备上下文感知能力：每个节点都会生成一个Python字典，其中包含其相应步骤的全面科学元数据。这些元数据记录随后被汇总，并在知识图谱创建阶段用于构建知识图谱，从而提供工作流及其执行上下文的结构化且机器可解释的表示，详见第2.2节。

列表1展示了一个示例，其中说明了如何使用指定的原子间势，计算给定系统的状态方程的完整工作流。在本例中，性质计算阶段包括生成初始结构，并使用 LAMMPS 在一系列体积下进行分子动力学模拟，每一步的输出都会直接传递给下一步。随后汇总每一步的元数据，并在知识图谱创建阶段将其解析到知识图谱中。

```
来自 workflows.pyiron.build导入块体 来自 workflows.pyiron.evcurves导入计算_能量-体积_曲线 来自 pyiron_工作流导入工作流, 作为_函数_节点 来自 概念_字典导入ConceptualDict

    pair_样式 = "eam/alloy"pair_系数 = "* *
    workflows/potentials/Fe_Ack.eam Fe"

    #ontology-aware metadata aggregation
    cd = ConceptualDict()
     #property calculation workflow wf =Workflow('ev1') wf.structure =块体('铁',立方=True,cdict=cd) wf.ev_曲线 =计算_ev_曲线(wf.structure, pair_样式,pair_系数,cdict=cd) wf.run() cd.to_yaml("metadata.yaml")


    #kg creation来自 atomrdf 导入 WorkflowParser,
    KnowledgeGraph

    kg =KnowledgeGraph() 工作流 =WorkflowParser(kg) 工作流。解析("metadata.yaml")
```

列表1：状态方程计算的完整工作流。在性质计算阶段，会创建一个原子结构，并计算状态方程。每一步都会汇总元数据，然后将其添加到知识图谱中。

在本研究中，我们提供了可重复使用且可复现的工作流，用于计算多种材料性质，包括状态方程、弹性常数（剪切模量、体积模量和泊松比）、热膨胀系数，以及比热和自由能等热力学量，这些量可用于确定相变温度。我们还提供了用于压缩和拉伸力学测试的工作流。此外，我们还提供了用于生成原子结构和多晶结构的辅助节点。第2.3节简要介绍了这些节点背后的科学方法论，并在第3.1节讨论了具有代表性的结果。

2.2 语义标注与知识图谱创建

遵循 FAIR 原则 [9] 需要使用元数据描述数据。具体而言，就互操作性而言，元数据应使用一种正式、可访问、共享且广泛适用的知识表示语言，通常选择 RDF 或 OWL[20, 21]。在社区层面，NFDI-MatWerk [22], 等倡议致力于满足这些正式要求，其目标是支持面向可互操作材料研究数据的语义标准的开发与协调。本研究中使用的领域特定语义制品正是在这一背景下开发的。除了遵循 FAIR 原则之外，材料科学中的语义互操作性还具有能够表达结构-性能关系的优势，从而将物理知识注入数据本身。

为了描述本文所生成数据的元数据，我们使用计算材料样本本体（CMSO） [23] 和原子模拟方法本体（ASMO）[24]。计算样本相关的概念，包括材料规范、晶体结构和模拟盒，均包含在 CMSO 中。ASMO 则描述建模所用的方法——在本例中为分子动力学——以及原子间势和模拟算法。工作流相关内容通过溯源本体 PROV-O 表示。PROV-O 是一种轻量级的 W3C 标准本体，定义了一组概念，用于描述不同应用领域中的溯源信息 [25]。如图1所示，每次工作流运行的元数据

都存储在概念词典中。概念词典是一种 Python 字典，其键与所需的本体类相对应。在图3(a)所示的示例中， `ConceptualDict` 对象汇总了每个步骤生成的元数据，从而支持后续的语义集成。随后，字典的 JSON 或 YAML 表示会使用 atomRDF 软件序列化为 RDF 三元组 [26], ，其示意图见图3(b)。主要对象包括计算样本和模拟实例。

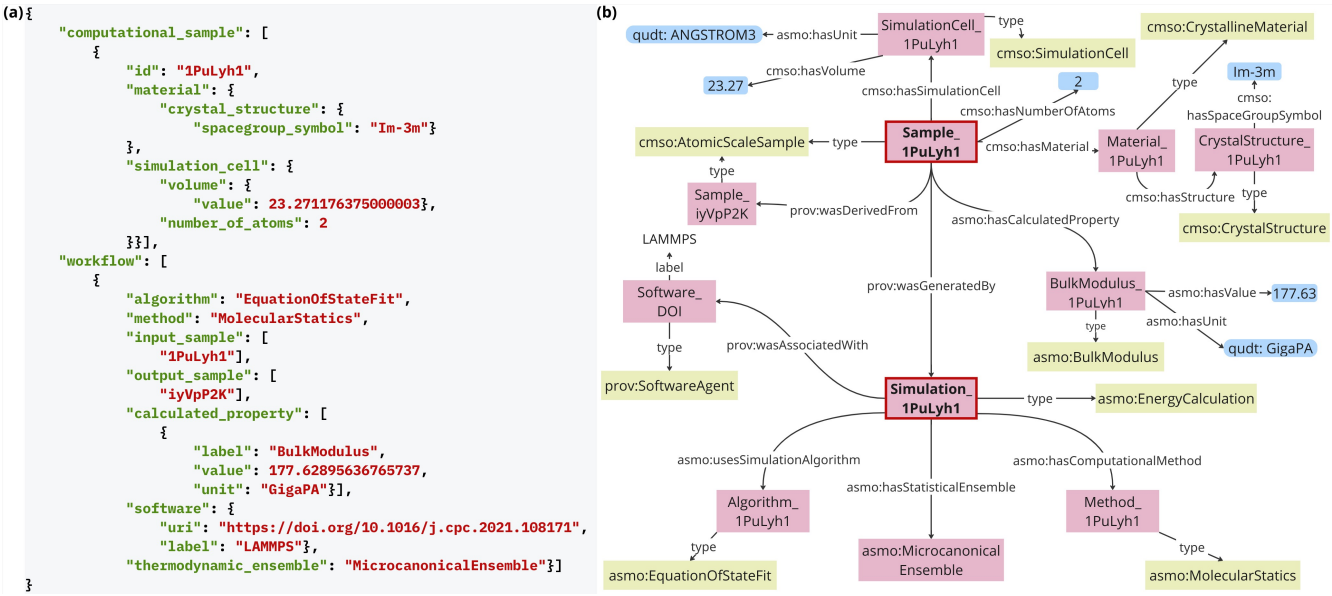

图3：（a）JSON 格式的概念词典片段。（b）JSON 中所示元数据字段的 RDF 序列化示意图。为便于展示，标识符已缩短；所有实例均分配了 DOI 或唯一标识符。

该 `ConceptualDict` 的一项关键设计原则是独立于任何特定工作流引擎。它提供了一种轻量级抽象，用于捕获语义相关的元数据，同时不限制工作流的定义或执行。因此，同一元数据结构可在不同的工作流环境中填充，从而确保对工作流执行过程及结果的描述保持一致。列表2展示了与列表1相同的工作流如何使用 jobflow 实现，体现了这种语义标注方法不依赖特定环境的特性。

```
1 从工作流.jobflow.构建块体 2 从工作流.jobflow.能量体积曲线导入计算_能量体积_曲线 3 从 jobflow 导入作业,本地运行_,流 4 从概念_字典导入
ConceptualDict 5 6 pair_样式 ="EAM/合金" 7 pair_系数 ="**workflows/potentials/铁_Ack.eamFe" 8
9   #ontology-aware metadata aggregation10 cd =ConceptualDict() 11
12  #property calculation workflow13 结构 =块体(名称 ='铁',立方=True,cdict=cd) 14 能量体积_曲线 =计
算_能量体积_曲线(结构.output, 15 pair_样式,pair_系数,cdict=cd) 16 流 = Flow( [结构, 能量体积_曲线]) 17 本地运行_(flow) 18
```

列表2：使用 jobflow 计算状态方程的工作流。这里只展示性质计算阶段，因为知识图谱创建阶段与列表1相同。

2.3 性质计算工作流

如第2.1节和第2.2节所述，我们提出了用于通过原子模拟预测材料性质的工作流，这些工作流具有可扩展性、可复现性，并且在技术和语义层面均可互操作。本节余下部分将简要介绍这些性质计算工作流背后的科学背景。完整的工作流集合（包括所有输入参数）可在相应的代码仓库中找到 [27]。

本研究中的所有原子模拟均使用大规模原子/分子大规模并行模拟器（LAMMPS） [28]完成。为确保不同工作流之间的可复现性和可比性，我们采用一套统一的基本构型参数。我们使用 `metal`单位制 [29], ，该单位制是金属材料的标准单位制，能量以电子伏特为单位，距离以 ˚A为单位，时间以皮秒为单位。

通常，模拟方案遵循标准化流程：首先根据目标温度从高斯分布中初始化速度；随后在等温等压系综中使系统达到平衡，以消除残余应力并建立热平衡；然后进行生产模拟，并在此过程中记录相关物理量。所有模拟均利用 LAMMPS 提供的区域分解和并行化功能，从而能够在大型系统和高性能计算集群上高效执行工作流。

2.3.1 状态方程：平衡体积和体积模量

状态方程（即能量-体积曲线）描述了系统在平衡点附近的总能量随体积的变化。状态方程可用于获取多种材料性质，例如平衡能量、平衡体积和体积模量。此工作流的第一步是创建原子结构。我们首先对结构进行弛豫，然后计算给定体积范围内的能量。我们计算输入原子结构在不同体积下的能量，并使用 Birch–Murnaghan状态方程 [30]拟合所得数据。

2.3.2 弹性性质：弹性常数和导出模量

弹性常数是晶体材料的基本性质，用于表征弹性区间内应力与应变之间的关系。各向异性材料的弹性张量包含 21 个分量，可表示为一个 6×6 矩阵 C_{ij}。体积模量 B 量化了材料抵抗均匀压缩的能力，而剪切模量 G_1 和 G_2 则表征材料在不同晶体学方向上抵抗剪切变形的能力。泊松比 ν 定义了材料在承受单轴应力时，横向收缩与纵向伸长之间的关系。根据公式 1，可以利用弹性常数计算这些量。我们首先创建原子结构。随后，按照 LAMMPS [31]中的方法，通过变形模拟盒并测量应力响应，计算 0 K 下的这些量。

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

2.3.3 机械加载：应力-应变响应

我们提供了两种可在分子动力学模拟中进行的常见力学响应测试的工作流：压缩测试和拉伸测试。这些测试既可采用单轴方式，也可采用静水压方式进行，并可用于计算多种力学性质，包括刚度、屈服应力、流动应力，以及相变或压力下相稳定性的相关指标。首先，创建原子结构。随后使用 `atomsk` [32]，根据给定的晶粒尺寸创建多晶结构。该多晶体先在高温下进行平衡，随后在目标较低温度下进行平衡，以消除初始应力并获得真实的

晶界结构。随后，使该样品以指定的应变率沿 z 轴进行静水压缩或单轴拉伸。在模拟过程中，同时记录应力响应和模拟晶胞的变形。

2.3.4 热学性质：热膨胀与比热

在此，我们计算热膨胀系数 α_P 和比热 c_P，用于描述温度与系统力学响应之间的耦合。为此，我们构建一个原子结构，并在等温等压系综（NPT）中使其达到平衡。平衡阶段结束后，在同一系综中进行分子动力学模拟，并记录每个时间步的内能。根据该系综中的涨落，可以确定恒压比热 c_P，其表达式为 [33]

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

其中，$\mathcal{U}$为内能，P和V分别为温度为 T时的瞬时压力和体积。k_B 表示玻尔兹曼常数。相同的涨落形式也可用于估算热膨胀系数 α_P

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

2.3.5 自由能计算：吉布斯自由能与相稳定性

吉布斯自由能决定压力下的相稳定性。我们提供用于计算吉布斯自由能和亥姆霍兹自由能的工作流。自由能计算从创建原子结构开始，随后使其在给定的温度和压力下达到平衡。然后，使用非平衡热力学积分方法计算晶体在该温度和压力下的自由能，并进一步确定自由能随压力的变化。两种方法均按照 `calphy` [34] 中的实现进行。

2.3.6 纳米压痕：硬度和载荷-位移响应

纳米压痕是一种广泛使用的技术，不仅可用于实验，也可用于分子动力学模拟 [35]，以研究硬度、弹性模量、位错突发条件以及应力诱导相变。该过程包括将纳米压头移至待测材料表面，然后压入材料，以评估其材料性质。纳米压痕模拟工作流旨在模拟实验测量流程，从而深入了解材料在原子尺度上的力学响应。

工作流的不同步骤包括创建原子结构、制备具有合适表面几何形状和真空层的基底，随后进行热平衡以消除任何残余应力，再使用垂直的球形压头以恒定速度压入材料表面。这样便可按指定间隔记录原始力-深度曲线以及带时间戳的原子构型，之后可对这些数据进行分析，拟合曲线以提取弹性模量和硬度，并检测突跳点及其幅度。

2.3.7 缺陷能量学：点缺陷形成能

与晶体学缺陷相关的一个重要量是形成能。这里，我们介绍用于计算基体金属中作为点缺陷的原子的形成能的工作流。该工作流首先创建完整原子结构并计算其总能量。随后，生成含有置换缺陷或间隙缺陷的结构，再次计算总能量。然后，缺陷形成能通过以下公式计算

$$E_f^X = E_{tot}^{nH+1X} - (E_{bulk}^{nH} + E_{ref}^{1X}), \tag{4}$$

其中，E_{tot}^{nH+1X}是含有 n 个基体（H）原子和一个（X）原子的构型的总能量，该（X）原子占据间隙点或替位点； E_{bulk}^{nH} 表示

具有相同数量（H）原子的相应块体基体金属超胞的总能量，而 E_{ref}^{1X} 表示参考构型中一个（X）原子的参考能量。

3 应用案例

如第 2.2 节所述，我们采用一个框架，能够在可复现的工作流中生成原子尺度模拟数据，同时用本体对数据进行标注，并将元数据序列化为 RDF 三元组。在本节中，我们通过三个应用案例展示该框架的实用性：(i) 我们将工作流应用于一个典型材料体系，并将计算得到的材料性质与现有数据进行比较；(ii) 我们展示如何利用工作流比较不同模型下的材料性质；(iii) 我们验证结构-性能关系。这些是计算材料科学领域中常规且重要的任务。在应用案例 (ii) 和 (iii) 中，我们特别展示了将本体与工作流结合使用，如何能够形式化关于用于生成数据的方法和材料体系之物理含义的知识。我们旨在说明，生成带有语义描述的 FAIR 数据对于加速材料科学研究进程所具有的附加价值。

3.1 典型材料体系示例

在本用例中，我们将工作流应用于铁这一典型材料体系。铁是一种成熟的参考材料，具有多个稳定相以及一系列经过验证的原子间势，因此适合用于基准测试并展示我们工作流的普适适用性。原子间势的选择对于准确表征材料性质至关重要。鉴于嵌入原子法 [36] 势函数（以下简称 EAM01 [37]）适用于预测压力下铁的性质，包括压力诱导的相变，我们在全文中均采用该势函数。

对于状态方程（第 2.3.1 节），我们使用相应的晶胞，对体心立方（bcc）、面心立方（fcc）和六方密堆积（hcp）晶体结构进行计算。以 0.5% 的间隔计算得到的 Birch-Murnaghan 状态方程结果如图 4（a）所示。正如预期，bcc 结构被确定为铁的基态，而 fcc 和 hcp 结构表现出非常相近的能量。使用该势函数得到的体积模量为 177 GPa，与实验计算值 166 GPa [38]符合得很好。还可以根据第 2.3.2 节所述的方法计算体积模量以及完整的弹性张量。张量分量和导出模量汇总于表 1。

表 1：使用嵌入原子法势计算的体心立方铁的弹性常数及其导出性质。

量	数值（GPa）
$C_{11} = C_{22} = C_{33}$	243
$C_{12} = C_{13} = C_{23}$	145
$C_{44} = C_{55} = C_{66}$	116
$C' = \frac{1}{2}(C_{11} - C_{12})$	49
体积模量 B	178
剪切模量 G_1	116
剪切模量 G_2	49
泊松比 ν	0.37*

*泊松比是无量纲的。

我们采用两种工作流来计算材料体系的力学响应：应力-应变曲线（第 2.3.3 节）和纳米压痕（第 2.3.6 节）。我们重点研究单轴拉伸测试（第 2.3.3 节），并创建具有指定晶粒尺寸的多晶体心立方铁结构，先在 600 K 下平衡，再在 10 K 的目标温度下平衡。随后，我们以 $5 \times 10^{-8}\,\mathrm{s}^{-1}$ 的应变率沿 z 轴进行单轴拉伸。该多晶体包含超过一千万个原子，

模拟持续 300 ps。两种晶粒尺寸（25 和 70 °A）对应的应力-应变曲线如图 4(b) 所示。曲线清晰地显示，在低应变下为弹性区，随后在较高应变下出现屈服点和流动应力区间。我们还参考 Luu 等人采用的模拟流程和硬度计算方法 [39]，进一步进行了纳米压痕（第 2.3.6 节）。图 4(f) 展示了一个含有 54000 个原子的铁体系中，随着压痕深度增加而计算得到的力和硬度。这些示例也表明，我们的工作流框架支持对规模极大的原子系统进行模拟，涵盖从数千到数百万个原子。每一步所记录的结构化元数据确保完整的模拟设置——包括晶粒生成参数、施加的加载条件以及所有热力学细节——都存储在知识图谱中。这些数据可以查询和分析，从而能够重复利用生成的数据。

比热和热膨胀系数可以使用第 2.3.4 节中的工作流计算。对于这些计算，我们使用由 6750 个原子组成的体系，这些原子在铁的体心立方晶格中排列。模拟在一系列不同温度下、零压力条件下进行。每次模拟运行 1 ns，并记录整个过程中的瞬时能量和体积。计算得到的 c_P 数值如图 4(c) 所示。我们的结果与实验观察到的 0.45 $Jg^{-1}K^{-1}$ [40] 符合良好。不同温度下的原子体积也已计算得出，如图 4(d) 所示。热膨胀系数既可由公式 3 得出，也可根据图 4(d) 中的斜率计算。通过这两种方法，我们分别计算得到 293 K 下的数值为 $53 \times 10^{-6}\,\mathrm{K}^{-1}$ 和$58 \times 10^{-6}\,\mathrm{K}^{-1}$。这两个数值均高于实验观测到的系数 $35 \times 10^{-6}\,\mathrm{K}^{-1}$ [41]，表明 EAM01 势能够捕捉温度依赖性的定性趋势，但在此温度范围内表现出更强的非谐性。

我们采用第 2.3.5 节所述的工作流，计算了 12–15 GPa 压力范围内铁的体心立方和六方密堆积晶体结构的自由能。计算在 100 K 下进行，两种晶体结构的系统规模均约为 2000 个原子。我们观察到，压力约为 13.5 GPa 时，体心立方结构会转变为六方密堆积结构。这一结果与先前的理论计算 [37] 和实验观察 [42]一致。

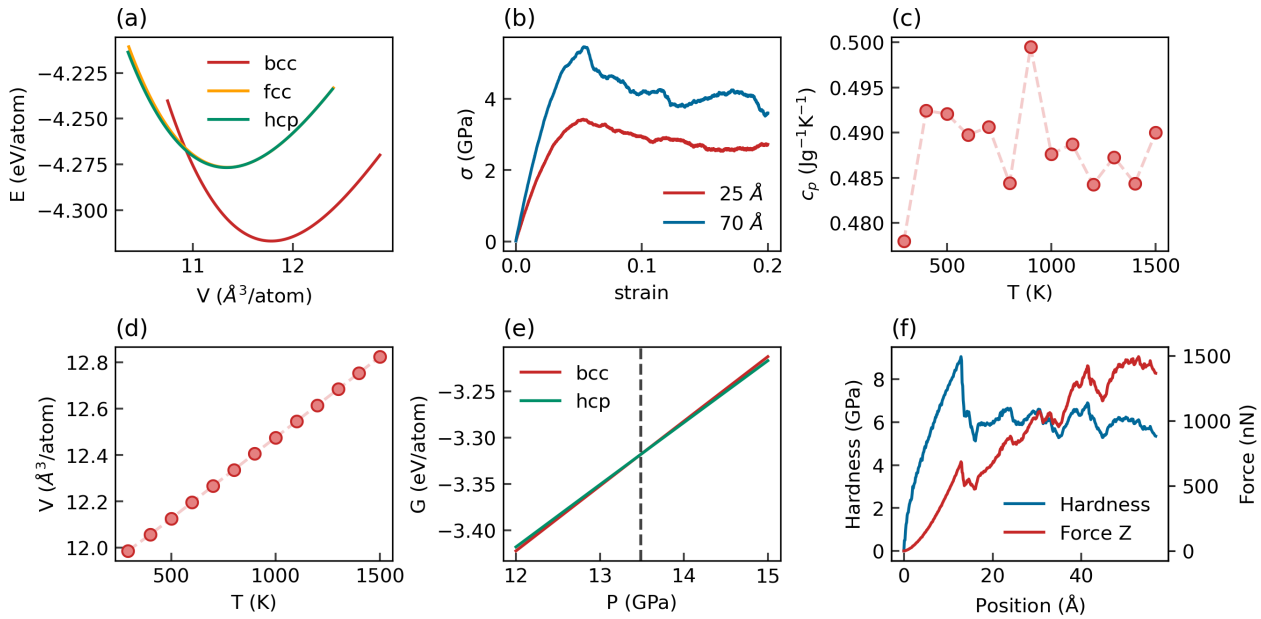

图 4：采用 EAM01 势计算得到的体心立方铁的热物理性质：(a) 能量-体积曲线，并与铁的不同晶体结构进行比较；(b) 流动应力随晶粒尺寸的变化；(c) 比热；(d) 热膨胀系数；(e) 压力对体心立方到六方密堆积相变的影响；(f) 纳米压痕得到的硬度和力。

最后，我们使用 th 第 2.3.7 节中描述的工作流，用于研究铁中碳 n

杂质的能量学。碳原子占据体心立方铁中的八面体位点和置换位置。该案例的计算能量列于表 2，模拟晶胞尺寸为 $5 \times 5 \times 5$，其中含有 250 个铁原子。为计算缺陷形成能，我们使用一种通用机器学习原子间势 GRACE [43], ，它能够描述铁和碳。计算形成能时，碳的参考数值取金刚石的能量（-9.1 eV/atom）。计算得到的缺陷形成能与密度泛函理论的数值吻合良好，间隙碳和置换碳的能量均处于先前研究报告的预期范围内，例如体心立方铁中八面体碳的形成能为 0.72 eV [44]。

表 2：以金刚石中的 C 为参考计算的 GRACE 势函数下体心立方铁中 C 的形成能 [43],

Case	数值（eV）
八面体间隙 C	0.665
置换型 C	2.935

3.2 Eval通过自动化工作流执行评估原子间势 n

在分子动力学模拟中，计算属性往往强烈依赖于所选的原子相互作用模型，即原子间势。因此，有必要进行系统的验证测试，以便为特定材料或关注的属性识别并选择合适的势函数。

在此，我们展示了工作流框架的两项关键优势。首先，工作流完全可复用，无需修改即可轻松适配不同的原子间势。其次，由于工作流的每个阶段都会记录元数据，因此有关势函数的所有相关信息（包括其类型、来源和文献参考）都会被自动捕获，并与生成的数据关联起来。

作为一个代表性示例，我们计算了九种不同原子间势的体积模量和弹性常数。除 EAM01 外，我们还采用了另外十种嵌入原子法势（EAM02 - EAM10） [45, 46, 47, 48, 49, 50, 51, 52, 53], ，以及一种原子簇展开（ACE）机器学习原子间势[54, 55], 和一种通用图原子簇展开（GRACE）势函数 [43]。

由此得到的数值存储在 `ConceptualDict` 对象中，随后被解析并纳入知识图谱，从而实现结构化数据存储和语义查询。这种方法支持直接使用 SPARQL 查询数据。不过，编写 SPARQL 查询可能较为复杂，通常需要深入了解底层本体 [56]。因此，我们使用 `tools4RDF` 库，根据所用本体 [57]辅助生成查询。尽管如此，我们仍明确展示 SPARQL 查询，以提高清晰度，并帮助理解本体谓词的应用方式。列表3展示了一个基于 CMSO 和 ASMO 本体的 SPARQL 查询示例，该查询可获取体积模量数值及其计算所用的相应势函数。基于本体的结构确保所有数据均以语义一致且可由机器解读的方式存储。

```
1 前缀 cmso: <http://purls.helmholtz-metadaten.de/cmso/>2 前缀 rdf: <
http://www.w3.org/1999/02/22-rdf-syntax-ns#>3 前缀 asmo: <
http://purls.helmholtz-metadaten.de/asmo/>4 选择不同的 ?
AtomicScaleSample ?BulkModulusValue ?Reference 5 条件 { 6 ?
AtomicScaleSample asmo:hasCalculatedProperty ?BulkModulus . 7 ?
BulkModulus asmo:hasValue ?BulkModulusValue . 8 9 ?BulkModulus
asmo:wasCalculatedBy ?Simulation . 0 ?Simulation
asmo:hasInteratomicPotential ?Potential . 1 ?Potential
cmso:hasReference ?Reference . 3 { ?AtomicScaleSample rdf:type
cmso:AtomicScaleSample . } 4 { ?BulkModulus rdf:type
asmo:BulkModulus .} 5 }
```

列表3：用于检索体积模量数值及其相关原子间势和文献参考的 SPARQL 查询。

我们采用的软件基础设施由工作流节点、工作流执行环境、概念词典和知识图谱创建流程组成，可确保所有模拟使用相同的晶体结构，并确保知识图谱中的每条模拟记录都保留完整的元数据，包括原子间势和计算参数的详细信息。列表2中的 SPARQL 查询所得的体积模量数值如图5(a)所示。

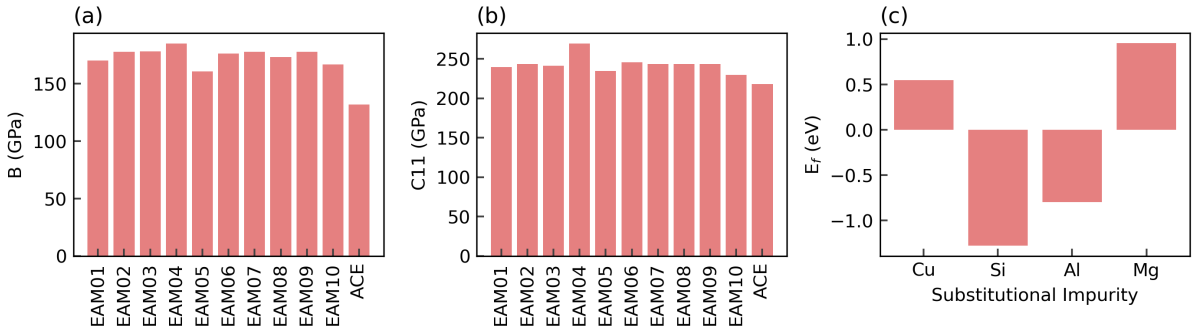

图5：知识图谱 SPARQL 查询的结果：(a) 使用不同 EAM 势计算的体积模量和 (b) 弹性常数 c_{11}。(c) 使用 GRACE [43]模型计算的铁中置换杂质（X=Cu、Si、Al、Mg）的形成能。

针对弹性常数 C_{11} 的类似 SPARQL 查询可提取相应数值，如图5(b)所示。这展示了如何通过程序从知识图谱中检索特定材料性质及其完整的数据溯源信息。这种基于本体的数据表示增强了不同模拟工作流之间的互操作性，并有助于在更广泛的材料科学研究生态系统中重复利用数据和元数据。

我们展示了另一个示例，说明如何查询使用 GRACE[43] 模型计算的缺陷形成能。相应的 SPARQL 查询见清单 4。此示例突显了采用我们的基于知识的工作流后可实现的功能。当通过替换一个原子引入置换缺陷时，工作流会自动标注所得样品含有置换缺陷。因此，可以使用 SPARQL 检索所有包含置换杂质的样品。在传统工作流中，如果只存储原子位置，这类信息很容易丢失，识别缺陷可能需要人工检查。SPARQL 查询会返回缺陷形成能以及系统中存在的化学物种。所得数值如图 5(c) 所示。除 Fe 基体中的 Si 外，所有形成能均处于预期的 0.1 至 0.4 eV 范围内 [43]，该范围来自密度泛函理论计算[58]。这些示例共同展示了本体感知工作流如何实现对原子间势的系统比较，并提供经过语义增强、适用于大规模

分析。

```
1 PREFIX cmso: <http://purls.helmholtz-metadaten.de/cmso/>2 PREFIX rdf: <
http://www.w3.org/1999/02/22-rdf-syntax-ns#>3 PREFIX asmo: <
http://purls.helmholtz-metadaten.de/asmo/>4 PREFIX podo: <
http://purls.helmholtz-metadaten.de/cdos/podo/>5 PREFIX cdco: <
http://purls.helmholtz-metadaten.de/cdos/cdco/>6 SELECT DISTINCT ?
AtomicScaleSample ?数值 ?标签 ?元素 7 (GROUP_CONCAT(DISTINCT ?symbol;
separator='', '') AS ?元素) 8 WHERE { 9 ?AtomicScaleSample
1asmo:hasCalculatedProperty ?property . 0 ?AtomicScaleSample
1cmso:hasMaterial ?材料 . 1 ?材料 cdco:hasCrystallographicDefect?defect . 2 ?
1AtomicScaleSample cmso:hasSpecies ?物种 . 3 ?物种 cmso:hasElement ?元素 . 4 ?
1元素 cmso:hasChemicalSymbol ?symbol . 5 ?property rdfs:label ?标签 . 6 ?
1property asmo:hasValue ?数值 . 8 { ?Defect rdf:type
1podo:SubstitutionalImpurity . } 9 FILTER (?标签='DefectFormationEnergy') 0 }
17 1 GROUP BY ?AtomicScaleSample ?数值 ?标签 2 }
1
1
2
2
2
```

列表 4：用于检索含有置换杂质的样品的缺陷形成能的 SPARQL 查询。该查询会识别原子尺度样品、其计算得到的缺陷形成能以及其中存在的化学物种

3.3 查询结构-性能关系

材料科学的基本原理之一，是理解结构-性能关系，并有效地将其用于材料设计。为此，材料的微观结构发挥着根本作用；晶体学缺陷的存在会直接影响材料的物理性质。我们提出，可以利用语义标注数据来表达这类关系，并支持直接查询，从而验证基本物理原理。

这里，我们以霍尔-佩奇效应为例。该效应给出了金属材料的晶粒尺寸与强度之间的关系，其描述如下：

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

其中， σ_y 为屈服应力（或较高应变下的流动应力）； σ_i 为使位错在单晶或非常大的晶粒中移动所需的应力； k 为霍尔-佩奇系数，是材料特有的常数； d为平均晶粒直径。验证霍尔-佩奇定律对工程应用具有重要意义；已有研究展示了通过从文献中自动提取晶粒尺寸和屈服强度数据进行类似的数据驱动分析 [59]。

根据我们使用单轴拉伸试验工作流（见第2.3.3节）针对采用EAM01势的多晶铁所生成的数据，我们查询具有不同晶粒尺寸的计算样品及其相应计算得到的流动应力，如清单5所示。

```
1 前缀cmso: <http://purls.helmholtz-metadaten.de/cmso/>2 前缀rdf: <
http://www.w3.org/1999/02/22-rdf-syntax-ns#>3 前缀asmo: <
http://purls.helmholtz-metadaten.de/asmo/>4
SELECTDISTINCT?AtomicScaleSample?FlowStress?FlowStressValue?Unit?GrainSize 5
WHERE { 6 ?AtomicScaleSamplecmso:hasSimulationCell?SimulationCell. 7 ?
SimulationCellcmso:hasGrainSize?GrainSize. 8 ?
AtomicScaleSampleasmo:hasCalculatedProperty?FlowStress. 9 ?
FlowStressasmo:hasValue?FlowStressValue. 10 ?FlowStressasmo:hasUnit?Unit. 11
{?AtomicScaleSample rdf:typecmso:AtomicScaleSample.} 12
{?FlowStressrdf:typeasmo:FlowStress.} 13 }
```

列表表 5：从知识图谱中提取平均晶粒尺寸和流动应力的 SPARQL 查询 h.

随后将所得数值绘制在图 6 中，该图展示了平均流动应力随平均晶粒尺寸倒数平方根的变化关系；这两个量均会在工作流阶段自动注释。我们可以看到，根据 Hall–Petch方程，晶界强化机制表现为：应力最初随晶粒尺寸减小而增加，直至达到强化极限，之后应力因晶界滑移而降低。这种软化现象称为反Hall–Petch效应。

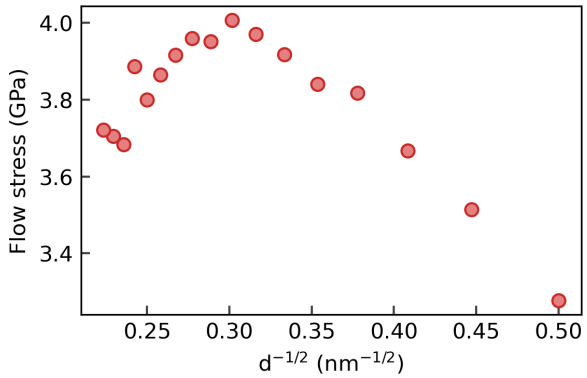

图 6：流动应力随平均晶粒尺寸倒数平方根的变化。

4 讨论

本研究表明，将工作流编排与语义元数据相结合，能够实现一种基于知识的方法，显著提升原子尺度模拟数据的互操作性和可解释性。我们提供了用于计算力学和热力学性质的工作流，并生成了针对铁和 Fe-X材料体系的标注数据集。

这些用例表明，语义标注工作流可用于基于数据验证物理定律，例如直接利用现有模拟数据验证霍尔-佩奇效应。本研究提出的框架还支持对不同原子间势进行系统比较，这对于原子尺度建模中的不确定性量化至关重要。这些工作流有意涵盖了更复杂的系统和模拟，例如拉伸和压缩测试、自由能计算、纳米压痕以及缺陷形成能计算，展现了语义工作流设计可扩展至与工程应用相关的力学行为和相变场景。

图 7 表明，基于知识的工作流通过引入模块化工作流节点、语义标注和 FAIR 数据，克服了传统临时模拟实践的局限。

所提出基础设施的附加价值，可以通过考察其对互操作性和可复用性的影响来最好地说明：

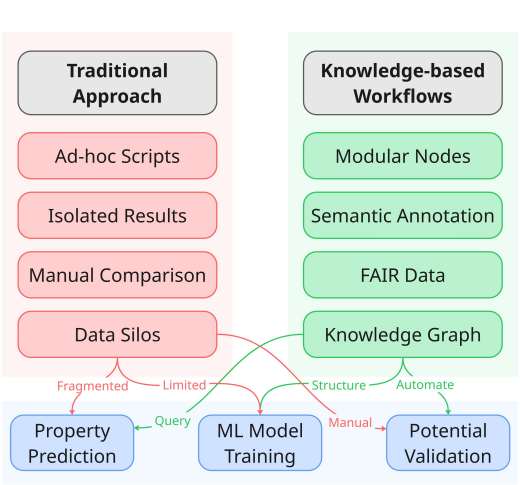

图 7：传统临时模拟脚本（左）与基于知识的工作流（右）的比较。传统方法会产生彼此孤立的结果，需要人工进行比较，并导致数据孤岛。语义工作流能够生成符合 FAIR 的数据和可查询的知识图谱，从而提升互操作性和可复用性。

i. 互操作性： 就 FAIR 数据而言，所捕获的元数据与应用层和领域层本体保持一致，为描述材料体系、计算样本、模拟工作流和计算属性提供了通用词汇。使用一致的语义描述符，可以在同一知识空间中将密度泛函理论计算的数据输出与经典分子动力学和机器学习原子间势模拟关联起来。该框架还支持与外部资源（如材料数据库）的语义集成，并有望实现与实验数据的互操作。就 FAIR 工作流而言，我们的节点不依赖于特定的工作流环境。文献表 1 和 2 展示了这一点，其中使用了两个不同的工作流系统 pyiron 和 jobflow [16], 。还可以进一步连接其他工作流系统，例如使用 Python工作流定义 [60]。总体而言，这表明本体感知工作流节点可以在不同的工作流系统之间无缝复用。

ii. 可复用性： 所提出方法在生成模拟工作流和数据方面的核心优势，是能够针对不同任务实现可复用性。无需修改代码，工作流即可使用不同的原子间势执行；第 3.2 节对 10 种不同势函数进行的系统评估证明了这一点。语义层同时捕获每次模拟所涉及的物理概念和流程步骤，因此只需调整相关元数据参数，同一工作流即可在不同材料体系或热力学条件下重复使用。这一能力显著提升了生成可比数据集以进行基准测试或模型验证的自动化潜力，并支持对原子模拟中的方法学不确定性开展系统研究。

尽管所提出的框架提升了 FAIR 数据和工作流生产过程中的互操作性和可复用性，但仍存在一些必须承认的局限性：

- 用于计算与工程应用相关性质的工作流仍然高度特定于具体应用，并且需要领域专业知识才能进行正确的构型与解释。分子动力学从根本上受可及时间尺度的制约，进而限制了现实的加载速率和相变动力学。因此，适当输入参数的选择仍取决于具备相关知识的用户，不当选择会直接影响结果质量。此外，对边界条件（如晶胞形状、加载模式或恒温器与恒压器算法）的敏感性，仍会影响模拟结果的可比性。语义层可以记录这些选择，但尚不能消除其背后的物理依赖关系。

- 工作流对全新模拟场景的通用性并非自动具备。准确的语义标注需要领域专业知识，而本体和工作流

模板都需要扩展，以适应新的物理过程或建模方法。尽管如此，整体架构仍可扩展到新的模拟场景。另一种方法是semantikon [61]，它允许使用任意本体进行序列化，从而提供更大的灵活性，但也将确保语义标注一致且有意义的责任转移给用户。

•最后，就融入更广泛的语义基础设施而言，我们目前最高层级的对齐依托于 PROV-O。尽管 PROV-O 提供了轻量且灵活的溯源表示方式，但与基本形式本体（BFO）等上层本体对齐，将进一步增强互操作性。现有的 PROV-O 与 BFO 映射表明，二者并非不兼容 [63], ，但采用与顶层本体对齐的设计，将有助于加强其在更广泛的材料科学语义生态系统中的集成。 [62]

未来的研究方向包括加强与外部知识图谱和更广泛研究数据基础设施的集成，以实现对模拟结果、实验测量数据和文献衍生知识的联邦搜索。另一项机遇是将语义工作流框架与新兴的智能体人工智能流程相连接，使其能够自主选择工作流、挑选原子间势函数、提出验证步骤并生成新的模拟数据。例如， LangSim 是一个将大语言模型与模拟工作流相结合的库 [64]。语义层能够提升此类人工智能驱动方法的透明度和可信度。此外，扩展底层本体，以涵盖微观结构演化、更复杂的缺陷相互作用以及先进的热力学描述符，将使该框架能够支持范围更广的物理现象和材料建模任务。归根结底，语义标注工作流不仅可作为文档工具，也可作为一种框架，加速从数据中理解和生成见解、验证模型并提高科学研究的可复现性。

5 结论与展望

在本研究中，我们引入了用于计算材料力学性质和热力学性质的基于知识的工作流，展示了元数据标注、本体对齐和溯源追踪如何将传统模拟脚本转变为结构化且具有互操作性的研究对象。通过将模拟分解为与领域本体和应用本体绑定的模块化工作流组件，我们实现了对各种原子模拟任务中的材料、计算样品、方法和计算属性进行一致描述。针对铁的用例所开发的工作流表明，这种方法能够提高可复用性，促进跨数据集的系统比较，并支持对材料性质进行透明解读。

本研究的一个核心成果是增强了互操作性和可复用性。语义层统一了不同模拟软件和建模方法中的元数据，为采用不同材料体系、势函数或边界条件的工作流所生成的数据提供一致标注。这种互操作性还可自然扩展至机构数据库以及 NFDI-MatWerk 框架内的国家研究数据基础设施解决方案，从而实现与实验数据集、外部模拟存储库以及不断发展的知识图谱生态系统的集成。

结构化元数据和模块化工作流设计也使该框架能够支持未来由人工智能驱动的材料发现。语义标注数据集支持针对物理概念的定向查询以及对结构-性能关系的验证，而智能体人工智能系统则可利用可重复使用的工作流节点来选择模拟方案、提出验证任务并协调大规模数据生成。这些能力为自动探索力学性质、高通量筛选相稳定性以及制定有依据的合金设计工作流提供了途径。

展望未来，该框架可扩展至更复杂的模拟场景，例如扩散过程和缺陷-缺陷相互作用，以及与介观或连续介质模型的多尺度耦合。随着可复用工作流库和材料科学本体不断丰富，该领域可以

从基于脚本的模拟转向知识驱动的计算材料科学。因此，我们鼓励学界采用、完善并扩展本文提出的工作流和本体，推动建立共享的语义基础，以支持可复现、可扩展且适用于人工智能的计算材料科学。

数据可用性声明

工作流见此处 https://github.com/pyscal/semantic-workflows。数据及相应的元数据见此处 https://doi.org/10.5281/zenodo.18380870。

致谢

本研究得到了 NFDI-MatWerk 联盟的支持，该联盟由德国研究基金会（DFG）资助，属于德国国家研究数据基础设施（NFDI）项目（NFDI 38/1，项目编号 460247524）。H.-T.L. 和 N.M. 感谢 DFG 的资助（项目编号 394563137——SFB 1368 和 ME 6073/20-1）。A.A.G. 衷心感谢获准使用 Forschungszentrum J¨ulich 的 JURECA 超级计算机进行计算 [65]，资助项目 IDdefectskg。

参考文献

[1] 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，*APL Materials* **2013**，*1*，1 011002。

[2]M. K. Horton, P. Huck, R. X. Yang, J. M. Munro, S. Dwaraknath, A. M. Ganose, R. S. Kingsbury, M. Wen, J. X. Shen, T. S. Mathis, A. D. Kaplan, K. Berket, J. Riebesell, J. George, A. S. Rosen, E. W. C. Spotte-Smith, M. J. McDermott, O. A. Cohen, A. Dunn, M. C. Kuner, G.-M. Rignanese, G. Petretto, D. Waroquiers, S. M. Griffin, J. B. Neaton, D. C. Chrzan, M. Asta, G. Hautier, S. Cholia, G. Ceder, S. P. Ong, A. Jain, K. A. Persson, *Nature Materials***2025**,*24*, 10 1522.[3] S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, D. Morgan, 计算材料科学 **2012**,*58* 218.[4] M. Scheidgen, L. Himanen, A. N. Ladines, D. Sikter, M. Nakhaee, A. Fekete, T. Chang, A. Gol- parvar, J. A. M´arquez, S. Brockhauser, S. Br¨uckner, L. M. Ghiringhelli, F. Dietrich, D. Lehm- berg, T. Denell, A. Albino, H. N¨asstr¨om, S. Shabih, F. Dobener, M. K¨uhbach, R. Mozumder, J. F. Rudzinski, N. Daelman, J. M. Pizarro, M. Kuban, C. Salazar, P. Ondraˇcka, H.-J. Bun- gartz, C. Draxl, 《开放源代码软件杂志》**2023**,*8*, 90 5388.[5]S. Menon, Y. Lysogorskiy, A. L. M. Knoll, N. Lei meroth, M. Poul, M. Qamar, J. Janssen, M. Mrovec, J. Rohrer, K. Albe, J. Behler, R. Drautz, J. Neugebauer, *npj ComputationalMaterials* **2024**,*10*, 1 261.[6] L. Himanen, A. Geurts, A. S. Foster, P. Rinke, *Advanced Science* **2019**, *6*, 21 1900808.[7] M.L.Evans, J.Bergsma, A.Merkys, C.W. Andersen, O.B.Andersson, D.Beltr´an, E.Blokhin, T. M. Boland, R. Casta˜neda Balderas, K. Choudhary, A. D´ıaz D´ıaz, R. Dom´ınguez Garc´ıa, H. Eckert, K. Eimre, M. E. Fuentes Montero, A. M. Krajewski, J. J. Mortensen, J. M. N´apoles Duarte, J. Pietryga, J. Qi, F. d. J. Trejo Carrillo, A. Vaitkus, J. Yu, A. Zettel, P. B. de Castro, J. Carlsson, T. F. T. Cerqueira, S. Divilov, H. Hajiyani, F. Hanke, K. Jose, C. Oses, J. Riebesell, J. Schmidt, D. Winston, C. Xie, X. Yang, S. Bonella, S. Botti, S. Cur- tarolo, C. Draxl, L. E. Fuentes Cobas, A. Hospital, Z.-K. Liu, M. A. L. Marques, N. Marzari, A. J. Morris, S. P. Ong, M. Orozco, K. A. Persson, K. S. Thygesen, C. Wolverton, M. Schei- dgen, C. Toher, G. J. Conduit, G. Pizzi, S. Graˇzulis, G.-M. Rignanese, R. Armiento, 数字探索**2024**,*3* 1509.

[8] M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Ban´aˇs, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Fer- guson, M. Filizola, J. S. Fraser, H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. L¨ohr, F. Marinelli, L. Martin-Samos, M. Masetti, R.Meyer, A.Michaelides, C. Molteni, T. Morishita, M. Nava, C.Paissoni, E. Papa- leo, M. Parrinello, J. Pfaendtner, P. Piaggi, G. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G. C. Sosso, ˇV. Spiwok, J. Sponer, D. W. H. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth, A. White, T. P. 联盟, 《自然·方法》**2019**,*16*, 8 670.

[9]M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez- Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G.Strawn, M.A.Swertz, M.Thompson, J.vanderLei, E.vanMulligen, J.Velterop, A. Waag-meester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, 《科学数据》**2016**, *3*, 1 160018。

[10] M. Scheffler、M. Aeschlimann、M. Albrecht、T. Bereau、H.-J. Bungartz、C. Felser、M. Greiner、A.Groß、C. T.Koch、K. Kremer、W.E.Nagel、M.Scheidgen、C.W¨oll、C.Draxl, 《自然》《自然》**2022**, *604*, 7907 635。

[11]L. M. Ghiringhelli, C. Baldauf, T. Bereau, S. Brockhauser, C. Carbogno, J. Cha manara,´S. Cozzini, S. Curtarolo, C. Draxl, S. Dwaraknath, A. Fekete, J. Kermode, C. T. Koch, M. K¨uhbach, A. N. Ladines, P. Lambrix, M.-O. Himmer, S. V. Levchenko, M. Oliveira, A. Michalchuk, R. E. Miller, B. Onat, P. Pavone, G. Pizzi, B. Regler, G.-M. Rignanese, J. Schaarschmidt, M. Scheidgen, A. Schneidewind, T. Sheveleva, C. Su, D. Usvyat, O. Valsson, C. W¨oll, M. Scheffler, 《科学数据》2023, 10, 1 626.[12] B. Bayerlein, T. Hanke, T. Muth, J. Riedel, M. Schilling, C. Schweizer, B. Skrotzki, A. Todor, B. Moreno Torres, J. F. Unger, C. V¨olker, J. Olbricht, 《先进工程材料》2022, 24, 6 2101176. [13] B. Bayerlein, M. Schilling, P. von Hartrott, J. Waitelonis, 《科学数据》2024, 11, 1 434. [14] A. Ghafarollahi, M. J. Buehler, 《先进材料》2025, 37, 22 2413523.[15] J. Janssen, S. Surendralal, Y. Lysogorskiy, M. Todorova, T. Hickel, R. Drautz, J. Neugebauer, 《计算材料科学》2019, 163 24.[16] A. S. Rosen, M. Gallant, J. George, J. Riebesell, H. Sahasrabuddhe, J.-X. Shen, M. Wen, M. L. Evans, G. Petretto, D. Waroquiers, G.-M. Rignanese, K. A. Persson, A. Jain, A. M. Ganose, 《开放源代码软件杂志》2024, 9, 93 5995.[17] S. P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. H¨auselmann, D. Gresch, T. M¨uller, A. V. Yakutovich, C. W. Andersen, F. F. Ramirez, C. S. Adorf, F. Gargiulo, S. Kumbhar, E. Passaro, C. Johnston, A. Merkys, A. Cepellotti, N. Mounet, N. Marzari, B. Kozinsky, G. Pizzi, 《科学数据》2020, 7, 1 300.[18] M. R. Crusoe, S. Abeln, A. Iosup, P. Amstutz, J. Chilton, N. Tijani´c, H. M´enager, S. Soiland- Reyes, B. Gavrilovi´c, C. Goble, CWL 社区, 《美国计算机协会通讯》2022, 65, 6 54.[19] L. Huber, J. Neugebauer, pyiron 工作流, `https://github.com/pyiron/pyiron_workflow`, 访问日期：2025-02-15.

[20] W3C RDF 工作组，RDF 1.1 概念与抽象语法，**2014**，网址 https://www.w3.org/TR/rdf11-concepts/，W3C 建议标准。

[21] W3C OWL 工作组，OWL 2 Web 本体语言文档概述，**2012**，URL https://www.w3.org/TR/owl2-overview/，W3C 建议标准。

[22]C. Eberl、M. Niebel、E. Bitzek、T. Dahmen、F. Fritzen、P. Gumbsch、T. Hickel、S. Klein、F. M¨ucklich、M. S. M¨uller、J. Neugebauer、C. Pauly、M. Politze、H. Sack、S. Sandfeld、R. Schwaiger、P. Slusallek、R. Stotzka、A. Streit、M. Zimmermann，NFDI-MatWerk 联盟提案，**2021**，URL https://doi.org/10.5281/zenodo.5082837。

[23] A. A. Guzm´an、S. Sandfeld，计算材料样品本体（CMSO），2024，URL https://doi.org/10.5281/zenodo.10805536。[24] A. A. Guzm´an、S. Sandfeld，原子模拟方法本体（ASMO），2024，URLhttps://doi.org/10.5281/zenodo.10805591。[25] W3C 溯源工作组，PROV-O：PROV 本体，2013，URL https://www.w3.org/TR/prov-o/，W3C 建议标准。[26] S. Menon、A. A. Guzm´an，atomRDF：用于基于本体创建、操作和查询原子结构的 Python 工具，2024，URL https://github.com/pyscal/atomRDF。[27] 语义工作流代码库，URL https://github.com/pyscal/semantic-workflows。[28] A. P. Thompson、H. M. Aktulga、R. Berger、D. S. Bolintineanu、W. M. Brown、P. S. Crozier、P. J. in 't Veld、A. Kohlmeyer、S. G. Moore、T. D. Nguyen、R. Shan、M. J. Stevens、J. Tranchida、C. Trott、S. J. Plimpton，《计算物理通讯》2022，271，108171。[29] LAMMPS 单位制，https://docs.lammps.org/units.html，LAMMPS 文档，访问于 2025-02-15。[30] F. Birch，《物理评论》1947，71，809。[31] LAMMPS 开发者，LAMMPS 示例：弹性常数计算，GitHub 代码仓库，2025，URL https://github.com/lammps/lammps/tree/develop/examples/ELASTIC，访问日期：2025-02-17。[32] P. Hirel，《计算物理通讯》2015，197，212。[33] M. P. Allen、D. J. Tildesley，《液体计算机模拟》，牛津大学出版社，牛津，第2版，2017。[34] S. Menon、Y. Lysogorskiy、J. Rogal、R. Drautz，《物理评论：材料》2021，5，103801。[35] S. Huang、C. Zhou，《JOM——矿物、金属与材料学会期刊》2017，69，11，2256。[36] M. S. Daw、M. I. Baskes，《物理评论B辑》1984，29，6443。[37] N. Gunkelmann、E. M. Bringa、K. Kang、G. J. Ackland、C. J. Ruestes、H. M. Urbassek，《物理评论B辑》2012，86，14，144111。[38] J. J. Adams、D. S. Agosta、R. G. Leisure、H. Ledbetter，《应用物理学杂志》2006，100，11，113530。[39]H.-T. Luu、S.-L. Dang、T.-V. Hoang、N. Gunkelmann，《应用表面科学》2021，551，149221。

[40] M. W. Chase、C. A. Davies、J. R. Downey、D. J. Frurip、R. A. McDonald、A. N. Syverud，*NIST- JANAF* 热化学表，NIST 标准参考数据库 13。美国国家标准与技术研究院，1.0 版，**1985**。

[41] Y. M. Kozlovskii, S. V. Stankus, 《物理学杂志：会议系列》**2019**，*1382*，1 012181。

[42] B. J. Jensen、I. Gray、G. T.、R. S. Hixson，《应用物理学杂志》2009，105，10 103502。[43] Y. Lysogorskiy、A. Bochkarev、R. Drautz，面向基础机器学习原子间势的图原子簇展开，2025，网址 https://arxiv.org/abs/2508.17936。[44] E. Hristova、R. Janisch、R. Drautz、A. Hartmaier，《计算材料科学》2011，50，3 1088。[45] Y. Sun、F. Zhang、M. I. Mendelev、R. M. Wentzcovitch、K.-M. Ho，《美国国家科学院院刊》2022，119，2 e2113059119。[46] S. Starikov、D. Smirnova、T. Pradhan、Y. Lysogorskiy、H. Chapman、M. Mrovec、R. Drautz，《物理评论：材料》2021，5 063607。[47] L. Proville、D. Rodney、M.-C. Marinica，Nature Materials 2012，11，10 845。[48] S. Chiesa、P. M. Derlet、S. L. Dudarev、H. V. Swygenhoven，《物理学杂志：凝聚态物质》2011，23，20 206001。[49]L. Malerba、M. Marinica、N. Anento、C. Bj¨orkas、H. Nguyen、C. Do main、F. Djurabekova、P. Olsson、K. Nordlund、A. Serra、D. Terentyev、F. Willaime、C. Becquart，《核材料杂志》2010，406，1 19，fP6 IP PERFECT项目：反应堆部件辐照损伤效应预测。[50] P. A. Olsson，《计算材料科学》2009，47，1 135。[51] H. Chamati、N. Papanicolaou、Y. Mishin、D. Papaconstantopoulos，《表面科学》2006，600，9 1793。[52] X. W. Zhou、R. A. Johnson、H. N. G. Wadley，《物理评论B辑》2004，69 144113。[53] M. I. Mendelev、S. Han、D. J. Srolovitz、G. J. Ackland、D. Y. Sun、M. Asta，《哲学杂志》2003，83，35 3977。[54]Y. Lysogorskiy、C. v. d. Oord、A. Bochkarev、S. Menon、M. Rinaldi、T. Hammerschmidt、M. Mrovec、A. Thompson、G. Cs´anyi、C. Ortner、R. Drautz，npj Computational Materials 2021，7，1 97。[55] B. Bienvenu、M. Todorova、J. Neugebauer、D. Raabe、M. Mrovec、Y. Lysogorskiy、R. Drautz，npj Computational Materials 2025，11，1 81。[56] A.-C. N. Ngomo、L. B¨uhmann、C. Unger、J. Lehmann、D. Gerber，《万维网》国际会议论文集，第22届，2013。[57] S. Menon、A. A. Guzm´an、O. Waseda、S. Sandfeld、T. Hickel，《开放源代码软件杂志》2026，11，117 9482。[58] B. Jelinek、S. Groh、M. F. Horstemeyer、J. Houze、S. G. Ki m、G. J. Wagner、A. Moitra、M. I. Baskes，《物理评论B辑》2012，85 245102。[59] P. Kumar、S. Kabra、J. M. Cole，《科学数据》2022，9，1 292。[60]J. Janssen、J. George、J. Geiger、M. Bercx、X. Wang、C. Ertural、J. Schaarschmidt、A. M. Ganose、G. Pizzi、T. Hickel、J. Neugebauer，Digital Discovery 2025，4 3149。[61]Osamu Waseda，Semantikon，网址 https://github.com/pyiron/semantikon，访问日期：2025-02-14。[62] J. N. Otte、J. Beverley、A. Ruttenberg，《应用本体论》2022，17，1 17。[63] T. Prudhomme、G. De Colle、A. Liebers、A. Sculley、P. K. Xie、S. Cohen、J. Beverley，《科学数据》2025，12，1 282。

[64] Y. Zimmermann, A. Bazgir, A. Al-Feghali, M. Ansari, J. Bocarsly, L. C. Brinson, Y. Chiang, D. Circi, M.-H. Chiu, N. Daelman, M. L. Evans, A. S. Gangan, J. George, H. Harb, G. Khaligh- inejad, S.T.Khan, S. Klawohn, M. Lederbauer, S.Mahjoubi, B.Mohr, S.M. Moosavi, A. Naik, A. B. Ozhan, D. Plessers, A. Roy, F. Sch¨oppach, P. Schwaller, C. Terboven, K. Ueltzen, Y. Wu, S. Zhu, J. Janssen, C. Li, I. Foster, B. Blaiszik, *arXiv* 预印本**2025**。

[65] 于¨利希超级计算中心，大型研究设施期刊**2021**，*7*，A182。