arXiv:2603.05984v1 [cs.DL] 6 Mar 2026

推动科学传播与航空航天工程领域的知识基础设施建设

Tim Wittenborg，L3S Research Center，汉诺威莱布尼茨大学，
德国汉诺威 tim.wittenborg@l3s.uni-hannover.de

摘要——**知识基础设施被定义为由人员、人工制品和机构构成的稳健网络，能够生成、共享并维护特定知识。然而，许多领域支离破碎，远未形成稳健的网络，例如科学传播或航空航天工程。尽管已有 FAIR（可查找、可访问、可互操作、可重用）数据管理工具，但这些工具在上述领域的应用仍然有限。从复杂的异构数据格式、缺乏结构化支持，到公开存在的反协作激励或法律障碍，诸多挑战阻碍了这些工具的应用。本博士研究以科学传播和航空航天工程为应用案例，阐述如何促进尚不完善的知识基础设施建设。通过分析这些问题并确定现有解决方案，可以实施并评估由工具支持、旨在实现协作基础设施的工作流。这些工作流包括用于信息提取和处理的人机协同人工智能（AI）支持型工作流、基于维基和知识图谱的数字图书馆，以及由利益相关方需求驱动的界面。尽管这些用于工作流自动化和知识表示的工具展现出良好前景，但仍面临重大挑战。未来工作必须超越技术问题的解决，处理社会和法律障碍，以释放特定领域的潜力。此外，任何领域新兴知识基础设施的倡导者，均可借鉴本研究成果，促进现有知识的联网。**

索引词——**知识基础设施、科学传播、航空航天工程、数字图书馆、视听内容、维基、媒体标注、工作流自动化**

I. 研究动机

科学传播与航空航天工程领域的研究面临一个共同问题：有价值的知识持续产生，却分散在不同格式、存储库和组织边界之中。这种碎片化削弱了 FAIR 数据原则的各个方面：**可发现性**、**可访问性**、**互操作性**和**可重用性** [1]。从简单的视频搜索到复杂的多学科设计优化，依赖数据的流程都因此耗费了不必要的资源。与此同时，最先进的技术不断进步，既包括知识表示技术（尤其是知识图谱），也包括工作流自动化，特别是人在回路中的人工智能（AI）集成。本论文旨在通过聚焦于兼顾自动化与人工验证的解决方案，并激励可持续数字图书馆中的协作式策展实践，推动新兴知识基础设施的发展。本文的核心研究问题（RQ）如下：

本研究由DFG SE2A卓越集群资助。

研究问题1： 哪些最好能独立于特定领域的工具，尤其是数字图书馆和工作流自动化工具，可用于促进知识基础设施的发展？**研究问题2：** 由模块化工具支持的工作流框架，在多大程度上能够满足提取、对齐和表示相关知识的要求？**研究问题3：** 在科学传播或航空航天工程领域应用知识基础设施工具时，会产生哪些特定领域的要求？**研究问题4：** 数字图书馆如何满足科学传播和航空航天工程等碎片化领域特有的知识基础设施要求？

II. 背景与相关工作

Edwards 将知识基础设施（KI）定义为 [2]“由人员、制品和机构构成的稳健网络，能够生成、分享并维护有关人类世界和自然世界的特定知识”。典型案例包括政府间气候变化专门委员会（IPCC）和维基百科社区，后者由个体贡献者、正式与非正式组织、子项目和项目、数字图书馆、工具等组成。这些网络的特点是能够超越任何单个项目的存续时间 [3]。然而，在科学传播和航空航天工程等领域，此类稳健网络并不存在，相关工作也较为零散，尤其是在考虑 FAIR数据原则时 [1]。但是，要将这些工作整合起来，不仅需要技术组件和标准，还需要关于溯源、治理和社区采纳的规范，以促进参与，并确保长期质量和再利用。因此，本节将重点介绍共同的制品，尤其是 数字图书馆，并说明它们在两个领域的应用： 科学传播 和 航空航天工程。

A. 数字图书馆

数字图书馆（DL）的定义各不相同，从“一套电子资源及相关技术能力，用于创建、搜索和使用信息”到将参与其中的工作人员和社区也纳入定义 [4]。着眼于其载体组成部分，现代数字图书馆越来越依赖结构化、机器可操作的数据，例如知识图谱。此类基于图的数字图书馆的著名例子是维基数据，它据称是全球规模最大的协作编纂知识图谱 [5]。如此庞大的数字图书馆很快便会面临扩展问题，在满足其社区需求之前就达到极限

[1], , 因此相关工作已从单体架构转向联合生态系统。围绕维基数据构建的这一生态系统包含一千多个专业化的 Wikibase 实例，称为 Wikiverse；而 Wikiverse 也是关联开放数据（LOD）环境的一部分。在这个 LOD 环境中，另一个值得关注的知识图谱是开放研究知识图谱（ORKG） [6],, 它专注于研究贡献的语义表示。这些载体对科学方法具有普遍意义，也可能与每一种知识基础设施息息相关。维基数据和 ORKG 每天都在推动数字图书馆的发展，因此本研究致力于帮助尚不发达的基础设施跟上这一不断演进的潜力。

B. 科学传播

Burns 等人 [7]将科学传播（SciCom）定义为“运用适当的技能、媒介、活动和对话，促使公众对科学产生一种或多种以下个人反应（*AEIOU* 元音类比）：认知、愉悦、兴趣、形成观点和理解”。其目标是提升公众理解、接受度、信任和支持，同时提供一个平台，收集公众对有价值的研究目标及其应用的广泛反馈、地方知识和公民需求 [8]。科学传播知识基础设施（SciCom KI）是专门服务于科学传播的知识基础设施，涵盖相关人员、机构和资料，例如视频和播客。MacKenzie [9] 整理了 952 个科学播客，wissenschaftspodcasts.de 386；Kikkawa 等人 [10], [11]整理了 23 万个视频；世界讲座项目[2] 59,634, TIB-AV 门户 [12] 50,690。此类视频经过细致研究，分析并标注了数百种不同的质量和属性 [13]。这些二手数据通常存储在孤立存储库中，与原始媒体相互分离，除非将其纳入 FAIR 基础设施——例如，通过在 TIB AV 门户中对媒体进行许可 [12]，或满足维基数据的知名度标准[3] [5]。然而，在评估共存于同一传播渠道、但在质量、目的、范围、评价标准、可访问性和创收方式上各不相同的信息时，捕获这些知识尤为重要 [14], [15]。

C. 航空航天工程

航空航天领域依赖形式化的工程流程和基于知识的工程（KBE），以捕获专家规则，并自动执行重复性的设计、分析和文档编制任务 [16]–[18]。随着系统涉及的学科日益多元，复用异构知识变得至关重要，这些知识涵盖显式规则和成果，也包括最佳实践与指南，有助于管理复杂性并加快迭代 [19], [20]。语义技术和本体日益广泛地用于提升互操作性，并支持在不同工具和团队之间进行自动推理与发现 [21]，[22]。核心交换格式也在不断开发，例如 CPACS 或 CM- DOWS [23]。然而，早期的领域研究早已列出构建下一代航空航天知识库的前提条件和挑战：“研究表明，改进知识的结构和表示将有助于提升工程设计活动的能力”[24]。研究还特别指出，“工程师往往花费大量时间寻找问题的解决方案，而这些问题或许早已得到解决。”，部分原因是“工程师有时不愿分享文档”。Procko 和 Ochoa 指出，对知识产权和治理的顾虑进一步限制了分享 [25] ；他们在文献综述中引用了 Harvey 和 Holdsworth 的研究 [26]：“除了这些挑战之外，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。”

本论文旨在总结这些发现，并提出可行的解决方案，以在这种易于共享的环境中构建知识基础设施。模块化解决方案需要兼顾基于知识的工程的要求与专家监督，并保留限制性治理需求。目标是在科学传播和航空航天工程领域构建协作式、平台无关的关联开放数据基础设施。

III. 研究方法与方法论

本研究采用混合方法研究，结合系统性文献综述（SLR）、需求获取和行动研究驱动的软件开发。其基础如图 1所示：开发一条数据处理流水线，该流水线可用于促进系统性文献综述，同时也包含可重新用于其他应用的模块。一项航空航天领域的系统性文献综述可以展示该流水线的能力，并为未来建立航空航天工程知识基础设施的开发提供参考。然而，这条基础流水线也可用于处理视听媒体，例如视频和播客，并以类似于系统性文献综述的方式丰富语料分析。将内容按此方式结构化后，还可实现更多应用，例如针对以类似方式处理的知识库开展计算事实核查。

这种方法是在动态发展中形成的，旨在满足不断出现的需求并利用已发现的协同效应。相关研究尤其对这一发展过程产生了影响，从 CMDOWS 等旨在将工作流规范化的领域特定方法，到联邦学习等外部启发；后者旨在为同样支离破碎的环境带来协调一致的改进。为推进开发，采用了行动研究，并进行了领域专家评估。评估通过问卷调查和访谈开展，同时也补充了文献综述的发现。

IV. 初步结果

A. SWARM-SLR

SWARM-SLR（面向机器可操作系统性文献综述的精简工作流自动化）是一种模块化、以需求为导向的工作流，可将 65 项系统综述

[1]https://www.wikidata.org/wiki/Wikidata:WikiProject 维基数据的局限 [2] https://world-lecture-project.org/ [3] https://www.wikidata.org/wiki/Wikidata:Notability

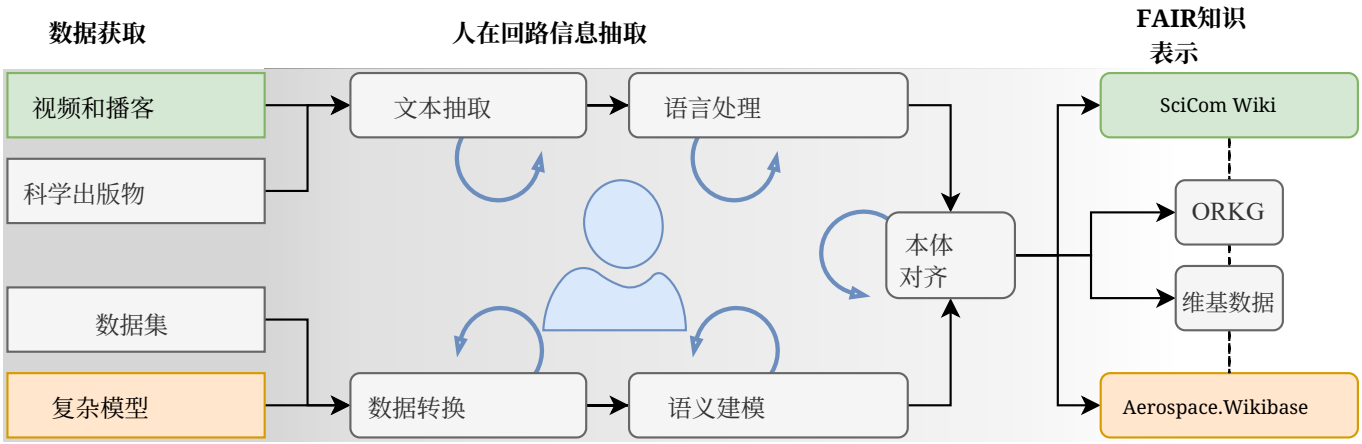


图1. 工作流程概览：对数据进行处理，以实现 FAIR 知识表示。视听内容也可以像文本出版物一样进行处理。基于知识的工程模型和数据集同样能够从语义信息提取和作为关联（开放）数据的知识表示中获益。

需求形式化，并将参考文献管理器、自然语言处理（NLP）流程和数字图书馆等任务专用工具映射到从规划到报告的完整系统综述生命周期。它由四个 Jupyter笔记本构成，通过人在回路验证引导用户完成系统综述，从而实现可互操作、可复用的半自动化文献综述。论文已于 [27] 发表。

a) ExtracTable：ExtracTable改进了SWARM-SLR的语料库创建和数据提取工作流。它结合大语言模型（LLM）与人在环路中定义的模式，旨在支持下游集成至知识图谱（KG）。论文已被接收 [28]。

B. 基于知识的航空航天工程

我们使用SWARM-SLR筛选了1,000多份文献，并对164篇核心文章进行了详细分析，以描绘航空航天领域当前基于知识的工程（KBE）实践。我们发现该领域拥有坚实的技术基础，尤其是CPACS和CMDOWS，但应用情况较为零散。现有文献建议优先采用开放交换格式和协作数据环境，并作出承诺以打破数据孤岛。数据复用应兼顾航空航天工程严格的治理需求。期刊文章正在审稿中 [29]。

C. SciCom Wiki

SciCom Wiki是一个基于Wikibase的开放数字图书馆，收录科学视频和播客。它根据利益相关者需求设计，并采用原型微服务，旨在提升非文本媒体的可发现性、标注和FAIR元数据采集。早期评估（问卷调查、访谈和原型测试）表明，它显著提升了可发现性和策展能力，同时也指出了更广泛推广所需解决的法律、用户引导和可扩展性问题。论文已被接收 [30]。

a) 计算事实核查：一种计算事实核查原型将基于大语言模型的陈述提取与针对精选真实知识图谱的符号验证相结合，从而为媒体内容生成可解释的科学准确性评分。专家和用户评估表明，这种方法能够提供有用的真实性指标，但目前受到非 FAIR 的真实值表示和低效提取流程的限制。论文已被接受 [31]。

五、挑战与讨论

本研究旨在探讨工具支持的工作流如何推动 FAIR 数据发展，进而促进知识基础设施建设。若干研究发现和已识别的挑战值得在未来进一步研究：

研究问题1 探讨了促进知识基础设施建设所需的各类制品，并识别出许多合适的候选项。从大型、基于图谱的数字图书馆到小型专用实例，再到成熟的标准、新兴的可互操作交换格式和工作流自动化工具，这些制品构成了一份全面的目录。然而，其中大多数仍面临社区采纳方面的挑战。

研究问题2 旨在评估模块化工具支持型工作流框架的能力。SWARM-SLR 表明，采用人在回路验证的笔记本驱动框架能够满足知识提取、结构化和表示方面的许多要求。ExtracTable 和 SWARM-SLR AIssistant 或可应对部分采纳和可用性方面的挑战，但抽象和人工智能权衡意味着当前方案仍有很大的改进空间，尚未充分满足所有要求。

研究问题3 旨在探究领域特定的知识基础设施需求。科学传播领域拥有易于获取的视听数据，这些数据需要进行结构化处理；相关法律考量看似简单，实则并不轻松，同时还需要整合各方贡献者。航空航天工程领域有多种结构化方法和模式，但受到自身施加的限制。核心挑战在于设计既能遵守法规、又能促进互操作性的解决方案。

研究问题4 探讨数字图书馆如何满足领域需求。Aerospace.Wikibase 是迈向共同基础的第一步，适用于其高度受限的领域，至少提供了通用的分类体系和持久标识符。然而，SciCom Wiki 能够提供同样的功能，并进一步发展为一个开放访问的数字图书馆，

因为在科学传播领域，数据共享并非问题。显然，基于联邦式关联开放数据实例构建的解决方案，尤其是 Wikibase 生态系统中的解决方案，能够满足许多需求。现成可用的资源十分丰富，也有模块化的工作流框架可供使用和定制。这些发现尚属初步，反映了正在进行的工作；未来的迭代和评估或许会揭示当前覆盖范围之外的更多内容。

致谢

AI 工具使用声明：在准备本作品期间，作者使用了 **GPT-5 mini (GitHub Copi-lot)**、**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**，依据 CEURGenAI 使用分类体系[4]。使用这些工具/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] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Ax- ton, A. Baak 等，《科学数据管理与数据保管的 FAIR指导原则》，*Scientific Data*，第 3 卷，2016 年 3 月。出版方：自然出版集团。[2] P. N. Edwards，《一台宏大的机器：计算机模型、气候数据与全球变暖的政治》。麻省理工学院出版社，2010 年。[3] H. Karasti、K. S. Baker 和 F. Millerand，《基础设施时间：协作开发中的长期议题》，计算机支持的协同工作（*CSCW*），第 19 卷，第 377–415 页，2010 年 8 月。[4] C. L. Borgman，《什么是数字图书馆？相互竞争的愿景》，信息处理与管理，第 35 卷，第 3 期，第 227–243 页，1999 年。[5] D. Vrande˘ci´c 和 M. Kr¨otzsch，《维基数据：一个免费的协作知识库》，*Commun. ACM*，第 57 卷，第 78–85 页，2014 年 9 月。[6] S. Auer、A. Oelen、M. Haris、M. Stocker、J. D'Souza、K. E. Farfar、L. Vogt、M. Prinz、V. Wiens 和 M. Y. Jaradeh，《借助知识图谱提升科学文献的获取》，*Bibliothek Forschung und Praxis*，第 44 卷，第 516–529 页，2020 年 12 月。出版方：De Gruyter。[7] T. W. Burns、D. J. O'Connor 和 S. M. Stocklmayer，《科学传播：当代定义》，*Public Understanding of Science*，第 12 卷，第 183–202 页，2003 年 4 月。出版方：SAGE Publications Ltd.。[8] K. Kappel 和 S. J. Holmen，《为何开展科学传播，它是否有效？科学传播目标的分类体系及实证证据综述》，传播学前沿，第 4 卷，2019 年 10 月。出版方：Frontiers。[9] L. E. MacKenzie，《科学播客：2004 至 2018 年的全球制作与产出分析》，皇家学会开放科学，2019 年 1 月。出版方：英国皇家学会出版部。[10] J. Kikkawa、M. Takaku 和 F. Yoshikane，《提升对 YouTube 上学术参考文献的识别》，载于数字图书馆理论与实践（A. Antonacopoulos、A. Hinze、B. Piwowarski、M. Coustaty、G. M. Di Nunzio、F. Gelati 和 N. Vanderschantz 编），（Cham），第 326–341 页，Springer Nature Switzerland，2024 年。[11] J. Kikkawa、M. Takaku 和 F. Yoshikane，《YA领域数据集：YouTube 视频上的学术书目参考文献数据集》，2024 年 7 月。[12] P. Mar´ın Arraiza 和 S. Strobel，《TIB|AV Portal：未来的关联媒体生态系统》，载于第 *24* 届万维网国际会议论文集，WWW '15 Companion，（美国纽约州纽约市），第 733–734 页，美国计算机协会，2015 年 5 月。[13] E. Navarrete、A. Nehring、S. Schanze、R. Ewerth 和 A. Hoppe，《深入探讨近期基于视频的学习研究：视频特征、工具、技术与学习效果的全面综述》，2023 年 8 月。arXiv:2301.13617 [cs]。

[14] B. F¨ahnrich、E. Weitkamp 和 J. F. Kupper，《探索在线科学传播中的"质量"：专家对如何评估和提升数字媒体环境中科学传播质量的看法》，Public Understanding of Science，第 32 卷，第 605–621 页，2023 年 7 月。出版方：SAGE Publications Ltd.。[15] S. Hagenhoff、L. Seidenfaden、B. Ortelbach 和 M. Schumann，《科学传播的新形式：一项案例研究》。Universit¨atsverlag G¨ottingen，2007 年。[16] J. Stjepandi´c、W. J. C. Verhagen、H. Liese 和 P. Bermell-Garcia，《基于知识的工程》，载于《21世纪的并行工程：基础、发展与挑战》，第 255–286 页，瑞士 Cham：Springer International Publishing，2015 年。[17] E. J. Reddy、C. Sridhar 和 V. Rangadu，《基于知识的工程：概念、方法与未来趋势》，American Journal of Intelligent Systems，第 5 卷，第 1–17 页，2015 年。[18] S. Quintana-Amate、P. Bermell-Garcia 和 A. Tiwari，《将专业知识转化为基于知识的工程工具：工程设计背景下知识获取的调查》，Knowledge-Based Systems，第 84 卷，第 89–97 页，2015 年。[19] G. La Rocca 和 M. J. van Tooren，《支持飞机设计中有限元模型自动生成的基于知识的工程方法》，载于第 45 届 AIAA航空航天科学会议与展览，航空航天科学会议，（内华达州里诺），2007 年 1 月 8—11 日。[20] I. Staack 和 P. Krus，《在飞机概念设计中集成机载电力系统仿真》，载于第四届 CEAS欧洲航空航天会议论文集，第 709 页，欧洲航空航天学会理事会（CEAS），Link¨oping University Electronic Press，2013 年，2013 年。[21] L. Kn¨o¨os Franz´en，《产品早期开发中的系统之系统视角：一种基于本体的方法》。博士论文，Link¨oping Uni- versityLink¨oping University，流体与机电系统，科学与工程学院，2023 年。[22] A.-S. Dadzie、R. Bhagdev、A. Chakravarthy、S. Chapman、J. Iria、V. Lan- franchi、J. magalh˜aes、D. Petrelli 和 F. Ciravegna，《在航空航天工程中应用语义网技术促进知识共享》，Journal of Intelligent Manufacturing，第 20 卷，第 5 期，第 611–623 页，2009 年。[23] I. van Gent、G. La Rocca 和 M. F. M. Hoogreef，《CMDOWS：一种存储和交换多学科设计优化系统的新提议标准》，CEAS航空学杂志，第 9 卷，第 607–627 页，2018 年 12 月。[24] I. Sanya、E. Shehab 和 D. Lowe，《航空航天设计工程中语义知识管理面临的挑战》，载于 DS 68-6：第 18 届工程设计国际会议（ICED 11）论文集——通过工程设计影响社会，第 6 卷：设计信息与知识，丹麦 Lyngby/哥本哈根，2011 年 8 月 15—19 日，第 241–248 页，2011 年。[25] T. Procko 和 O. Ochoa，《利用关联数据进行知识管理：面向航空航天工业的提案》，SSRN 4703237，2022 年。[26] D. Harvey 和 R. Holdsworth，《航空航天工业中的知识管理》，载于 IPCC 2005：国际专业传播会议论文集，2005 年，第 237–243 页，2005 年 7 月。ISSN：2158-1002。[27] T. Wittenborg、O. Karras 和 S. Auer，《SWARM-SLR——面向机器可操作系统性文献综述的精简工作流自动化》，载于《数字图书馆理论与实践》（A. An- tonacopoulos、A. Hinze、B. Piwowarski、M. Coustaty、G. M. Di Nunzio、F. Gelati 和 N. Vanderschantz 编），（Cham），第 20–40 页，Springer Nature Switzerland，2024 年。[28] L. John、A. M. Ghanmi、T. Wittenborg、S. Auer 和 O. Karras，《ExtracTable：将科学语料库转换为结构化知识的人在回路中的转换》，载于《数字图书馆理论与实践》（W.-T. Balke、K. Golub、Y. Manolopoulos、K. Stefanidis 和 Z. Zhang 编），（Cham），第 470–487 页，Springer Nature Switzerland，2026 年。[29] 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》，2025 年 5 月。arXiv:2505.10142 [cs]。[30] T. Wittenborg、N. Stehr、O. Karras 和 S. Auer，《SciCom Wiki：支持视频和播客科学传播知识基础设施的数字图书馆》，2025 年 11 月。arXiv: . cs . [31] T. Wittenborg、C. S. Tremel、O. Karras 和 S. Auer，《在线话语的计算事实核查：对气候变化相关新闻文章的科学准确性进行评分》，载于 2025 年 IEEE知识图谱国际会议（ICKG），第 371–378 页，2025 年。

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