arXiv:2601.09456v1 [cs.SE] 14 Jan 2026

面向能源研究软件的元数据模式

Stephan Ferenz [1, 2]，Oliver Werth [2]，以及 Astrid Nieße [1,2]

[1]计算机科学系，卡尔·冯·奥西茨基奥尔登堡大学，奥尔登堡，德国¨能源部，OFFIS 信息技术研究所，奥尔登堡，德国[2]

2026年1月14日

摘要

领域特定的元数据模式对于提高科研软件的可发现性和可复用性，以及遵循 FAIR4RS 原则至关重要。然而，许多领域，包括能源研究，都缺乏已建立的元数据模式。为弥补这一空白，我们基于需求分析开发了一种能源研究软件元数据模式，并通过用户测试对其进行了评估。结果表明，该模式在满足形式化和互操作性需求的同时，也满足了能源研究人员的特定需求。此外，测试显示，对必需信息进行良好呈现是使研究人员能够创建所需元数据的关键。本文介绍了设计能源研究软件元数据模式所面临的挑战与机遇。

引言

FAIR4RS（研究软件可查找、可访问、可互操作和可重用）原则 [4]为提升科研软件的整体可重用性提供了关键框架。研究表明，领域特定的元数据模式对于提高科研软件的可发现性和可重用性至关重要 [1, 4]。尽管一些科研领域已经为科研软件建立了领域特定的元数据模式，例如生命科学 [27]和地球科学 [22],，但其他领域目前仍落后于此。要在这些领域实现 FAIR 研究软件，开发领域特定的元数据模式是关键一步。

其中一个领域是能源研究，其重点在于理解、分析、改进和设计能源系统及组件，涵盖社会、经济、环境和技术层面。这类研究高度依赖科研软件，例如用于能源组件或整个能源系统的数据分析和仿真工具 [16]。目前已经出现了一些面向能源研究软件的元数据模式初步方案，例如 开放能源平台（OEP）上的事实表。这些方法在形式化、与通用元数据模式的互操作性以及作为基础的深入需求分析方面仍有局限 [16]。因此，现有方法不足以支持能源领域中的 FAIR 研究软件。

*通讯作者：stephan.ferenz@uol.de

为解决这些问题并推动 FAIR4RS，我们首先基于对能源研究人员进行的 32 次访谈，提出了需求分析 [16]。除了这些需求之外，能源研究软件的元数据模式还应满足其他要求，例如与现有模式的互操作性，以及便于用户依据该模式创建元数据。本文探讨了如何设计能源研究软件的元数据模式，以平衡这些不同的要求。为回答这一问题，我们开发了一种能源研究软件元数据模式（*ERSmeta*），并让能源研究人员为自己的软件创建元数据，以此对其进行评估。

本文其余部分安排如下：首先，我们介绍“相关工作”。随后，“方法”部分介绍用于开发元数据模式的流程，并描述评估的设置。接着，我们介绍“元数据模式：ERSmeta”，并展示“元数据模式评估”的结果。最后，“讨论”部分概述我们方法的优势与局限性，并指出未来的研究方向。

相关工作

本节基于我们的前期工作，概述现有的科研软件元数据模式，见 [16] 和 [13]。所有介绍的方法汇总于表1。首先，我们考察通用方法以及来自其他领域的方法。

CodeMeta [28]是一项由社区推动的科研软件元数据标准，基于*schema.org*。目前已有各种与其他元数据模式的交叉映射，例如与 GitHub 应用程序编程接口（API）元数据的交叉映射。*CodeMeta* 包含多个元素，其中一些侧重于文件大小或支持的操作系统等技术细节，另一些则涵盖许可证等管理信息。该元数据标准没有必需元素。它支持使用统一资源标识符（URI）来标识作者、贡献者和许可证。其内容特定元数据仅限于应用类别和关键词。

“引文文件格式（*CFF*） [7] 为正确引用科研软件提供了必要信息。尽管它涵盖了一些相关元数据信息，但并未定义恰当的元数据模式。此外，*CFF* 仅限于一般信息，不包括任何领域特定元素。

Kuckertz 等人 [30] 提出了*DataDesc*，这是一种专注于科研软件接口的元数据模式。该模式复用了*CodeMeta* 中的多个元素，并支持对科研软件接口进行详细描述。除关键词和描述外，它并未涵盖详细的领域特定方面。

针对地球科学，Gil、Ratnakar 和 Garijo [21]开发了*OntoSoft*，这是一种通过六个类别描述科研软件的本体：识别、理解、执行、开展研究、获取支持和更新。Gariko 等人 [19, 20] 通过创建软件描述本体扩展了这一方法，其中基于科学变量本体增加了对输入和输出数据的描述。他们还将这一方法与*CodeMeta* 对齐，并支持在开放知识图谱中发布元数据，其中包含指向语义网中其他实例（如维基数据）的链接。

Ison 等人 [27] 为生命科学软件注册库 bio.tools 开发了元数据模式*biotoolsXSD* 。该模式以 XML 模式的形式表示，包含 55 个元素，其中 10 个为必需元素。部分元素（如功能、输入和输出）必须使用*EDAM* 本体作为值词汇表。该模式还包含软件专属元素，例如编程语言、许可证和操作系统，这些元素无需使用本体。

其次，我们考察能源领域的方法。Schwarz 和 Lehnhoff [32] 描述了一个能源协同仿真组件目录。他们使用语义媒体维基收集有关仿真器和信息

表1：科研软件元数据模式概览（基于 [13] 和 [16]）

✓：已满足　(✓)：部分满足　✗：未满足

		Domain	Mandatory elements	Reuse elements from existing approaches	Use of domain ontologies as value vocabulary
元数据模式	*CodeMeta* （[28]）	通用	✗	✓	✗
	CFF [7]	通用	✗	✗	✗
	DataDesc [30]	通用	✓	✓	✗
	BiotoolsSchema [27]	生物信息学	✓	✗	✓
本体	*OntoSoft* [21]	地球科学	✓	✗	✗
	软件描述本体 [19]	通用	✗	✓	✓
Not 形式化的 ap- 方法	能源协同仿真组件目录组件 [32]	能源	✗	✗	(✓)
	开放能源建模倡议（openmod）维基	能源	✗	✗	(✓)
	关于框架的 OEP 事实表	能源	✓	✗	(✓)

功能模型接口（FMI）的信息，以补充仿真接口的描述。然而，目录中的元素可用于构建元数据模式，但文中并未对其进行更详细的描述。

Topenmod 在 其维基中列出了一系列能源模型。然而，元数据模式 is 未正式化，也未对元素或数值使用受控词表。

The OEP 包含有关能源研究中模型和框架的信息。其元数据元素与 openmod 维基中的类似。对它们的描述有限，元数据也未经过形式化。

能源领域的所有方法均未被形式化为元数据模式。此外，这些方法缺少相关描述，未说明元素是如何选定的，以及在此过程中纳入了何种程度的需求 [16]。

方法

本节将介绍我们制定元数据模式*ERSmeta* 及对其进行评估的过程。首先，我们介绍用于开发该元数据模式的方法，随后概述我们的评估方法。



图1：我们开发元数据模式的方法概览。白色方框表示专家交互和测试。深灰色方框表示开发步骤。浅灰色方框表示（中间）产物。

元数据模式的开发

图1 展示了我们开发 *ERSmeta* 所采用的流程。我们采用了 Curado Malta 和 Baptista 提出的 Me4MAP（元数据应用配置文件开发方法），并据此调整了我们的方法 [5]。我们首先从三个来源收集信息：基于访谈的需求分析、对科研软件文献中需求的分析，以及对现有元数据模式中元素的分析。这些需求共同促成了一个广泛的领域模型。接下来，我们对领域模型中的元素进行了细化和优先级排序。我们邀请多位专家评审这一版本的 *ERSmeta*，以收集初步反馈。随后，我们将领域模型中的元素与现有元数据模式进行对齐，以实现高度互操作性。然后，我们为多个元素定义了值词汇表。我们使用两款软件测试了由此得到的 *ERSmeta* 版本。最后，我们将该模式形式化，并进行了最后一轮测试。下文将详细介绍所有步骤。

基于访谈的需求分析 由于*ERSmeta* 应涵盖相关领域的特定信息，我们决定广泛收集信息需求。因此，我们对能源研究人员进行了 32 次专家访谈 [16]。我们在 [16]中发布了需求分析结果及由此形成的领域模型。

表 2：所分析的元数据模式概览

	软件	Data
通用	CodeMeta [28],DataDesc [30],R软件包元数据, Python软件包元数据，CFF [7], RSD数据模型, 软件描述本体 [19]	
其他领域	*ontosoft* [22],*BiotoolsSchema* [27]	
工程		m4i [26]
能源	openmod，OEP 框架与模型事实说明	OEMetadata [25]

研究软件文献中的需求 预期的元数据模式不仅应满足能源研究人员当前的需求，还应有助于实现 FAIR 研究软件。因此，我们还分析了 FAIR4RS 原则 [4] 和其他研究软件文献，以确定*ERSmeta*的更多需求。特别是， [23] 对这些需求进行了重要的概述。这些需求对应于一些具体元素，我们已将其添加到领域模型中。

现有模式中的元素 此外，我们分析了相关元数据模式中的元素，以识别相关元数据元素，从而弥补空白，并明确需求分析中提取的元素。 表 2 概述了我们在此步骤中分析的元数据模式。因此，我们向领域模型中添加了其他元素。

改进领域模型 我们通过减少领域模型中的冗余，对领域模型进行了整合。此外，我们通过将元素划分为必需、推荐和可选，进行了第一轮优先级划分。我们为所有元素添加了定义。如果元素取自现有模式，我们则沿用了现有定义。

专家反馈轮次 为了进一步改进 *ERSmeta*，我们收集了五位专家的专家反馈；他们的研究领域是研究软件元数据和能源研究软件。我们将包含所有元素及其描述和优先级划分的领域模型以 Excel 表格的形式发送给专家，并请他们填写一份简短的调查问卷以提供反馈。目标是识别缺失的元素、不恰当的描述和不一致之处。我们将收到的反馈纳入了领域模型。

与现有元数据模式对齐 为了实现高度互操作性，我们将领域模型中的元素与现有元数据模式进行了对齐。Curado Malta 和 Baptista 将这一过程称为词汇对齐 [5]。我们将领域模型映射到以下现有模式：*CodeMeta* [28], *CFF* [7], *DataDesc* [30],*OEMetadata* [25], *OEP 框架与模型事实说明*、*m4i* [26], *ontosoft* [22], *BiotoolsSchema* [27], *Bioschemas ComputationTool*以及软件描述本体 [19]。我们检查了每个对齐元素的定义，以确定能否将现有本体或元数据模式中的元素复用于*ERSmeta*。我们优先复用 *schema.org* 中的元素，因为它是最常用的本体。如果找到匹配的元素，我们就复用其名称和描述，并添加其命名空间链接。通常，如果数据类型通用，我们也会尽量复用该类型。在某些情况下，为了更好地与现有定义保持一致，我们需要调整元素的层级结构。

值词汇表的定义 为了限制某些元素允许使用的数值，从而提高所生成元数据的互操作性，我们为多个元素定义了值词汇表。值词汇表的数据来源各不相同。在可能的情况下，我们使用现有本体，例如开放能源本体 [2]，或使用维基数据等语义知识库。否则，我们会依据文献或其他现有来源创建允许使用的术语列表。对于某些元素，我们扩展了现有列表；对于另一些元素，我们则根据专家知识构建了完整列表。

1st 使用两款软件进行测试为了进一步改进*ERSmeta*，我们根据两款能源研究软件的元数据模式创建元数据，以测试其可用性。对于每款软件，我们都与该软件的开发者共同创建元数据。根据这次测试，我们改进了模式（例如，描述、允许的数值等）。

元数据模式的形式化为了形式化*ERSmeta*，我们决定使用两种语法类型。我们使用了形状约束语言（SHACL）[29] 的 Turtle 语法，以确保与语义网具有良好的兼容性。SHACL 可以为所有元素设置复杂约束，例如，可以将某个元素限制为特定类别，这对动态值词表很有用。此外，我们还使用了 JSON-LD 模式，以提高可读性并增强与CodeMeta的互操作性；CodeMeta 也以 JSON-LD 模式提供。在这种形式化中，只有元素词汇与语义网兼容，约束和数值则不兼容。我们首先定义了 SHACL，随后将其转换为 JSON-LD 模式，并在 JSON-LD 模式中补充了其他必需信息。此外，我们还在代码库中添加了与其他元数据模式的交叉映射。

2nd 使用两款软件进行测试我们使用同样的两款软件进行了第二轮测试，也对形式化过程进行了测试。根据第一轮测试中获得的非正式元数据，我们按照 SHACL 将正式元数据创建为 Turtle 格式，并按照 JSON-LD 模式创建为 JSON-LD 格式。我们测试了这两种格式是否都能表达所有信息，也评估了依据模式进行验证是否成功。我们修复了一些错误，并改进了描述。

评估设计

开发 *ERSmeta* 后，我们着手开展更广泛的评估。由于我们的开发工作已纳入有关 FAIR 研究软件的文献要求（例如贡献者角色），以及寻找能源研究软件的研究人员的需求，因此我们将重点放在基于元数据模式创建元数据这一方面。能源研究软件的元数据通常应由相应软件的开发者创建，例如研究软件工程师或编写代码的研究人员。为评估这一方面，我们将评估研究设计为现场实验：参与者为自己的能源研究软件创建元数据，并通过问卷以结构化的方式分享体验。问卷是评估元数据模式的常用方法 [31]。该评估研究已获 OFFIS 研究委员会批准（批准编号：2025E021）。

为创建元数据，我们使用了自己的工具软件元数据提取与整理软件（SMECS）[12, 17]。SMECS 是一款轻量级后端的网络工具，可根据提供的网址从 GitHub 和 GitLab 存储库中提取结构化元数据，例如名称和关键词。SMECS 会通过用户友好的接口展示提取的元数据及其他元数据元素，用户可在其中整理提取的元数据并输入其他元数据。创建的元数据可下载为 JSON 文件。这样，SMECS 实现了半自动化的元数据创建。SMECS 的第一个版本基于 *CodeMeta*。我们



图2：元数据模式评估问卷概览

综合将*ERSmeta*（V0.8）集成到 SMECS 中，以便用户创建特定领域的元数据 eir
能 源领域科研软件。

我们设计了这份问卷，以收集参与者使用 SMECS 和 *ERSmeta*创建元数据的体验。问卷可在 Zenodo 上获取 [15]。问卷的整体结构见 图2，由若干部分组成。

首先，参与者被要求提供一般的人口统计信息，例如从事研究工作的年限，以及他们在元数据和能源研究软件方面的经验。接下来，参与者需要打开 SMECS，并为自己的能源研究软件创建元数据。之后，他们需要回答有关使用 SMECS 体验的问题，包括系统可用性量表（SUS） [3]、基于技术接受与使用统一理论（UTAUT）对使用 SMECS 的行为意向进行评估的问题 [36]以及基于技术接受模型（TAM）对提取质量进行评估的问题2 [35]。此外，我们还询问参与者对某些功能构想的看法，并设置了开放性问题，供参与者以自由文本作答，谈谈 SMECS 的用法，并提出进一步改进建议。

随后，我们请参与者评估他们使用*ERSmeta*的体验，包括元素的数量和分布、元素的实用性、元素的描述，以及值词汇表的实用性。对于不同类别，我们根据技术接受模型 [6], TAM2 [35], TAM3 [34], UTAUT [36], 和UTAUT2 [36]设计了问题项。我们还添加了一些开放性问题，让参与者可以填写自由文本，例如提出需要增加的元素。最后，我们也请参与者提供所创建的元数据，以便进一步评估。为此，我们使用了单独的表单，以避免将问卷答案与所提供的元数据关联起来。

我们通过不同的邮件列表邀请参与者（Strommarkttreffen、新闻通讯NFDI4Energy、openmod、社交媒体（NFDI4Energy 在 LinkedIn 上），并直接联系已知能源研究软件的开发者，鼓励他们参与本研究。

元数据模式：ERSmeta

所开发的能源研究软件元数据模式（*ERSmeta*）如图3所示。该模式以 JSON Schema（JSON-LD 格式）和 SHACL（Turtle 格式）提供，可在 GitHub 上获取 [9] 以及 Zenodo 上获取 [10]，便于重复使用。

该模式由 86 个顶层元素组成，分为十个主题领域，有助于

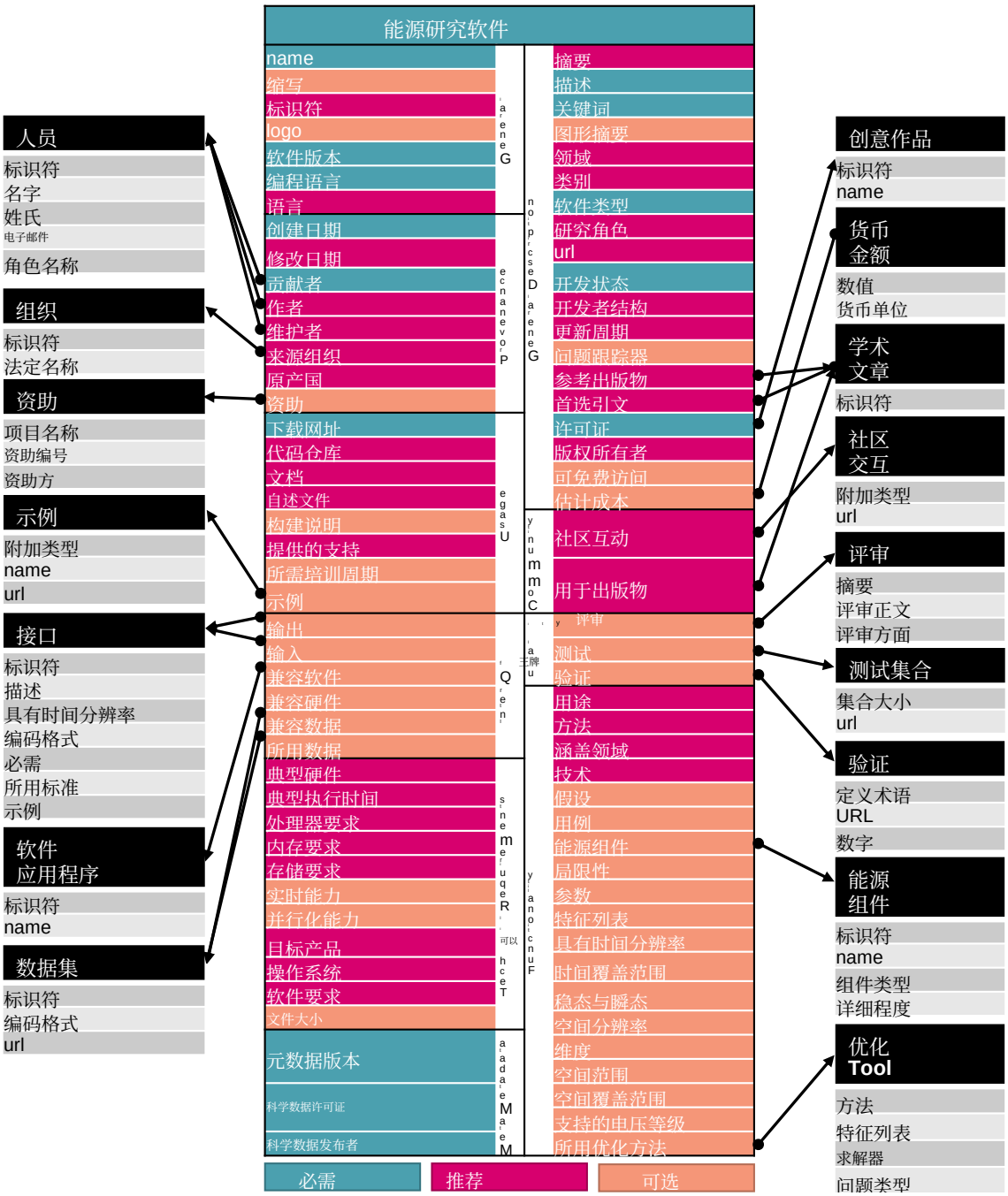


图3：*ERSmeta*：能源研究软件的元数据模式（评估前 ）

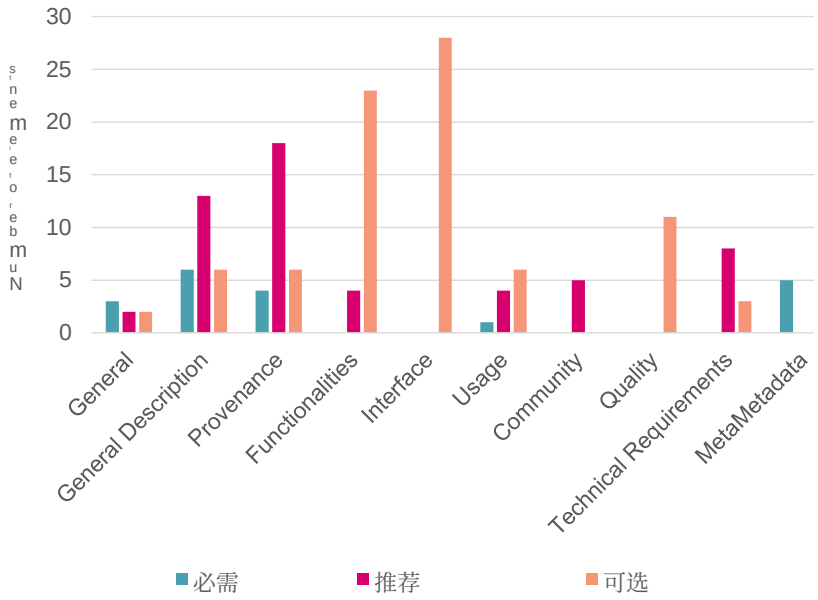
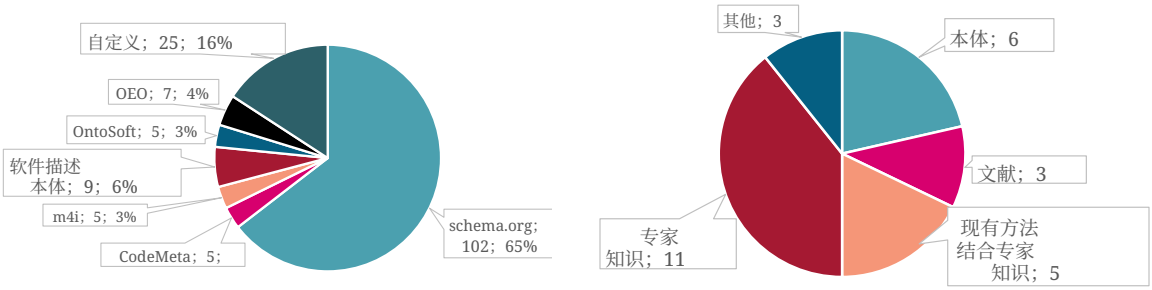


图4：必需、必需和可选元素的分布

更好地了解该模式。主题领域涵盖软件的不同方面，例如，功能领域侧重于软件功能的领域特定方面。顶层元素中有23个元素使用了需要由其自身定义模式的值类型，例如，作者。为此，另外定义了16个包含48个元素的元数据模式，如图3左右所示。考虑到每个顶层元素都至少需要填写一次，能源研究软件的元数据模式中总共包含158个元素。

为了说明各元素的重要性，我们将其分为三类：必需、推荐和可选。19个必需元素必须填写。它们有助于确保元数据达到一定的质量水平。示例包括软件的名称和编程语言。54个推荐元素能够很好地概述软件，因此模式的用户应填写这些元素。参考出版物和用途是推荐元素的示例。85个可选元素对于描述软件很有用，但重要性较低，例如资助、输入和输出。图4概述了必需、推荐和可选元素在十个主题领域中的分布情况。

我们复用了许多现有元素，以实现与其他元数据模式的高度互操作性。图5a概述了我们从哪些现有元数据模式或本体中复用了多少元素。大多数元素复用自 schema.org；它是一种广泛使用的本体，最初主要用于描述网络资源，以供搜索引擎使用。正如相关工作中所述，CodeMeta [28] 是科研软件最常用的元数据模式，并基于 schema.org。属于 CodeMeta 和 schema.org 的复用元素均计为 schema.org 的元素。我们还复用了 CodeMeta 中尚未纳入 schema.org 的一些元素，例如 developmentStatus。此外，我们从软件描述本体 [20], 中添加了科研软件通用方面的元素，例如 preferredCitation；并从 OntoSoft [22],中添加了 compatibleSoftware 等元素。我们还从 Open Energy Ontology [2], 中添加了涵盖领域等更具领域针对性的方面，从 m4i [26], 中添加了 method 等方面。此外，我们还定义了 25 个新元素，



(a) 元素的复用　　（b）值词汇表的来源

图5：ERSmeta 中重用的元素及所用值词汇表的统计

这些元素涵盖了 developerStructure 等通用方面，以及 energyComponent 等领域特定方面。

为提高所创建元数据的互操作性，我们为多个元素整合了值词汇表。值词汇表将某个元素允许的数值限制为受控词汇表，可以是预定义列表、本体或本体中的特定类 [37]。它们有助于创建便于比较和筛选的元数据。我们使用了不同来源的值词汇表，如图5b所示。我们纳入了本体中的现有概念，例如开放能源本体 [2]，以及语义名称权威来源中的概念，例如维基数据。我们还采用了文献中的分类，例如Hasselbring 等人提出的 研究角色 [24]，以及Shin 和 Ramanathan 提出的 实时能力 [33]。对于某些元素，我们还运用自己的专家知识扩展了现有来源，例如对于软件类型，我们采用并扩展了来自*BiotoolsSchema* [27]的值词汇表。此外，我们还运用自己的专家知识创建了一些术语预定义列表，例如支持的电压等级。

元数据模式评估

我们按照方法部分所述，通过调查问卷对 *ERSmeta* 进行了评估。调查问卷的答案已发布在 [14]，但由于隐私限制，调查参与者创建的元数据未公开。用于生成本节图表的分析脚本也已发布在 [11]。

共有十二人参与调查（n=12），并完成了整份问卷，其中十人提供了其创建的元数据供进一步分析。参与者在各自研究机构中担任不同职位，包括科研支持人员、博士生、博士后研究人员和其他科研人员。他们来自不同领域，其中一人来自经济学，五人来自计算机科学，六人来自工程领域。参与者对元数据模式的经验各不相同，其中两人此前完全不了解元数据模式。值得注意的是，这两位参与者对工具和元数据模式的评价更为苛刻。例如，图6所示的 SUS 分数表明，有元数据经验和没有元数据经验的参与者之间存在差异。在其他调查问卷条目中也观察到了这一趋势，这表明对元数据模式缺乏了解或兴趣可能会导致对 *ERSmeta*产生更负面的看法。

图6所示的 SUS 分数图 6 表明，参与者认为该系统的平均可用性尚可，但其可用性仍有进一步提升的空间。具体而言，与

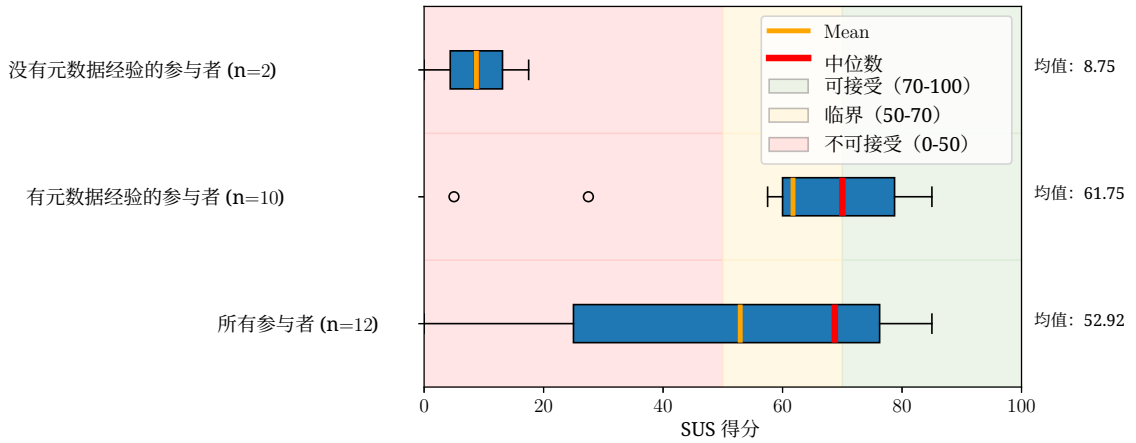


图6：熟悉和不熟悉元数据模式的参与者及总体的系统可用性量表（SUS）得分。

复杂性（“我觉得该系统复杂得不必要”）和频繁使用意愿（“我认为我会愿意频繁使用该系统”）相关的调查问卷问题（条目）显著拉低了 SUS 分数。我们认为，这种批评主要源于$ERSmeta$中的元素数量过多，因为开放式文本字段显示，参与者担心元素数量太多，且元数据创建过程耗时。这也反映在参与者对元素数量相关条目的回答中，如图 7所示。值得注意的是，参与者同意“元素数量太多”，这表明用户需要填写的元素应更少。同时，参与者也部分认同，元素总数有助于描述他们的科研软件。

然而，尽管参与者批评元数据模式包含的元素数量过多，他们仍认为这些元素有助于描述其能源研究软件，且与之相关。如图 7所示，参与者对这些元素的实用性评价为正面，表明它们能够有效地收集必要信息。参与者还认为所创建元数据的质量很高。开放式文本字段中的反馈进一步印证了这一点：参与者提到这些元数据相关且有帮助。在分析提交的元数据集时，我们发现只有三个元素（典型执行时间、示例和估算成本）未用于任何软件。这说明参与者认为大多数元素都相关且实用。当被问及是否仍缺少元素时，参与者大多表示不同意，说明现有元素集足以满足他们的需求。在开放式文本字段中被要求指出缺少的元素时，参与者仅提到四项属性（相关出版物、相关项目、数据来源和目标受众）。其中两项已纳入该模式，一项超出范围（数据元数据），只有第四项可能会被加入元数据模式。不同主题领域中元素的分布如图 8所示。值得注意的是，与功能相关的元素经常未被填写；我们认为这是因为某些元素可能并不适用于每一种能源研究软件。

元素描述的实用性呈现在图9中。参与者的评分显示平均结果为中立，这表明这些描述是一个良好的起点，但仍需进一步改进。具体而言，参与者在开放文本字段中指出，

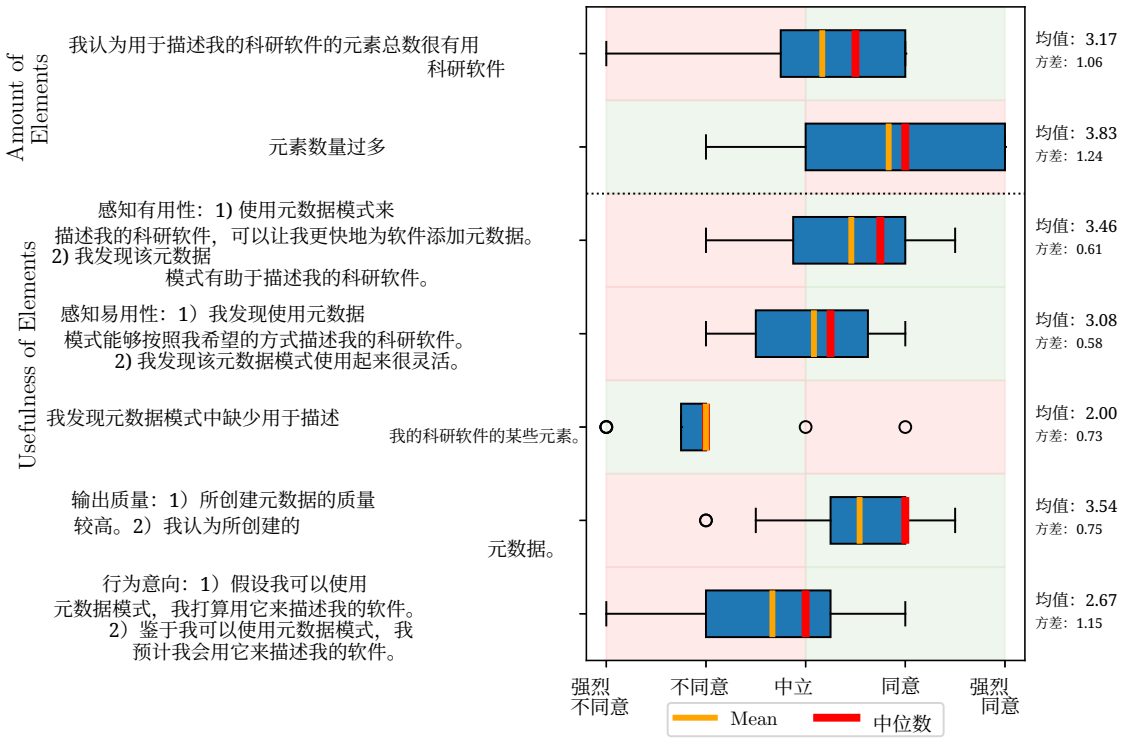


图 7：对元素数量和元素实用性的印象（n=12）。对于元素的实用性，同一构念（针对同一方面的问题）中的各项取平均值。计算均值和方差时，将李克特量表映射为数字（强烈不同意 $= 1$，……，强烈同意 $= 5$）

如果能提供更多示例，这些描述将有助于提升*ERSmeta*的可用性。

图9还展示了参与者认为值词汇表有效且有帮助。参与者对值词汇表的实用性给予了积极评价，表明它们是*ERSmeta*的一项宝贵资产。此外，在开放文本字段中，值词汇表也被列为元数据模式的一项优点。尽管参与者总体上对值词汇表感到满意，但他们也指出，仍需进行一些小幅扩展，使其内容更加全面。

对 *ERSmeta* 的评估显示，参与者的评价褒贬不一。虽然他们认为这些元素有用且相关，但元素数量过多以及系统的复杂性是显著的缺点。尽管如此，本次评估的结果将为改进和完善元数据模式及软件奠定宝贵基础，最终提升其充分满足能源研究软件社区中研究人员和研究软件工程师需求的能力。

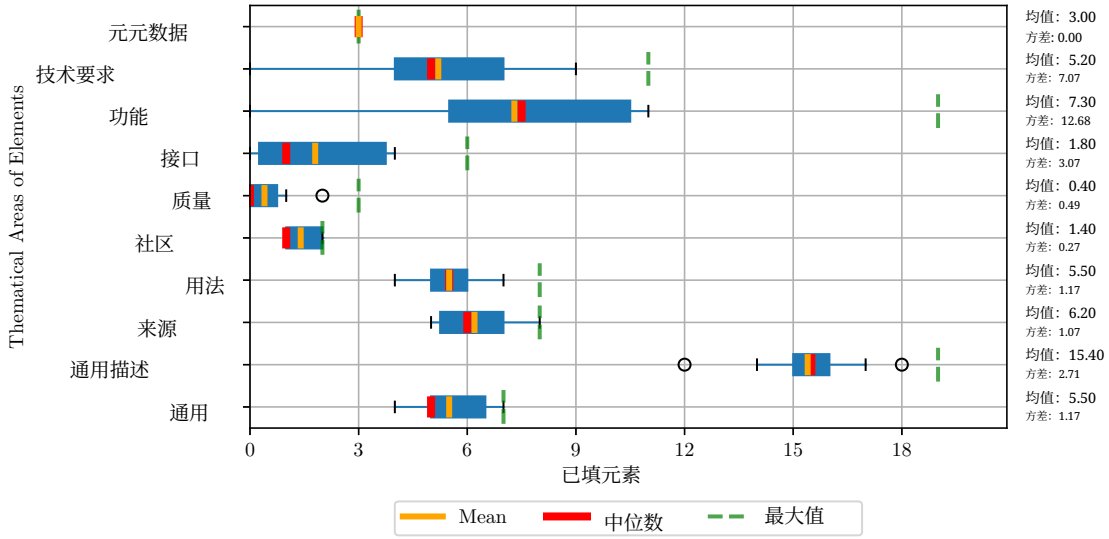


图 8：十个元数据集各主题领域的已填元素

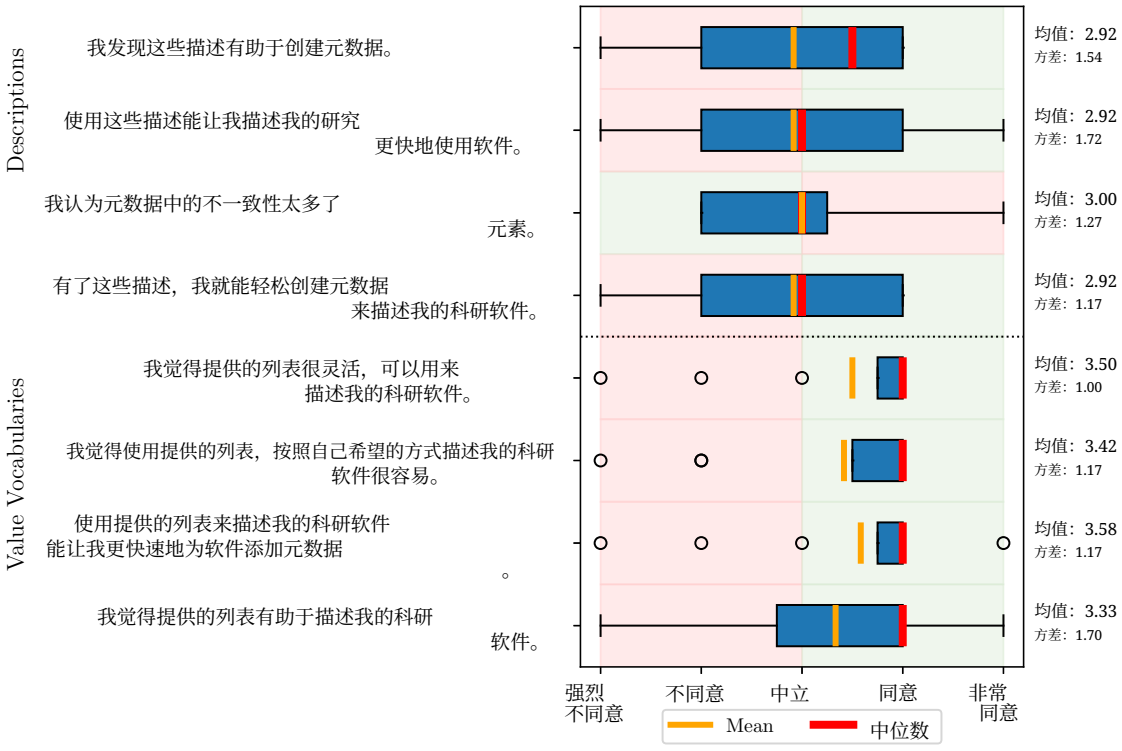


图9：描述和值词汇表的实用性（n=12）

讨论

ERSmeta 的开发与评估为了解在该领域创建元数据模式所面临的挑战与机遇提供了宝贵见解。然而，这项研究也存在一些应予以说明的局限性。

其中一项局限性是评估的地域范围有限，大多数参与者来自德国。此外，评估中涵盖的不同能源研究软件在研究主题和规模方面也较为有限，例如，大多数参与者都参与了规模较大的能源研究软件的开发。由于能源研究是一个高度跨学科领域，因此要纳入所有必需的视角颇具挑战性。虽然我们已经在需求分析中纳入了多样化的视角，但这可能仍不足以深入覆盖能源研究的所有部分。

此外，同时对工具（SMECS）和元数据模式进行评估可能引入了混杂因素，使得难以单独分析参与者的体验。尽管我们试图通过先询问他们对 SMECS 的看法来减轻这一影响（因为他们对 SMECS 的体验更为直接），但他们对 SMECS 和 ERSmeta 的印象逐渐交织在一起。具体而言，参与者对 ERSmeta 的看法在很大程度上受其在 SMECS 中呈现方式的影响，而后者又受 ERSmeta 的塑造。

尽管评估存在局限性，*ERSmeta* 仍展现出支持能源研究软件元数据创建和管理的潜力。值得注意的是，调查问卷参与者并未发现模式中存在任何重大缺漏，几乎没有人建议增添其他元素。此外，元素的描述被认为具有可用性，而调查问卷参与者尤其认可值词汇表，并将其视为*ERSmeta*的一项优点。

元数据模式的呈现方式已被证明是影响其对能源研究人员可用性的关键因素。尽管当前工具向用户展示了所有元素，但下一步的关键是减少展示的元素，仅保留最重要且最适用的部分。此外，探索根据提取信息的数量调整所展示元素数量的可能性也会有所裨益，从而有望带来更精简、更易用的体验。

ERSmeta 主要定义了新的领域特定元素，同时也定义了一些非领域特定元素，例如 communityInteraction。值得仔细分析这些非领域特定元素是否适用于通用地描述科研软件；如果适用，则应将其纳入*CodeMeta*等通用标准。

ERSmeta 是一种形式化的模式，因此比我们在相关工作中介绍的能源领域现有的非形式化方法更加灵活。此外，我们在评估中展示了，*ERSmeta* 复用了许多已建立的元数据模式和本体中的现有元素，从而使所创建的元数据能够与这些方法实现互操作。*ERSmeta* 基于一个清晰的流程，我们在模式开发部分对该流程进行了概述。然而，仍需进一步评估，以更深入地考察该模式的可用性和有效性。未来研究的一个潜在方向，是将该元数据模式集成到软件注册库中，理想情况下遵循 [18] 中所述的最佳实践。这样可以更详细地评估该模式的可用性，以及其支持能源研究软件搜索与发现的能力。

需要仔细关注科研软件通用元数据的未来发展，以确定是否有必要进一步修改*ERSmeta* 。例如，*ERSmeta* 无法以良好的机器可读方式描述接口；如果希望*ERSmeta* 同时便于人类使用（和创建），这将是一项挑战。Kuckertz 等人 [30] 介绍了一种方法：将接口信息包含在源代码中，然后自动提取到元数据中。

这可能是应对该挑战的潜在解决方案。由于*ERSmeta* 主要基于*CodeMeta*构建，因此应当能够快速将新方法应用于*ERSmeta*。

总之，*ERSmeta* 是创建能源研究软件元数据标准的重要第一步。要制定这一标准，还需要开展更多研究，以评估元数据模式的可用性，尤其是在软件注册库中的可用性。然而，*ERSmeta* 已经可以用于描述能源研究软件，因此有助于实现 FAIR 能源研究软件这一重要目标。

致谢

作者谨此感谢德国联邦政府、德国各州政府和联合科学会议（GWK）对NFDI4Energy联盟所提供的资助与支持。本项工作由德国研究基金会（DFG）资助，项目编号为 501865131，属于德国国家研究数据基础设施（NFDI，www.nfdi.de）的一部分。

参考文献

[1] Michelle Barker 等人：《研究软件 FAIR 原则简介》。载于：Scientific Data 9.1（2022年10月），第 622 页。 ISSN：2052-4463。 DOI：10.1038/s41597-022-01710-x。[2] Meisam Booshehri 等人：《开放能源本体简介：增强能源系统分析中的数据解释与接口》。载于：Energy and AI 5（2021年9月），第 100074 页。ISSN：2666-5468。 DOI：10.1016/j.egyai.2021.100074。[3] John Brooke：《SUS：一种快速简便的可用性量表》。载于：Usability Eval. Ind. 第 1 版。第 189 卷。CRC Press，1995年11月。 ISBN：978-0-429-15701-1。 DOI：10.1201/9781498710411-35。[4] Neil P Chue Hong 等人。《研究软件 FAIR 原则（FAIR4RS 原则）》。技术报告。FAIR4RS 工作组，2022年5月，第 32 页。 DOI：10.15497/RDA00068。[5] Mariana Curado Malta 和 Ana Alice Baptista。《社会团结经济元数据应用配置文件的开发过程》。载于：《开发元数据应用配置文件》。编者：Mariana Curado Malta、Ana Alice Baptista 和 Paul Walk。美国宾夕法尼亚州赫尔希：IGI Global，2017年，第 98–117 页。 ISBN：978-1-5225-2221-8。 DOI：10.4018/978-1-5225-2221-8.ch005。[6] Fred D. Davis：《信息技术的感知有用性、感知易用性与用户接受度》。载于：MIS Quarterly 13.3（1989年），第 319–340 页。 ISSN：0276-7783。 DOI：10.2307/249008。JSTOR：249008。[7] Stephan Druskat 等人。《引文文件格式》。技术报告。Zenodo，2021年8月。 URL：https://zenodo.org/records/5171937。[8] Muhammad Shoaib Farooq 等人：《高管商学中课堂录播系统（LCS）的接受与使用：扩展 UTAUT2》。载于：Interactive Technology and Smart Education 14.4（2017年11月），第 329–348 页。 ISSN：1741-5659。 DOI：10.1108/ITSE- 06-2016- 0015。[9] Stephan Ferenz。ERSmeta。2025年3月。 URL：https://github.com/NFDI4Energy/ERSmeta（访问于 2025年4月15日）。[10] Stephan Ferenz。ERSmeta。Zenodo。2025年10月。 DOI：10.5281/zenodo.17465772。

[11] Stephan Ferenz。ERSmeta 评估脚本。Zenodo。2025年12月。 DOI：10 . 5281 / zenodo . 17936836。（访问于 2025年12月15日）。[12] Stephan Ferenz 和 Aida Jafarbigloo。软件元数据提取与整理软件（SMECS）。Carl von Ossietzky Universit¨at Oldenburg。2025年5月。 URL：https://github.com/ NFDI4Energy/SMECS（访问于 2025年3月18日）。[13] Stephan Ferenz 和 Astrid Nieße：《通过引入基于元数据的注册库，提高能源研究软件的可发现性》。载于：ing.grid 1.2（2023年11月）。 ISSN：2941-1300。 DOI：10.48694/inggrid.3837。[14] Stephan Ferenz 和 Oliver Werth。能源研究软件元数据模式评估数据集。2026年1月。 DOI：10.5281/zenodo.17935022。（访问于 2026年1月5日）。[15] Stephan Ferenz 和 Oliver Werth。《用于评估能源研究软件元数据模式的问卷》。载于：（2025年10月）。 DOI：10.5281/zenodo.17465939。（访问于 2025年10月29日）。[16] Stephan Ferenz、Oliver Werth 和 Astrid Nieße：《实现 FAIR 能源研究软件的元数据方案要求》。载于：EASST电子通信 83（2025年2月）。 ISSN：1863-2122。 DOI：10.14279/eceasst.v83.2594。[17] Stephan Ferenz 等人：《SMECS：软件元数据提取与整理软件》。载于：EASST电子通信 85（2025年12月）。 ISSN：1863-2122。 DOI：10.14279/ eceasst.v85.2708。（访问于 2025年12月16日）。[18] Daniel Garijo 等人：《研究软件注册库和存储库的九项最佳实践》。载于：PeerJ计算机科学 8（2022年8月），e1023。 ISSN：2376-5992。 DOI：10.7717/peerj- cs.1023。[19] Daniel Garijo 等人：《OKG-Soft：包含机器可读科学软件元数据的开放知识图谱》。载于：2019年第15届国际电子科学会议（eScience）。2019年9月，第 349–358 页。 DOI：10.1109/eScience.2019.00046。[20] Daniel Garijo 等人。《软件描述本体》。2021年5月。 URL：https://w3id.org/ okn/o/sd/1.9.0（访问于 2025年7月25日）。[21] Yolanda Gil、Varun Ratnakar 和 Daniel Garijo：《OntoSoft：捕获科学软件元数据》。载于：第八届国际知识捕获会议论文集。K-CAP 2015。美国纽约州纽约市：计算机协会，2015年10月，第 1–4 页。ISBN：978-1-4503-3849-3。 DOI：10.1145/2815833.2816955。[22] Yolanda Gil 等人：《OntoSoft：科学软件的分布式语义注册库》。载于：2016年IEEE第12届国际电子科学会议（e-Science）。2016年10月，第 331–336 页。DOI：10.1109/eScience.2016.7870916。[23] Morane Gruenpeter 等人：《D4.4——EOSC 范围内研究软件推荐元数据标准指南》。载于：（2023年6月）。 DOI：10.5281/zenodo.10786147。[24] Wilhelm Hasselbring 等人：《多维研究软件分类》。载于：Computing in Science & Engineering（2025年），第 1–10 页。 ISSN：1558-366X。 DOI：10.1109/MCSE.2025. 3555023。[25] Ludwig H¨ulk 等人。开放能源元数据（OEMetadata）。2025年1月。 URL：https://github. com/OpenEnergyPlatform/oemetadata（访问于 2025年12月16日）。[26] Dorothea Iglezakis 等人：《使用 M4i本体对科学过程建模》。载于：研究数据基础设施会议论文集 1（2023年9月）。 ISSN：2941-296X。 DOI：10.52825/cordi.v1i.271。

[27] J. Ison 等人。“BiotoolsSchema：生物信息学软件描述的形式化模式”。载于：GigaScience 10.1（2021 年）。 ISSN：2047-217X。 DOI：10.1093/gigascience/giaa157。[28] Matthew B. Jones 等人。CodeMeta：软件元数据交换模式。版本 3.0。2023 年。 URL：https://codemeta.github.io/terms/（访问于 10/28/2025）。[29] Holger Knublauch 和 Dimitris Kontoskostas。形状约束语言（SHACL）。2017 年 7 月。 URL：https://www.w3.org/TR/shacl/（访问于 12/16/2025）。[30] Patrick Kuckertz 等人。“DataDesc：用于创建和共享科研软件接口技术元数据的框架”。载于：Patterns（2024 年 10 月），第 101064 页。 ISSN：2666-3899。 DOI：10.1016/j.patter.2024.101064。[31] Nikos Palavitsinis、Nikos Manouselis 和 Salvador Sanchez Alonso。“有机农业学习资源元数据应用程序描述文件的评估”。载于：元数据与语义研究。编者：Fabio Sartori、Miguel Angel Sicilia 和 Nikos Manouselis。计算机与信息科学通信丛书。柏林、海德堡：施普林格，2009 年，第 270–281 页。 ISBN：978-3-642-04590-5。 DOI：10.1007/978-3-642-04590-5_26.[32] Jan Schwarz 和 Sebastian Lehnhoff。“能源场景协同仿真中语义与领域知识的本体集成”。载于：第十一届知识发现、知识工程与知识管理国际联合会议论文集。奥地利维也纳：SCITEPRESS - 科学与技术出版公司，2019 年，第 127–136 页。ISBN：978-989-758-382-7。 DOI：10.5220/0008069801270136。[33] K.G. Shin 和 P. Ramanathan。“实时计算：计算机科学与工程的新学科”。载于：IEEE会议录 82.1（1994 年 1 月），第 6–24 页。 ISSN：1558-2256。DOI：10.1109/5.259423。[34] Viswanath Venkatesh 和 Hillol Bala。“技术接受模型3与干预措施研究议程”。载于：决策科学 - DECISION SCI 39（2008 年 5 月），第 273–315 页。 DOI：10.1111/j.1540-5915.2008.00192.x。[35] Viswanath Venkatesh 和 Fred D. Davis。“技术接受模型的理论扩展：四项纵向实地研究”。载于：管理科学 46.2（2000 年 2 月），第 186–204 页。 ISSN：0025-1909。 DOI：10.1287/mnsc.46.2.186.11926。[36] Viswanath Venkatesh 等人。“信息技术的用户接受度：迈向统一观点”。载于：MIS Quarterly 27.3（2003 年），第 425–478 页。 ISSN：0276-7783。 DOI：10.2307/30036540。JSTOR：30036540。[37] Marcia Lei Zeng 和 Jian Qin。元数据。第三版。伦敦：Facet Publishing，2022 年。 ISBN：978-1-78330-588-9。