Aerospace.Wikibase：迈向航空航天工程知识基础设施

Tim Wittenborg*, Ildar Baimuratov*, Jamal Eldemashki
*L3S研究中心，汉诺威莱布尼茨大学，德国汉诺威{tim.wittenborg, ildar.baimuratov}@l3s.uni-hannover.de

arXiv:2603.05192v1 [cs.DL] 5 Mar 2026

摘要——航空航天工程可从协作式知识管理中获益良多，但其基础设施仍然支离破碎。弥合这一差距对于减少当前重复劳动的做法，以及应对航空数据快速增长带来的挑战至关重要。本研究介绍了一个基于 Wik- ibase 构建的易用平台，旨在支持航空航天工程知识的协作共享与整理；该平台最初填充了近期一项系统性文献综述中的数据。作为坚实的基础，Aerospace.Wikibase 提供了 700 多个与流程、软件和数据相关的术语，并开放供未来扩展。将项目特定概念与持久、独立的基础设施相连接，使航空航天工程师能够围绕通用知识开展协作，同时避免项目资料被据为己有，从而在认识到工业界局限性的同时，推动为现代挑战提供可持续的解决方案。

索引词—**航空航天工程、Wikibase、知识管理、众包、关联数据、维基**

一、引言

航空航天业融合复杂系统与前沿研究，推动设计、制造和运营效率的进步。然而，灵活的数据与信息复用在航空航天领域的潜力尚未得到充分发挥。研究指出，尽管航空航天业长期以来一直采用基于知识的方法，但在互操作性、标准化和应用方面仍存在显著差距 [1], [2]。跨学科协作与知识共享的局限、统一数据标准的缺失，以及对FAIR原则 [3],**可**发现性、**可**访问性、**互**操作性和**可**复用性的遵循不足，都阻碍了基于知识的工程潜力的实现。此外，领域工程师与知识工程师之间的隔阂凸显了建立协作框架的必要性，以衔接实践与理论知识的应用[4], [5]。实施这些标准能够促进创新、优化工作流程并推进可持续发展目标。

我们的目标是找到一种更可持续的方式，使航空航天工程知识的众包不受项目周期和机构边界的限制。由此，我们提出了以下研究问题：

研究问题

航空航天工程师如何在开放访问且能够长期持续的基础设施上分享知识？

我们的贡献包括：**1)** 航空航天工程知识交流既有方法的概述。**2)** 基于近期关于基于知识的航空航天工程的系统性文献综述而建立的明确数据模型。**3)**将Aerospace.Wikibase 实现为开放且持久的知识基础设施。

本文其余部分的结构如下： 第二节介绍航空航天知识基础设施的研究现状，第三节阐述我们改进现有方案的方法。第四节详细介绍所开发的 Wikibase实例。这些发现将在第五节中讨论，第六节则总结全文。

II. 背景与相关工作

Edwards [6]将知识基础设施（KI）定义为“……由人员、制品和机构构成的稳健网络，用于生成、共享和维护关于人类世界与自然世界的特定知识”。维基百科社群就是一个显著的例子：维基媒体基金会（机构）提供数字基础设施（制品），将“来自学术界、工业界、商业界和公众等不同领域的各类参与者、组织和观点汇聚在一起” [7]（人员）。航空航天工程领域中的各类机构很容易识别，既有美国国家航空航天局（NASA）和西门子这样的组织，也有 AGILE[1]、可持续与高能效航空（SE2A）[2],和SynTrac[3]这样的项目。制品，尤其是可公开访问的制品，则难以识别得多，因为 Harvey 和 Holdsworth [1]指出，“航空航天与国防工业或许是所有行业中监管最严格的，受到众多国家和地区机构的监管”。尽管有人针对航空航天领域提出了各种知识管理系统和本体 [5], [8]–[10], ，但据我们所知，目前还没有开放访问的协作式航空航天知识库。最接近的例子可能是维基百科上的航空航天工程术语表[4] 。其他所有开放协作式知识库

[1]https://www.agile4.eu/[2]https://www.tu-braunschweig.de/en/se2a[3]https://www.trr-syntrac.com/[4]https://en.wikipedia.org/wiki/航空航天工程术语表

本研究由德国研究联合会 SE2A卓越集群资助。

似乎都在各自项目期限结束后停止维护了 [11]。

在 [12], 对基于知识的航空航天工程进行的系统性文献综述中，作者也表达了这一观点：知识通常未被形式化，而是分散在人力资源层面、报告、会议和直接沟通中，并且不对外开放。为解决这一问题，作者提取了一个涉及航空航天工程流程、软件和数据的术语数据集，并将其导入开放研究知识图谱（ORKG） [13], ——一个支持以结构化、机器可读的方式呈现研究贡献的平台。这使跨学科比较、检索和关联科学知识变得更加容易。尽管这一知识基础设施开放可用，便于协作，但 ORKG 以研究贡献为中心的模型对工程师而言似乎过于受限；他们需要一个能够规范领域术语及其相互关系的系统。

因此，建议转而采用专用的 Wikibase.Cloud 实例。Wikibase[5] 是一款用于存储和管理结构化、关联式和协作式数据的免费开源软件，与维基数据类似。与 ORKG 一样，Wikibase 以结构化、机器可读的方式呈现内容，但不受特定数据模型的限制，适合分享未以研究贡献形式呈现的航空航天工程知识。不过，Wikibase 只是一款软件；要用它来分享航空航天工程知识，就必须进行适当的配置和部署。

III. 方法

我们的目标是通过开发Aerospace.Wikibase.Cloud 平台，弥合航空航天领域知识基础设施现状与理想之间的差距。这主要包括两个步骤：（1）数据获取和建模；（2）Wikibase实例的实施与填充。

A. 数据获取和建模

我们重新使用了系统性文献综述[12],中的数据，该综述确定了700多个航空航天工程领域术语及其相互关联关系。我们设计了一个语义模型，用于构建正在开发的Aerospace.Wikibase中的这些数据。该模型包含若干类，这些类源自以下研究问题：“哪个航空航天工程流程由哪个软件使用哪个数据，以哪种格式，并采用哪个模式完成？”对于数据格式规范

对于数据项，我们从信息制品本体 [14]中导入相应的类。数据模型类与一个维基数据实体[6]相关联。流程类从基础形式本体导入，软件类则从软件本体 [15]导入。除了这些类之外，我们还遵循 ORKG 数据模型，引入了贡献类，用来表示展示航空航天领域研究贡献的学术出版物。

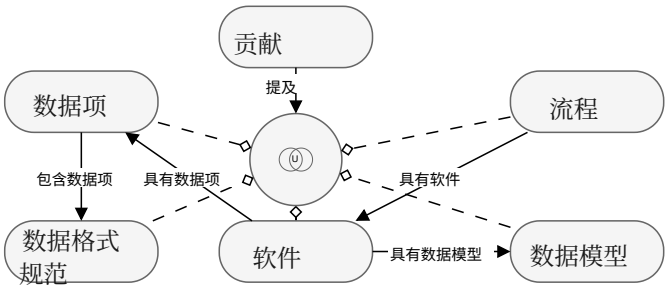


图 1. 数据模型

该模型包含若干关系。其中大多数关系定义了类的实例之间的连接，例如“具有流程”“具有软件”“具有数据项”“具有数据格式规范”和“具有数据模型”。此外，模型还包含“提及”关系，用于连接贡献与其中引用的领域实体。我们引入的另一种关系是“具有部分”，例如，它将流程与其子过程连接起来。最后，我们使用“实例”关系，在领域实体之间建立实例—类关系（例如，三维模型是模型的一个实例）。

该模型还采用了若干注释属性。第一个，别名，用于存储同一实体的替代名称或同义名称。第二个，来源，表示信息的来源，例如文献综述、调查或访谈。最后，实体与维基数据统一资源标识符对齐，以增强 Wikibase 内容的 FAIR 合规性。图 1 展示了所设计的数据模型，该模型以 OWL 文件形式提供。

B. Wikibase 实现与填充

航空航天知识库构建于专用的 Wikibase实例之上，采用包含系统部署与自动化数据摄取的结构化工作流程。除 Wikibase 的通用功能外，我们还开发了一种定制方法，以确保数据被一致地导入，在利用 Wikibase 灵活性与用户界面的同时保留类层次结构。

1) 映射到 *Wikibase*：数据模型的 OWL构造按如下方式映射到 Wikibase 数据模型：

- **类** → Wikibase项目，其中 rdfs:subClassOf 映射到子类属性。
- **个体** → 通过实例关联的条目，其中每个 rdf:type 三元组都进行了相应转换。
- **对象属性** → 具有适当数据类型的 Wikibase 属性。
- **标签、描述和别名** →用于提高可搜索性的 Wikibase 字段。
- **外部标识符和来源信息** → 维基数据 URI、DOI 和专用的 本体 *IRI* 属性确保了对源本体的追溯能力。

Wikibase 将条目和属性都作为具有标签、描述和使用文档的一等实体公开。我们的映射确保了诸如实例

[5]https://wikiba.se/ [6]https://www.wikidata.org/wiki/Q1172480

、子类、本体 IRI 或维基数据 URI 等陈述具有明确的语义，使导入的本体在用户界面中易于理解。

2) 数据摄取：导入流程使用 Python 实现，采用 rdflib 进行解析，并使用 WikibaseIntegrator 摄取数据，基于 Stehr 最近提出的方案 [16]。属性按照 Wikidata.org 的惯例手动创建（明确的数据类型、描述性标签和使用说明），因为它们定义了模式，必须保持稳定。随后，这些属性被引用到映射字典中，该字典将本体注释与 Wikibase 中对应的内容关联起来，确保能够一致地重复使用。

数据摄取分为三个阶段。首先，导入所有类，建立子类层级。然后，创建所有个体，并通过“实例”将其与所属类关联，补充本体_iri 等元数据，并为其分配以 Wikibase 别名形式存储的别名。完成这些步骤后，我们使用了 Wikibase 查询服务。该服务在用户界面中提供 SPARQL 端点，我们通过运行 SPARQL 查询来验证标签、层级链接和标识符是否正确导入。

3) 数据质量与错误处理：数据摄取过程中偶尔会发生暂时性故障，例如数据库锁定或保存失败错误，这些错误由复制延迟或 API 超时引起。为缓解这些问题，机器人被设计为幂等：创建每个实体前，都会先在本地缓存中检查；如果有外部标识符，则优先通过这些标识符进行匹配。这样可确保在故障后重新运行流程时不会产生重复实体，而是从一致的状态继续摄取。每个批次处理完毕后，系统都会自动运行 SPARQL 查询以验证数据质量，检查是否存在缺失标签、重复条目，以及缺少子类或实例链接的孤立类。这些质量控制循环能够系统地验证内容完整性和结构完整性。

4) 增量可扩展性：该系统在设计时便考虑了集成初始导入之外的其他本体。通过使用新的注释属性或对象属性扩展映射字典，并在模式中创建任何缺失的 Wikibase属性，便可纳入新的数据来源，而不会干扰现有知识库。这种增量式方法使知识图谱能够以受控且一致的方式不断扩展，从而确保其具备长期的可扩展性和适应性。此外，由于所有本体都必须遵循相同的属性映射策略，该方法还可以防止模式碎片化。

IV. 结果

生成的 Aerospace.Wikibase 实例共包含 14,886 个节点，构成 37,897 个三元组，其中 8,202 个（55%）节点是国际化资源标识符，6,660 个（45%）是字面量。字面量分为四种数据类型：3,607 个（54%）属于 `rdf:langString`，1,271 个属于 `xsd:integer` （19%），954 个（14%）属于`xsd:string`，828 个（12%）属于 `xsd:dateTime`。在这些三元组中，15,513 个（41%）涉及 III-A 小节中描述的领域属性，10,393 个（27%）与 Wikibase 本体，包括排名、时间戳、陈述、站点链接 和 标识符；11,990 个（32%）则具有其他属性，例如 `rdf:type`、`skos:altLabel`、`schema:dateModified`、`schema:version`、`rdfs:label` 和 `schema:description`。领域条目和属性的使用情况汇总于表 I。Aerospace.Wikibase 页面数接近 1,000，按页面数计算，其排名位于 Wikibase 实例的前 13%。

表 I Aerospace.Wikibase 中的领域实体。请注意，“边”统计的是边的数量，而不是边类型的数量。

Type		数量	总计
Node	实例	637	710
	类	7	
	歧义项	66	
Edge	对象属性（内部）	1,533	4,466
	对象属性（外部）	1,231	
	注释属性	1,533	

V. 讨论

Wikibase 现在为数据访问和数据整理提供了更灵活的环境。导入的数据严格区分为实例和类。虽然这种做法仍可将软件实例与流程实例等连接起来，但这些内容原本也可能会用类和子类表示。在 Wikibase 中，这种严格的建模方式被逐渐形成的多层结构所取代，从而允许条目之间建立实例关系，而这些条目此前本身被视为实例。例如，三维模型是数据项和模型的一个实例，或可能是其子类。可以推测，这些结构最终大多会融合为一个多层类与元类体系。放宽这些限制是 Wikibase 环境的主要优势之一，此外，它还拥有简单、熟悉的用户界面。由于它与维基数据相连，却不隶属于维基数据，工程师如今可以在无需熟悉维基数据额外惯例的情况下，对其领域知识进行建模和扩展。

然而，提供技术上的可能性或许无法克服工程领域中显而易见的、根深蒂固的反对自由共享知识的倾向，尤其是在航空航天工程领域。多年来的研究指出，该领域是“所有领域中监管最严格的” [1]，其工作重点是在这种明显的抵制态度赶上之前，获取并隔离知识，以确保经济优势，甚至军事优势。然而，指出这一根深蒂固立场的研究也表明，要应对航空航天领域日益复杂且时间紧迫的挑战，跨越系统、领域和机构边界进行协作正成为一种必要。AGILE、SE2A 或 SynTrac 等项目表明，这些解决方案并非不受欢迎，而是解决未来问题所必需，甚至可能不可或缺。

我们的 Wikibase 倡议并非建立在“每位工程师都会向同一个平台贡献内容”这一假设之上，而是致力于让他们能够协作使用通用工件

和知识库。维基百科每次编辑对应 2000 次阅读[7]，也就是说，内容被使用的频率远高于贡献的频率。只要领域工程师就一种通用语言达成一致，他们的私下协作就能通过公开维护的术语体系关联起来。Aerospace.Wikibase 展示了一个示例：可自主管理、又从根本上互联互通的知识图谱，如何成为满足特定领域需求的核心知识共享基础设施。理想情况下，航空航天领域的代表性中央机构也会采用此类 Wikibase，例如已将 MediaWiki 用于内部知识管理的美国国家航空航天局、美国航空航天学会（AIAA）、德国航空航天中心（DLR）、代尔夫特理工大学（TU Delft）等机构。即使没有这些机构参与，单个工程师也可以为 Aerospace.Wikibase 或自己的 Wikibase 填充内容，或者立即开始为维基数据和 ORKG 贡献内容。前人作者所要求的知识基础设施，如今已触手可及。

研究发现

航空航天工程师可以在专门设立、开放访问的 *Aerospace.Wikibase.Cloud* 实例中分享知识。这一可持续的基础设施不受项目资金和领域限制，并提供通用知识节点，以连接私有系统。

VI. 结论与未来工作

本研究通过开发专用的 Wikibase 实例，解决了航空航天工程领域缺乏协作式、开放且可持续的知识基础设施这一问题。我们使用一项近期系统性文献综述的数据，该综述识别了领域专属术语及其相互联系，并将这些信息建模为 OWL 本体，复用了基础形式本体、信息制品本体、软件本体和维基数据中的类。随后，我们实施了一套工作流程，包括系统实现和用于自动化数据摄取的程序化接口。该工作流程确保数据导入一致、保留层级结构，并优化用户交互。最终的 Wikibase 包含 14,886 个节点，构成 37,897 个三元组，其中分别有 1,025 个节点和 4,759 个三元组与领域专属实体相关。这表明，即使只投入少量整理工作，也能大幅提升相关资源的可用性；此外，Wikibase 生态系统中已有一百多个其他平台蓬勃发展。因此，我们认为未来工作主要有两个方向：一是拓展数据范围并提高数据质量，二是推进航空航天工程知识基础设施建设。前者可通过整合基础数据集之外的其他制品来实现；而要推进知识基础设施建设，则需要个人和机构充分利用这些制品。要克服航空航天工程知识管理面临的挑战，就必须开展这项未来工作。

致谢

AI 工具使用声明：在准备本研究的过程中，作者使用了 **DeepL**、**Grammarly（浏览器插件）**、**LanguageTool（浏览器插件）**，用于：**翻译文本、语法和拼写检查、改写和重新措辞**，具体依据 CEUR生成式人工智能使用分类体系[8]。使用该工具/服务后，作者对内容进行了必要的审阅和编辑，并对出版物的内容承担全部责任。

参考文献

[1] D. Harvey 和 R. Holdsworth，《航空航天业的知识管理》，载于 *IPCC 2005. Proceedings. International Pro- fessional Communication Conference, 2005.*，第237–243页，2005年7月。ISSN：2158-1002。[2] I. Sanya、E. Shehab 和 D. Lowe，《航空航天设计工程中的语义知识管理挑战》，载于 *DS 68-6: Proceedingsof the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 6: Design Informationand Knowledge,Lyngby/Copenhagen, Denmark,15.-19.08. 2011*，第241–248页，2011年。[3] M. D. Wilkinson、M. Dumontier、I. J. Aalbersberg、G. Appleton 等，《科学数据管理与管理的 FAIR指导原则》，*Scientific Data*，第3卷，第160018页，2016年3月。编号：1。出版商：Nature出版集团。[4] E. Haney，《航空航天系统设计与预测中的数据工程》。博士论文，2016年。[5] T. Procko 和 O. Ochoa，《利用关联数据进行知识管理：针对航空航天业的提案》，*Available at SSRN 4703237*，2022年。[6] P. N. Edwards，《《一台庞大的机器》：计算机模型、气候数据与全球变暖政治》。麻省理工学院出版社，2010年。[7] H. Karasti、F. Millerand、C. M. Hine 和 G. C. Bowker，《知识基础设施：第一部分》，*Science & Technology Studies*，第29卷，第1期，第2–12页，2016年。[8] A.-S. Dadzie、R. Bhagdev、A. Chakravarthy、S. Chap man、J. Iria、V. Lan- franchi、J. Magalh˜aes、D. Petrelli 和 F. Ciravegna，《将语义网技术应用于航空航天工程中的知识共享》，*Journal of Intelligent Manufacturing*，第20卷，第611–623页，2009年10月。[9] R. Curran、W. Verhagen 和 M. Van Tooren，《在航空航天生产中整合多学科工程知识的 KNOMAD 方法》，载于 *48thAIAA Aerospace Sciences Meet-ing Including the New Horizons Forum and Aerospace Exposition*，（佛罗里达州奥兰多），第1315页，美国航空航天学会，2010年1月。[10] I. Sanya 和 E. Shehab，《航空航天业中用于开发平台无关的知识工程系统的本体框架》，*International Journal of Production Research*，第52卷，第6192–6215页，2014年10月。[11] H. Karasti、K. S. Baker 和 F. Millerand，《基础设施时间：协作开发中的长期问题》，*Computer Supported Co- operativeWork(CSCW)*，第19卷，第377–415页，2010年8月。[12] T. Wittenborg、I. Baimuratov、L. K. Franz´en、I. Staack、U. R¨omer 和 S. Auer，《基于知识的航空航天工程——系统性文献综述》，*arXivpreprintarXiv:2505.10142*，2025年。[13] S. Auer、A. Oelen、M. Haris、M. Stocker、J. D'Souza、K. E. Farfar、L. Vogt、M. Prinz、V. Wiens 和 M. Y. Jaradeh，《利用知识图谱改善科学文献的获取》，*Bibliothek Forschung und Praxis*，第44卷，第3期，第516–529页，2020年。[14]《信息制品本体（IAO）》。最后访问日期：2025年9月10日。[15]《软件本体》。最后访问日期：2025年9月10日。[16] N. Stehr，《用于提供有关科学视频与播客的信息的数字知识基础设施》，2025年4月。出版地：汉诺威：Gottfried Wilhelm Leibniz Universit¨at。

[7] https://en.wikipedia.org/wiki/维基百科:Statistics

[8] https://ceur-ws.org/GenAI/Taxonomy.html