通过引文元数据和时空元数据丰富学术元数据公地，以支持负责任的研究评估和研究发现

Daniel Nüst[1*], Gazi Yücel[2], Anette Cordts[2], Christian Hauschke[2]

[1]地理信息学研究所，明斯特大学，德国

[2]TIB——莱布尼茨科学与技术信息中心，汉诺威，德国

* 通讯作者：Daniel Nüst，daniel.nuest@uni-muenster.de

关键词：开放科研信息、引文元数据、时空元数据、学术出版、研究评估、发现

摘要

本文聚焦于开放科研信息作为透明且负责任的研究评估以及研究成果发现基础的重要性。我们介绍了一项工作，通过支持开放科研信息公地，尤其使独立的小型开放获取期刊能够向多个开放数据枢纽（Open Citations、维基数据、开放研究知识图谱）提供元数据。在此背景下，我们提出 The OPTIMETA Way：一种以有效且经过质量保证的方式整合元数据收集、丰富和分发的方法，使其即使在小型学者主导的出版渠道中也能得到采用。我们针对这一方法设计了一套实施策略，以最广泛使用的期刊出版软件开放期刊系统（OJS）中的两个插件为载体。这些插件会收集、丰富并自动交付论文的引文元数据和时空元数据。我们利用关联开放书目数据促进研究评估和发现的贡献体现在三个方面。第一，我们倡导通过开放 API，在出版时收集经过丰富且用户验证的元数据，从而扩大开放科研信息数据池。第二，我们将目前因语言因素或缺乏大型出版机构支持而未被纳入标准科学计量实践的数据平台和期刊整合进来。第三，我们支持基于位置和时间元数据的新用例，使其超越通常使用的发现功能，尤其是利用空间覆盖范围评估研究活动，以及探索研究成果之间新的跨学科联系。



本文档采用知识共享署名 4.0 国际许可协议：知识共享署名

1 引言

开放科研信息（ORI）为透明且负责任的研究评估以及有效发现研究成果提供了基础。因此，ORI 的质量、开放性、广度和范围都应尽可能高。然而，尤其对于小型独立开放获取期刊而言，收集出版元数据并将其存入开放科研信息数据枢纽十分困难。对于更复杂的元数据而言，情况尤其如此；这类元数据超越了单条记录的标题或出版日期等简单属性，通过建立科学出版物之间的联系来发挥作用。在这项工作中，我们提出了一种概念，旨在于出版流程中最有效且高效的时刻，促成经过验证的开放出版元数据的快速发布。我们将这一概念称为 The OPTIMETA Way，并通过 Public Knowledge Project（PKP；https://pkp.sfu.ca/）开发的出版软件开放期刊系统（OJS）中的两个插件予以实现。这两个插件涵盖两类具有极高潜力的元数据，可支持更负责任的研究评估和更强大的研究发现，但目前尚未得到广泛提供：经过验证且完整的开放引文元数据，以及时空元数据。

本研究的主要贡献在于提出了一种有效收集和发布出版元数据的理念，该理念能够：（a）兼顾数据质量与人工用户交互的需求；（b）瞄准出版流程中利益相关者具备动机和专业知识的特定环节；（c）使独立期刊能够创建超越大型商业出版商通行标准的创新元数据。本文通过原型实现——OPTIMETA 服务——展示了这一理念。OPTIMETA 服务聚焦于出版元数据的两个领域，充分发挥研究成果之间空间和时间关系的整合作用，并促进开放引文信息的获取，以应对数量庞大的科学文献。

本文结构如下。首先，我们简要介绍基础背景工作，尤其关注受众可能存在的多样性。接着，我们提出创新出版元数据的理念和实施策略。最后，在讨论本研究的益处、局限性和未来方向之前，我们将这些理念和成果与学术元数据公地、研究评估和研究发现概念联系起来。

2 相关工作

2.1 开放书目元数据：负责任的研究评估的基础

研究需要大量资源，组织和资助研究的参与者有兴趣筛选和评估这些投入所产生的研究成果。因此，开展研究评估有各种原因，其中包括学术界存在“缺乏信任”的问题，以及希望促进资源分配得到改善。这类评估通常被描述为基于复杂框架和评估规则、对研究流程进行元数据驱动的量化。评估通常依据产出与接收类指标，而只有在具备适用于此目的的元数据时，才能生成这些指标。

近年来，传统的研究评估方法受到了来自各方的根本性挑战。书目元数据的作用是其中一个受到关注的重要问题，例如，两项广受讨论的倡议就提出了这一问题，因为必须通过开放许可提供此类材料，才能实现公平且负责任的科研评价。两项倡议都谈到了负责任评估所需的数据基础。2012年，学术出版领域的几位重要参与者在美国加利福尼亚州旧金山举行的美国细胞生物学学会年会上会面，讨论与科研评价相关的新兴问题（美国细胞生物学学会，2012年）。此次会议的成果是《旧金山科研评价宣言》（DORA，参见 Cagan，2013年）。截至2022年7月，已有来自158个国家的近22,000名个人和组织签署了该宣言。DORA明确要求出版商和指标供应商做到“开放透明”。DORA建议的一项关键行动是，出版商应“取消对研究论文参考文献列表的一切再利用限制，并依据知识共享公有领域贡献许可提供这些列表”（建议9）；指标供应商则应依据开放许可公开其指标所依据的数据和方法（建议11和12）。第二项倡议于2015年发起，即《莱顿宣言》（Hicks 等，2015年），它进一步提出了更为具体的指导，说明评估所用元数据应如何共享，尤其是为何应当共享。由此形成的两项原则——第4项（“保持数据收集和分析流程开放、透明且简单”）和第5项（“允许被评估者核实数据和分析结果”）——与本文的主题尤为相关。

为了遵循负责任的研究评估原则和建议，数据源必须满足各种标准。研究信息——即与研究活动相关的参与者、事件、流程和产出的元数据——必须可发现且可访问，才能对出版物进行评估；记录还必须包含元数据，且其形式应便于合法且在技术上轻松地重复使用。Wilkinson 等人（2016年）将这些标准（可发现性、可访问性、互操作性和可重用性）称为“科学数据管理与管理服务的FAIR指导原则”。顺着这一逻辑推至结论，任何研究评估都应完全基于从公开可用的数据源中获取、采用开放许可，并使用开放标准发布的元数据。

2.2 开放书目元数据和时空元数据作为开放科研信息的一部分

开放科研信息（ORI）是一个新兴术语，用于描述符合上述标准、适用于负责任的研究评估的数据源元数据。近期，人们采取了各种行动并发起了多项倡议，旨在更准确地界定和推广ORI。例如，2018年，第14届国际当前研究信息系统会议（euroCRIS 2018）以“科研信息的FAIR性”为主题。2020年，一个德乌项目与国际专家探讨“开放基础设施中的FAIR研究信息”这一主题（Kaliuzhna 和 Altemeier 2021），促成了适用于研究信息的 FAIR原则高级标准的制定（Hauschke 等，2021a）。Bijsterbosch 等（2022）介绍了“开放研究信息的七项指导原则”，并对“可信且透明的溯源”“

丰富学术元数据公地

元数据”“算法开放性”“持久访问与可用性”“开放标准与互操作性”“与第三方开放协作”以及“通过治理实现学术自主权”进行了更为详细的分析。

开放书目元数据是 ORI 的重要组成部分，尤其与以产出为导向的研究评估相关。从更广泛的角度来看，这涉及描述已发表作品的任何元数据，例如期刊论文、会议论文集、专著或编著。作为出版界和图书馆界的基础与动力，书目元数据由作者、编辑、图书馆员以及许多参与传播学术成果的其他人员生成。传统的书目元数据类型包括参考文献列表、摘要、作者所属机构、作者标识符和许可协议等。除研究评估外，书目元数据还被用于多种其他用途，例如创建学术知识图谱（如 Jaradeh 等，2019年；Manghi 等，2021年；Priem 等，2022年；Turki 等，2021年），以及用于图书馆目录和书目发现服务（Gonzales，2014）。

近年来，其他书目元数据类型——时空元数据——也日益受到关注（Niers 和 Nüst，2020年）。时空元数据能够提供出版物所涵盖的位置和时间段的精确信息。这些元数据使不同研究成果之间能够建立联系；然而，目前 ORI 中时空元数据的可用性和使用情况仍较为有限。对于出版物（Niers 和 Nüst，2020年）、数据（Garzón 和 Nüst，2021a）以及当前研究信息系统（CRIS）平台而言，时间与空间的整合潜力都尚未得到充分利用。

多个倡议和项目正致力于丰富学术元数据公地。这里无法全面评述所有相关工作，但一些重要案例包括 Rasberry 等人（2019）的研究，他们展示了如何通过维基数据的前端 Scholia，将维基数据用作学术元数据的来源。他们还讨论了位置数据，不过主要涉及作者及其机构地址的坐标。Nielsen 等人（2018）进一步探讨了基于位置信息进行发现的应用场景，并介绍了如何利用维基数据的点和多边形要素，在 Scholia 中查询文章主题中提及的地点。Lauscher 等人（2018）认为，图书馆应在学术元数据的制作和整理中发挥重要作用，并以社会科学领域印刷书籍的引文元数据为例，证明了这一策略的可行性和有效性。乌克兰开放引文索引是全国范围内收集学术出版商引文元数据的一个案例（Nazarovets，2019）。在国际层面，两项倡议已取得广泛关注：开放引文倡议（https://i4oc.org/）和开放摘要倡议（https://i4oa.org/）。尽管这些项目和倡议各有价值，但仍有许多工作有待完成，尤其是对于小型、独立和学者主导的期刊而言。对这些期刊来说，引文元数据和时空元数据潜力巨大，因为当这些元数据以结构化、机器可读的格式提供，并可通过开放书目数据库访问时，就能建立不同出版物之间的联系，从而提升大量学者主导的开放获取期刊所作贡献的可见度。

2.3 引文元数据

引文元数据是指描述一份文档如何引用另一份文档的元数据。早在1952年，Eugene Garfield就在美国化学学会马里兰分会的一次演讲中提出了记录文档间交叉引用的想法。他表示：“如果作者在每条书目引文中都提供 CA 摘要编号，我可以保证，未来的 CA 摘要将通过提供与相关摘要的交叉引用而包含更多信息”（Garfield，1952年）。后来，他着手创建了科学引文索引，该索引逐渐发展为引文数据库科学网，并启发了随后几十年中出现的许多类似产品。

引文元数据的描述是两个对象之间的一种简单关系，随着时间的推移，人们已经针对不同应用开发了多种标准和定义。为了说明各种工作成果的多样性，我们讨论三个涉及不同类型引文元数据的例子。Starr 和 Gastl（2011）介绍了 DataCite 描述出版物与研究数据集之间关系的初始方式。其基本假设是，DataCite 元数据模式中的每个实体，即每项研究成果，均使用最少数量的必填字段进行描述：创建者、标题、出版者、标识符和出版年。随后，可以通过各种方式构建单个实体之间的关系。IsCitedBy（及其逆属性 Cites）跟踪相互引用的作品之间的关系。这种方法与内容无关。Peroni 和 Shotton（2012）开发了引文类型本体（CiTO），其中引文仍然连接两个作品，但引文本身被视为一个可以独立通过各种属性进行描述的实体。这使得人们能够更加详细地描述各种类型的关系，并表示诸如文内引文频率等特征。

随着时间推移，描述特定类型作品的引文关系标准也逐渐形成。一个近期的例子是研究软件的引用，Smith 等人（2016年）为此制定了相关原则，旨在捕捉此类作品的特点。例如，必须能够处理软件开发中常见的版本管理，并明确说明引用的是软件的某个特定版本、任意版本，还是最新版本。

2.4 时空元数据

地理空间元数据是关于地理数据和信息的元数据（维基百科 2022年）。从用于编码地理空间元数据的公共标准和行业标准（例如复杂的 ISO 191** 系列标准），到在线收集和提供地理空间元数据的目录（联邦地理数据委员会），相关标准、格式和工具种类繁多。虽然这些类型的数据通常具有相关性，但研究成果和发布数据集正变得越来越普遍，也越来越受到认可。我们的研究聚焦于更为经典的研究成果——研究论文——的地理空间属性。因此，在提及描述文本和图形研究成果的空间或时间性质的元数据时，我们使用 时空元数据一词。例如，这类元数据可以指科学论文所研究的空间范围或所谓的感兴趣区域，也可以指数据被分析所对应的时间段。JournalMap（Karl 等，2013年）此前已采用过这种方法，尽管

存在相当大的局限性（Hauschke 等，2021b）。此外，最常用的研究数据仓储（例如 Zenodo、OSF 和 Figshare）并不明确支持空间元数据。只有少数研究数据仓储专门用于处理地理配准数据，例如 Pangaea（https://www.pangaea.de/about/）以及带有空间扩展的 CKAN（https://ckan.org/features/geospatial/）。Dataverse 仅支持采用过时格式的矢量数据 Shapefile（https://guides.dataverse.org/en/latest/developers/geospatial.html?highlight=地理空间）。这种支持的缺乏可能是因为，地理空间数据的处理并不是数据库软件的常见功能（除非使用额外的软件扩展），也不是专业领域专长的重点。

这类元数据与广泛的科学领域息息相关。在自然科学与生命科学中，观测数据、模型数据、栖息地或发现地点是最显而易见的关联点（例如涉及地球科学、地质学、海洋学、气象学、生态学、动物学和植物学）。对于社会科学与应用科学而言，人类及其实体活动区域与这些领域的联系十分重要，例如医学与健康科学、农业科学、经济学、工程学、社会学、政治学等。正如预期的那样，形式科学（如数学、逻辑学、计算机科学）中的理论研究，或物理科学（物理学、化学）中小尺度且偏理论的研究，对地理空间元数据的兴趣较低；不过，跨学科研究或包含某些应用元素的研究通常与现实世界有关，也就是说，涉及时空中的某个位置。

让我们把注意力重新转向用于处理研究出版物的平台：无论是用于 CRIS 的大型商业解决方案，例如 Pure（爱思唯尔）或 Converis（科睿唯安），还是广泛使用的开放项目，例如 DSpace（Smith 等，2003），都不支持任何记录类型的时空元数据。VIVO（Conlon 等，2019）有一个用于描述给定研究活动或产出的空间组成部分的属性（vivo:geographicFocus），但并未得到广泛使用。开发者文档和代码仓库表明，人们对这一主题抱有兴趣；例如，DSpace 列出了多个曾有人提出或制作原型的空间搜索实例（参见 https://github.com/DSpace/DSpace/pull/511）。在发现平台（例如 ScienceDirect、Google Scholar）和出版商网站中，也存在同样缺乏支持的情况。在出版领域，对时空信息进行明确建模的唯一实例，是都柏林核心规范*DCMI* 元数据术语中的**覆盖范围**字段，其中该字段最为重要（https://www.dublincore.org/specifications/dublin-core/dcmi-terms/terms/coverage/）。然而，这一字段对于机器可读信息交换而言并不是特别有用，因为其中可以存放任何类型的空间或时间元数据，例如散文式描述、坐标对，或复杂几何图形及时间段的文本编码。DCMI 元数据术语还使用了一些替代术语，例如管辖区或位置，但这些术语似乎并未在此处提到的平台中得到实现。因此，所有这些平台都包含未明确说明或无法直接使用的空间信息，例如研究人员的地址、论文摘要中的位置名称、全文中的发掘地点坐标，或补充材料中作为图表出现的遥感影像。时间元数据更为常见，这主要是因为其性质更简单，并且任何数据库管理系统都能轻松支持。然而，时间字段通常侧重于出版元数据（资源是在何时创建或出版的？），而非内容本身。类似地，这类时间信息也隐藏在标题、摘要或全文中。

这完全忽视了空间和时间信息作为整合器的潜力，例如 Kuhn（2012）在跨学科研究方面所论述的那样。过去已有研究指出了这种潜力，既涉及在出版物之间建立新的联系（Niers 和 Nüst，2020），也涉及数据（Garzón 和 Nüst，2021a；Garzón 和 Nüst，2021b）。总体而言，时空元数据目前在 ORI 中并未发挥重要作用。

3 The OPTIMETA Way

3.1 方法

引言介绍了丰富学术元数据公地所涉及的挑战与益处，以及这如何与负责任的评估和有效发现研究成果相关联。为应对捕获和传播高质量且实用的研究成果元数据时遇到的部分挑战，我们制定了一种通过开放引文和时空元数据来强化开放获取出版系统的方法——*The OPTIMETA Way*。这一理念承认，在研究周期的出版阶段，元数据创建与使用之间存在冲突。

首先，创建高质量元数据给作者带来的益处并非显而易见，尽管作者是大多数相关信息最了解的来源。然而，如果清楚地说明提供元数据的理由，这种情况或许会有所改变，例如，这有助于更负责任地评估他们的工作，并加强与其他学科的联系，以利于评估和发现。尽管如此，元数据的采集方式应直观且富有吸引力，同时牢记 James Frew 定律：*“Frew* 第一定律：科学家不会撰写元数据。*Frew* 第二定律：任何科学家都可以被迫撰写糟糕的元数据。”（Hey 2015）。其次，目前大型学术出版商是元数据的主要所有者。他们强烈关注发展自身业务，并利用此类数据获利（Pooley 2022; Brembs 2021; Franceschi-Bicchierai 2022），却不太愿意尽其所能为知识创造贡献，或回应技术创新的需求。相反，创新必须由那些希望拥有开放、免费的学术元数据公地的人推动，例如大学出版社或独立期刊，尽管他们的资源有限。

鉴于这些冲突与挑战，The OPTIMETA Way 以产出为导向，致力于让学术出版中至关重要、尽管目前相对缺乏话语权的利益相关者能够自行采集和传播学术出版元数据。这种方法之所以称得上**“OPTImal”**，是因为它从小幅改进入手，而这些改进有望带来最大的收益：高质量引文元数据对研究评估和质量保障型元科学的潜在影响巨大。此外，时空元数据的新颖性及其整合潜力，即使在初步应用阶段也颇具吸引力，例如，在同一张地图上展示代表论文的简单几何图形，以供目视检查和发现。

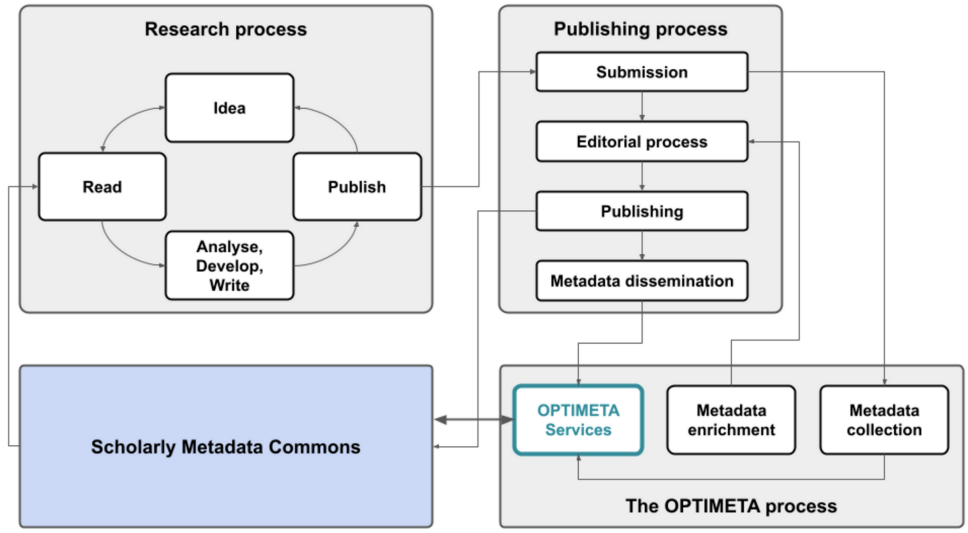


图 1. 出版流程、通用研究流程与 The OPTIMETA Way 的各个阶段及其相互联系。

这种方法**以产出为导向**，通过自动化和直观的用户界面为元数据创建者提供帮助，从而避免给他们增加额外、繁重、耗时或不受欢迎的任务。该软件利用开放数据源丰富用户输入内容，随后提供这些内容供人检查。与所需投入相比，这种方法还着眼于以最快速度取得成效，因此无须面面俱到。元数据将在出版流程中的一个特定时点创建：作者最愿意满足所有行政要求的时候，也就是提交论文供评审、并最终发表之时。这也是积极参与的专业人员（审稿人、编辑、出版工作人员）批判性审查元数据的最后时机；同时，发表产出（包括元数据）本身已是流程的核心环节。此外，在这一流程中，产出始终由人工核查；由于数据并非由算法生成，因此可以将质量更高的元数据回传至用于丰富数据的来源。

最后，通过 The OPTIMETA Way 实现**促进**。这种方法使独立期刊和小型出版商（他们通常使用开源软件平台）能够积极参与结构化元数据的收集与分发，而无需具备专业知识。实现这一目标的关键步骤是将元数据存入开放数据枢纽，从而让肩负传播知识这一共同使命的期刊为构建更宏大的开放学术出版元数据共享库贡献力量。因此，The OPTIMETA Way 的影响将大于各部分影响之和。

图1是 The OPTIMETA Way 的示意图。一侧展示了简化的研究周期，包括研究活动及其转化为书面成果的过程。另一侧展示了期刊的出版流程，从投稿、编辑到出版，以及相关或后续的元数据传播。The OPTIMETA Way 意味着我们支持在期刊出版流程中元数据本已受到关注的阶段生成、

丰富和传播元数据。通过这种方式，该方法避免了研究人员进行任何追溯性编辑和出版后的任务。

3.2 实施示例

3.2.1 概述

开放期刊系统（OJS）是全球应用最广泛的期刊出版软件。它旨在让科学出版更加轻松、高效（Willinsky，2005年）。它涵盖从投稿到评审、从校对到制作的完整出版工作流程。该软件为开源软件，并由全球社区持续改进。目前推出的版本为 3.3。OJS 提供上述核心功能，包括元数据创建、编辑和导出。通过插件还可以为 OJS 添加更多功能，这些插件既可手动安装，也可通过所谓的插件库（https://docs.pkp.sfu.ca/plugin-inventory/en/）安装。这些插件可以扩展或更改 OJS 的任何投稿和出版工作流程步骤，并为编辑后台或系统前端添加新功能。

我们通过两个插件实现了 The OPTIMETA Way：OPTIMETA 引文插件和 OPTIMETA 地理插件（简称“引文插件”或“地理插件”）。这两个插件共同实现了**OPTIMETA 服务**（见图1）中所述的功能。两个插件都会在投稿工作流中收集元数据，并利用开放学术数据源中的数据对其进行丰富。随后，它们会在最终公开这些信息之前，向提交数据的人员展示增强后的信息；最终，这些信息会同时发布在 OJS 网站和外部数据接收端。与外部数据接收端的连接可以近乎实时地完成，即在出版工作流程发生相关事件时主动提交信息。或者，也可以使用其他平台作为采集机制，定期从 OJS 获取已发布的元数据。图2展示了这些插件当前支持的数据源和数据接收端，以及未来可能支持的对象。左侧的数据源以红色标示。两个插件都依赖用户贡献的元数据，并利用外部参考文献和其他数据对其进行丰富。右侧以绿色标示的是目标元数据平台和聚合器（即数据接收端）。与本项目合作的期刊和大学出版社列于底部（完整名称列表及链接请见https://projects.tib.eu/optimeta/en/）。

这些插件以公开测试版的形式提供，可从我们的公开 GitHub 代码库 https://github.com/TIBHannover/optimetaCitations/和 https://github.com/TIBHannover/optimetaGeo自由下载和安装。我们将根据OPTIMETA 项目合作伙伴和 OJS 社区（用户和开发者）的反馈改进这些插件，反馈尤其聚焦于用户体验。

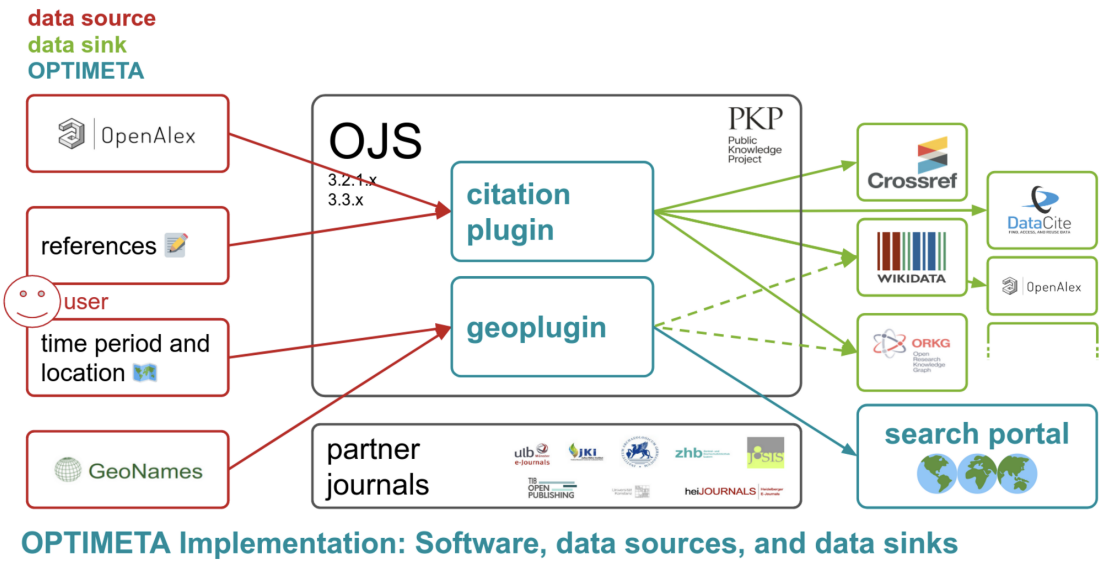


图 2. The OPTIMETA Way 的实施示例：适用于 OJS 的 OPTIMETA 引文插件和地理插件，以及相连的外部数据源、数据接收端和合作伙伴。

3.2.2 引文插件

引文插件旨在于出版流程中收集机器可读的引文元数据，以便将这些元数据发布到开放书目数据源中。该流程已整合到用于提交出版物的现有开放期刊系统工作流程中。当作者开始向期刊投稿时，OJS 会为参考文献提供一个特殊的元数据字段。我们假设该字段随后由作者或期刊编辑填写。引文以原始格式输入 OJS，即输入到一个可自由修改的文本字段中（图 3）。

Hauschke C, Nüst D, Cordts A, Lilienthal S (2021) OPTIMETA – Strengthening the Open Access publishing system through open citations and spatiotemporal metadata. Research Ideas and Outcomes 7: e66264. https://doi.org/10.3897/rio.7.e66264

图 3. 原始引文示例。

引文插件会解析原始参考文献，并在存在数字对象标识符（DOI）时将其提取出来。提取 DOI 后，查找算法会基于外部开放书目数据源，以语义结构化格式丰富元数据。目前，引文插件会查询 Crossref 和 OpenAlex 的开放 API，这两个平台都具有良好的覆盖范围和可观的元数据质量。我们选择从外部数据源收集元数据，而不是使用解析工具解析引文，因为我们发现，根据 Crossref 和 OpenAlex 中的 DOI 查询得到的结果总体上更加完整、准确。开发定制的解析服务或集成现有工具都会给插件带来不必要的复杂性。所有这些步骤都可以通过单击按钮手动触发。因此，所需的额外时间

用于丰富引文元数据的时间仅需几分钟甚至更短；与整个研究和出版流程所耗费的时间相比，这几乎可以忽略不计，如图1所示。

依赖 DOI 来查找完整参考文献信息是该插件目前的一项局限。由于非成员访问 API 受到限制，并且可靠的 SimpleTextQuery 表单（https://apps.crossref.org/SimpleTextQuery）仅供手动使用，因此未实现通过完整参考文献查询 Crossref 的替代方案。我们不知道 OpenAlex 是否提供可比的完整参考文献查询功能。在这种情况下，鉴于由此获得的元数据质量更高，且所需的手动交互更少，我们决定采用更高程度的自动化，其收益超过了遗漏没有 DOI 的参考文献这一缺点。对于新的投稿，插件将鼓励作者在投稿过程中补充缺失的 DOI，从而确保元数据具有合理较高的质量和完整性。由于 OpenAlex 从 Crossref 抓取数据，将这两项服务同时作为插件的数据源似乎有些多余。然而，这种冗余可以避免未来依赖某一项特定服务；此外，由于 OpenAlex 还会抓取其他数据源，同时使用两个数据源很可能提供额外且可能更加完整的信息。

完成这些步骤后，系统会将从外部开放 API 中提取的丰富结果呈现给作者进行评审。作者只需检查结果是否正确即可完成评审。引文的各个部分也可以手动编辑（图4）。随后，这些经过语义结构化的元数据（包括标题、作者及其对应的作者标识符（ORCID iD）等）会存储在 OJS 数据库中。

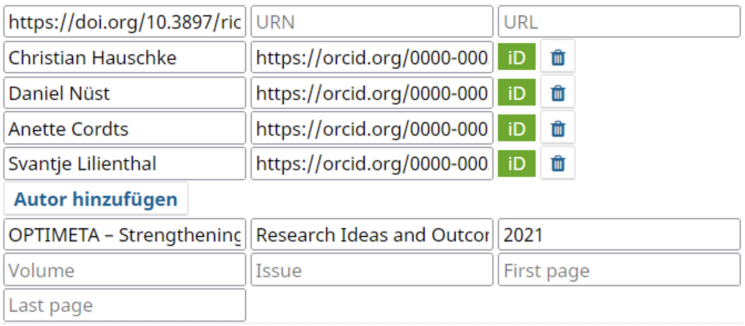


图 4. 语义结构化引文示例，可手动编辑。

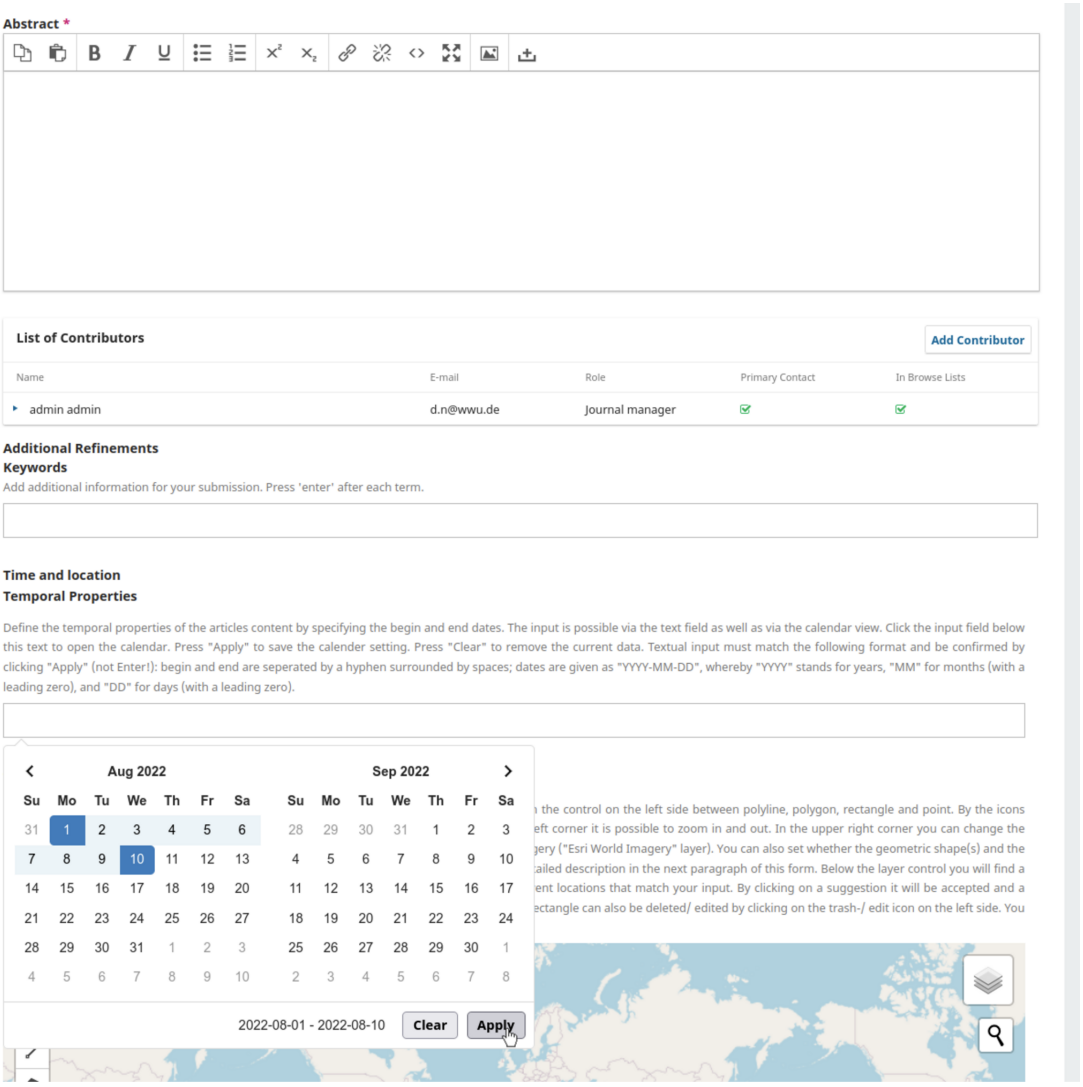
经过增强处理后，可以通过两种方式将引文提交至外部服务：手动点击提交按钮，或由 OJS 调度器管理的自动流程。目前，第一种工作流程已应用于 OpenCitations。整合后的元数据按照 Massari 和 Heibi（2022）所述的方式构建，并作为期刊期次提交至 OpenCitations 指定的 GitHub 仓库（https://github.com/GaziYucel/open_citations_croci_depot）。随后，OpenCitations 可以处理并采集该期次中的元数据。

插件的初始版本侧重于使用 DOI 作为受支持的出版物标识符。在未来的版本中，我们计划支持非 DOI 标识符。此外，支持通过 RIS 和 BibLaTeX 等常见书目标准导入书目元数据，将有助于改善用户体验。目前，插件开发的重点是在论文的出版流程中提供支持。不过，期刊运营者自然不只希望为新文章发布引文信息，也希望为过往出版目录中的文章发布引文信息。为满足这一需求，我们计划设计一个专门的概览页面，以便批量处理文章。最终，我们还希望接入更多数据接收端，从而尽可能广泛地传播引文元数据。例如，我们目前正致力于实现与维基数据的对接。

3.2.3 时空元数据插件

OPTIMETA 地理插件可在 OJS 实例中采集并展示单篇研究论文的时空元数据。名称中使用“geo”一词，是因为它比技术性更强的“spatio-temporal”更简短，也更容易被广泛理解。地理空间数据和地理数据通常包含时间方面的信息，也就是说，对“空间中的何处”这一问题的回答总是与“时间上的何时”相关联。由于多种原因，该插件中的“地理空间”元数据比时间元数据更为突出。首先，在地图上展示地理要素比用数字显示一个或多个时间段更直观，因此也更有趣。其次，地理空间元数据比时间元数据更具新颖性和效用，因为“2022年”或“白垩纪期间”之类的时间文本描述或分类更容易通过文本检索找到，而空间关系更难通过文本描述来捕捉，也很难通过文本检索找到。

地理插件尤其扩展了投稿工作流。由于输入地理空间元数据的方式简单直观，几分钟内即可快速提供时间和空间元数据。作者需要提供**时间元数据**，以简单时间段的形式注明开始日期和结束日期。日期可以手动输入，例如在表单字段中输入以连字符分隔的日期：“2021-01-01 - 2022-02-02”；也可以使用交互式日历弹出窗口输入，如图5所示。时间段是一个简单的字符串，可用于表示时间不确定性，并在需要或无法确定时允许一定程度的不精确。例如，历史学论文可能会将某种霸权的存续时间记为“753 - 1234”，而对新物种的观察则可能需要精确到具体日期，如“2022-08-08 - 2022-08-09”。



图。 5。投稿表单的屏幕截图，显示图像下方三分之一处用于输入时间段的表单字段和弹出窗口。

接下来，作者需要提供**空间元数据**，并使用交互式地图，如图6所示。作者可以添加多种类型的几何图形，例如点、折线、矩形和多边形。这些几何图形可用于恰当地表示与文章相关的空间特征，例如历史人物的居住地、动物踪迹、遥感观测区域或兽群领地。作者可以快速添加一个粗略矩形，提供较为不精确的数据；也可以放大地图并输入详细的单个数据点，具体取决于时间允许情况或投稿要求。作者可借助两个背景图层创建这些数据：OpenStreetMap提供的开放数据街道地图，以及Esri提供的免费航拍影像瓦片。这样，自然地物和人造建筑都可以作为方位参照。创建构成详细空间元数据的几何图形时，地理插件会持续查询GeoNames地名词典服务（https://www.geonames.org/），并使用几何图形的坐标自动推导出最小包围范围的边界矩形或边界框

行政区域。行政单位填写在地图下方的表单字段中。世界各国的行政区划单位数据差异很大，可以说，这类数据对全球北方国家最有用。

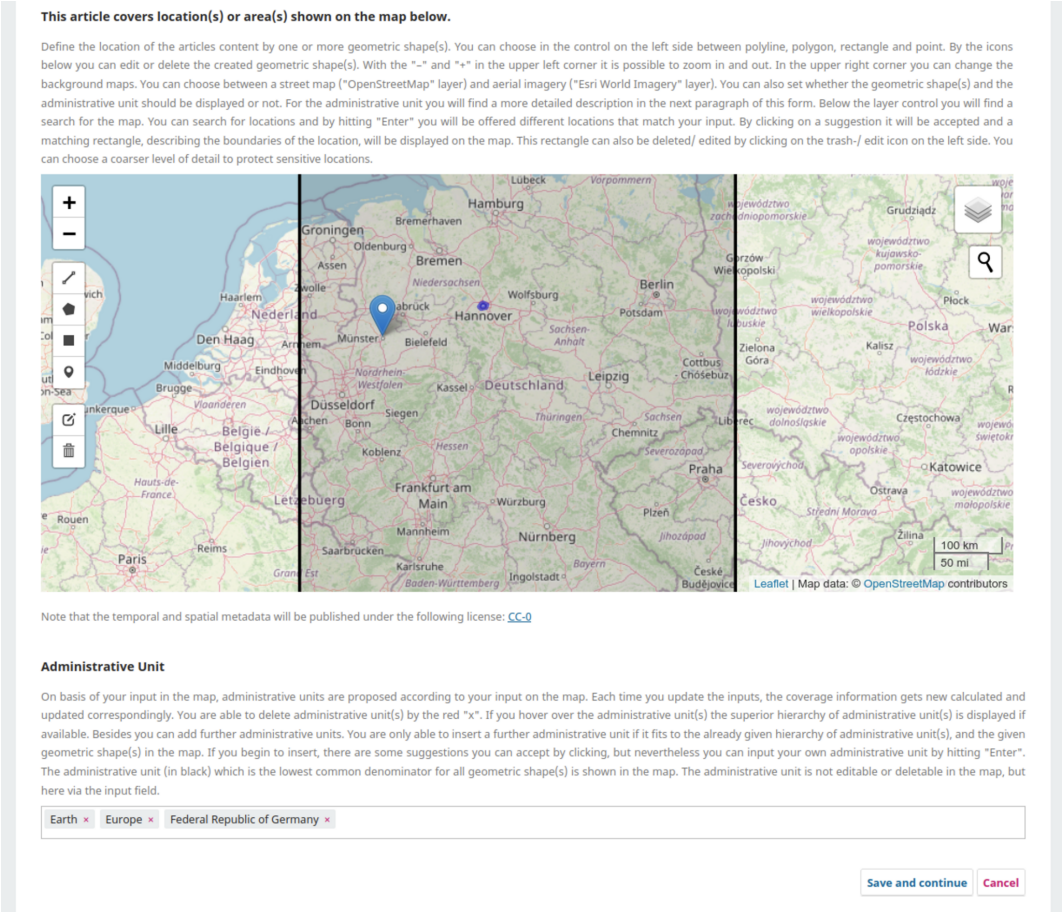


图 6.投稿表单截图，展示了用于收集空间元数据的交互式地图；本例中，明斯特市内有一个点几何，汉诺威市周围有一个蓝色多边形；这些几何图形均位于“地球、世界、德国”这一行政单位内，其边界矩形以黑色显示在地图上。

投稿过程中收集的信息随后可在编辑流程中接受评审。最终，数据以纯文本形式存储在 OJS 数据库中，格式为 GeoJSON（https://geojson.org/）。单个 要素集合 包含几何图形和一条简短的溯源声明，说明数据由谁创建或源自何处。尽管 GeoJSON 无法处理不同坐标参考系统，因此存在局限性，我们仍选择使用它，因为它应用广泛且简单易用。对于在全球范围内发现研究论文而言，“1984年世界大地测量系统”（WGS 84）坐标参考系统数米的精度已完全足够，尤其是考虑到大多数几何图形都是在交互式地图上手动创建的。

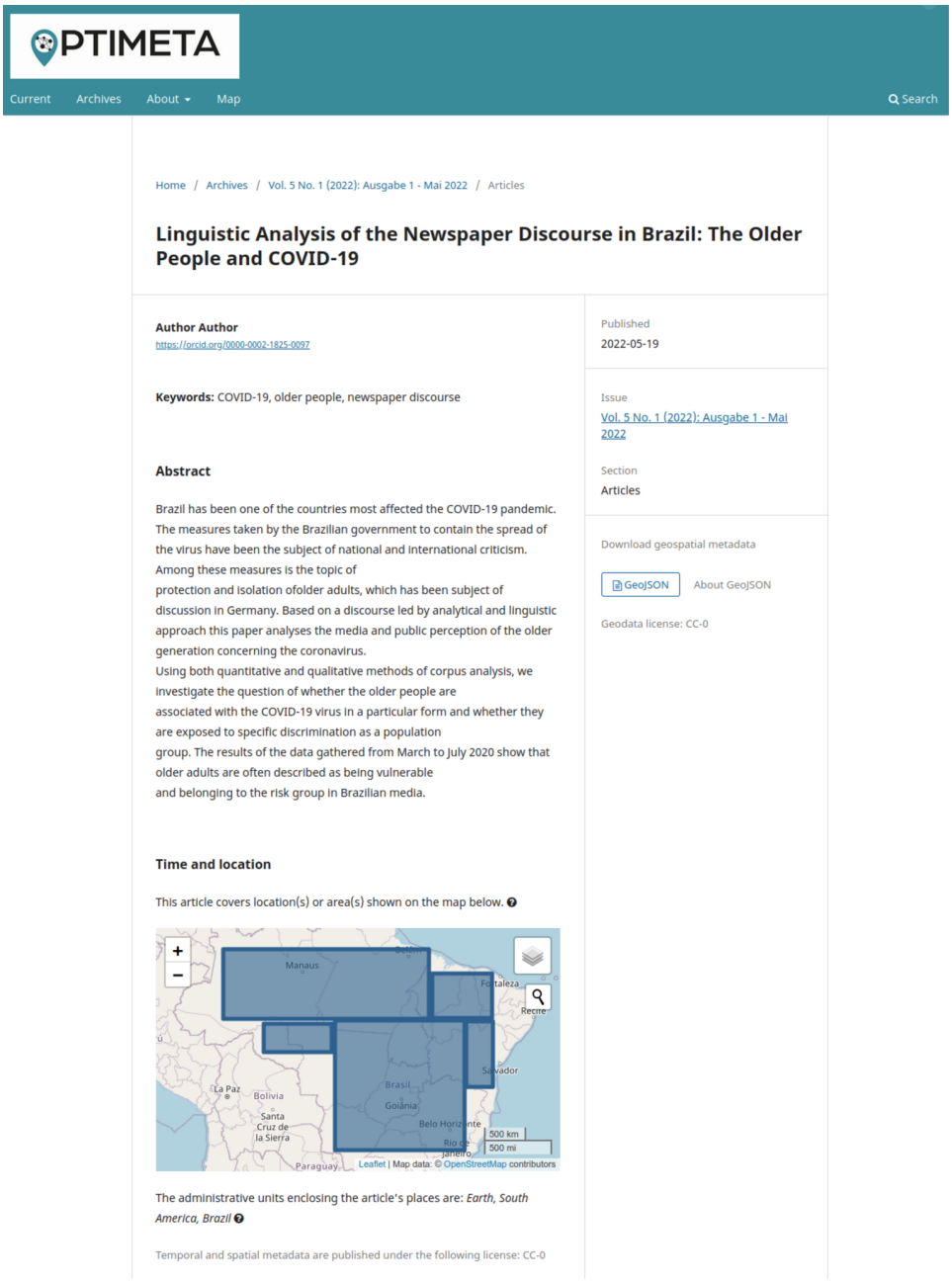
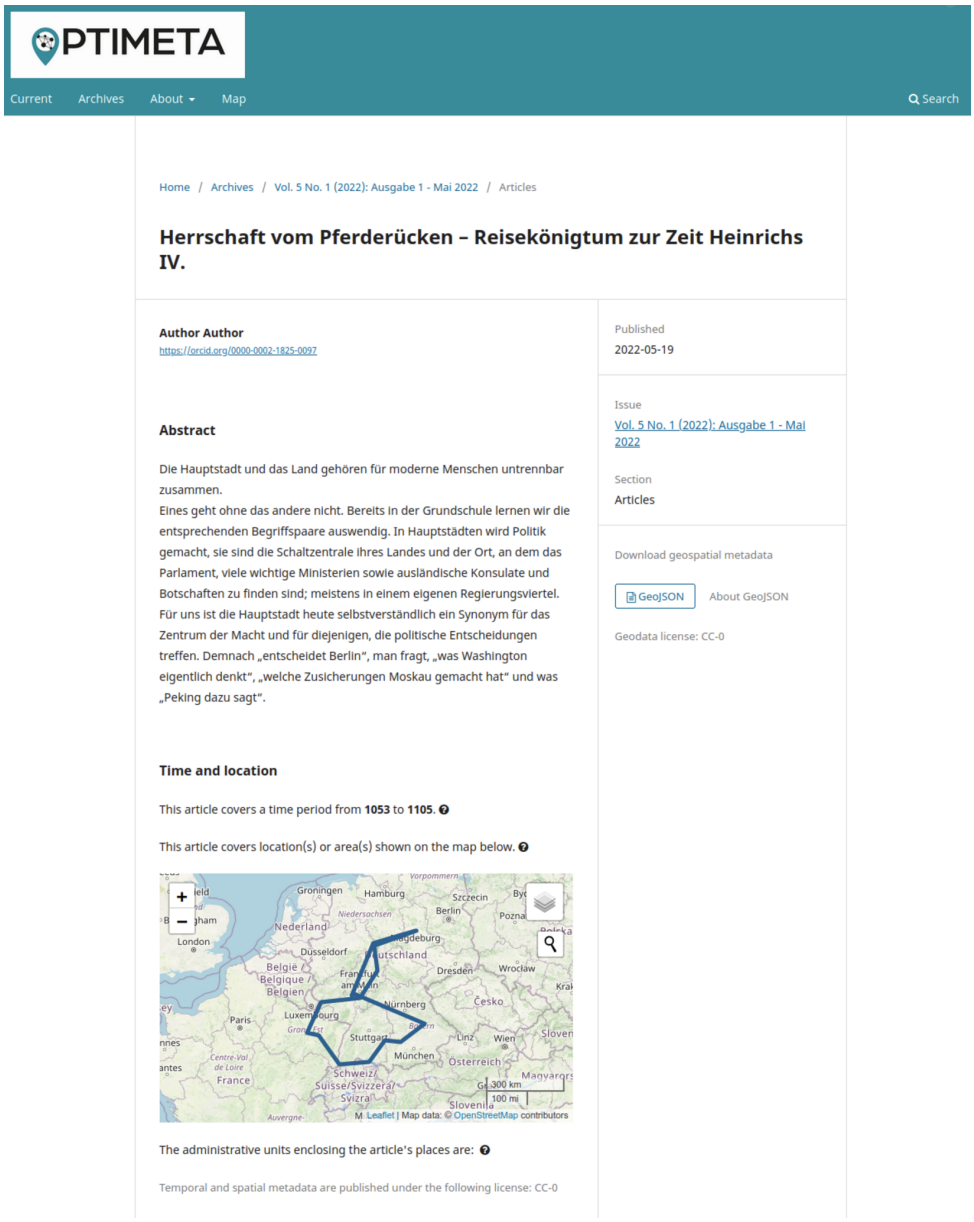


图 7. 含时空元数据的文章落地页截图；这篇论文包含多个覆盖巴西部分地区的多边形，地图下方有相应的文本描述，但没有时间段；右侧栏包含以 GeoJSON 格式下载元数据的按钮。



图。 8. 包含出版信息和时空元数据的文章落地页截图；地图上方标出了关注的时间段，描述该论文内容的几何图形由若干条代表旅行路线的折线构成。

随后，时空元数据会与论文一同发布在多个页面上：文章落地页（见图7和图8）、期刊期次落地页（见图9），以及期刊本身的独立页面。文章落地页还包含一个下载按钮，方便用户获取 GeoJSON 格式的时空元数据。如果用户将光标悬停在地图上的某个几何要素上，期刊落地页会显示如图9所示的同步高亮效果。地图上的要素以及上方列表中对应的论文都会以红色高亮显示。点击期刊期次地图或期刊地图，会打开一个小型弹出窗口（未显示），其中包含出版元数据（作者、标题等）以及指向文章落地页的链接。每幅地图下方都清楚地标明了

时空数据许可的说明，作者在创建元数据时需要同意该许可。许可固定为公有领域许可 CC-0，以确保尽可能广泛地使用。

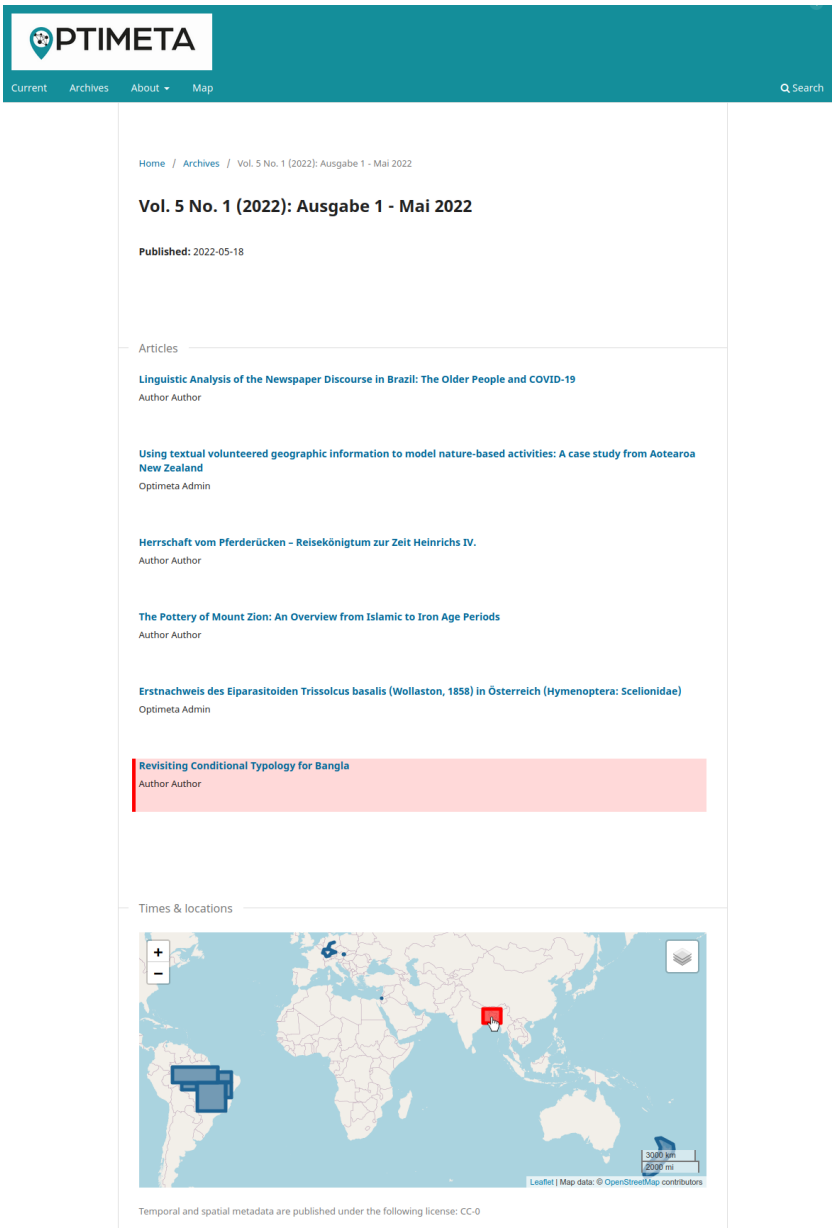


图 9. 公共演示期刊中期刊期次视图的屏幕截图，见 https://service.tib.eu/optimeta/index.php/optimeta/issue/view/1。标准 OJS 主题经过扩展，在期刊期次的文章列表下方增加了“时间与地点”部分。将鼠标光标悬停在地图底部的空间要素上时，几何图形及上方列表中对应的论文会以红色高亮显示。

上述页面均面向人类用户，但文章落地页的 HTML 网站中也以机器可读形式包含空间元数据和时间元数据。这些元数据字段

支持通过其他服务进行抓取和收集。图10展示了测试文章 HTML 页眉中显示的部分选定值，每个值均由一个名称定义，并在可用时指定明确的方案。除其他出版元数据外，空间元数据还以多种形式和模式纳入其中，包括都柏林核心元数据字段 `DC.SpatialCoverage`和 `DC.Coverage`。前者以完整 GeoJSON 记录的文本编码形式纳入（图10第10行），后者（第20行）则以行政区划单位的文本表示形式纳入，从最大的行政区划单位开始，逐级细化至更通用的字段`geo.placename`，其中包含可用的最小行政单位。在所提供的示例中，这一单位是国家名称，但也可以更加具体，例如城镇名称。最后，最小行政单位的边界矩形通过 `ISO19139` 字段给出，采用符合 ISO 19139 标准的地理边界框 XML 编码；同时也通过 `DC.box` 字段给出，采用以分号分隔的简单坐标列表表示限定该矩形的四个基点方向。时间元数据存储在字段`DC.temporal`和`DC.PeriodOfTime`中，二者均采用 ISO 8601 所定义的时间段文本表示形式。

初期开发阶段侧重于在投稿过程中收集元数据，并展示空间元数据。后续阶段将致力于开发更精细、更具交互性的时间元数据展示方式，具体包括：将论文的时间段放到同一时间轴上，以便查看特定期刊期次和整本期刊的时间分布；支持多个时间段；扩大所支持的历史日期格式范围（公元前、数百万年的时间跨度等）；增加直接输入坐标的功能；支持使用个性化参考数据集（与期刊的主题相关，例如生物圈、栖息地）来创建空间元数据；以及通过半自动方式推导时空元数据，例如从数据存储库中检索信息，或检查补充材料中的数据文件。

```
<!DOCTYPE html>
<html xml:lang="en-US" lang="en-US"><head>
<meta http-equiv="content-type" content="text/html; charset=UTF-8">
<meta charset="utf-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Test 3: Three | Journal of Optimal Geolocations</title>

<meta name="generator" content="Open Journal Systems 3.3.0.11">
<meta name="DC.temporal" scheme="ISO8601" content="2022-06-27/2022-06-30">
<meta name="DC.SpatialCoverage" scheme="GeoJSON" content="{&quot;type&quot;:&quot;FeatureCollection&quot;,&quot;features&qu
<meta name="geo.placename" content="Italian Republic">
<meta name="DC.box" content="name=Italian Republic; northlimit=47.091783741544; southlimit=35.49285259236; westlimit=6.6266
<meta name="ISO 19139" content="&lt;gmd:EX_GeographicBoundingBox&gt;&lt;gmd:westBoundLongitude&gt;&lt;gco:Decimal&gt;6.6266
<meta name="DC.PeriodOfTime" scheme="ISO8601" content="2022-06-27/2022-06-30">
<meta name="gs_meta_revision" content="1.1">
<meta name="citation_journal_title" content="Journal of Optimal Geolocations">
<meta name="citation_author" content="C Contributor">
<meta name="citation_title" content="Test 3: Three">
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
<meta name="DC.Coverage" xml:lang="en" content="Earth, Europe, Italian Republic">
<meta name="DC.Creator.PersonalName" content="C Contributor">
<meta name="DC.Title" content="Test 3: Three">
<meta name="DC.Type" content="Text.Serial.Journal">
<meta name="DC.Type.articleType" content="Articles">
```

图 10. 文章落地页源代码的截图，展示了 HTML 页眉中给出的部分 HTML 元属性，包括时空元数据的不同表示形式。

3.3 丰富学术元数据公地

我们将学术元数据公地视为知识公地的一个特殊子集（Hess 和 Ostrom，2006年；Mansell，2013年）。在这一子集中，开放许可、因而由集体所有的学术元数据集合由相关学术界及其他利益相关者共同治理和共享。这种公地有多种表现形式，通过网站或 API 等形式提供便于用户使用的界面，并支持数据的贡献与提取。维基数据就是这类界面的一个广为人知的例子。

我们　　既是……的组成部分，也通过这两个插件为学术元数据公地作出贡献：

1. 引文插件：借助此插件，我们通过开放 API，在出版时收集并分发经过丰富和用户验证的元数据，从而扩大研究信息的开放数据池。我们还纳入了目前未包含在标准科学计量数据源中的数据源，因为这些数据源使用的语言不同，或不受大型出版机构支持。2. 时空元数据插件：借助此插件，我们支持新的应用场景，例如基于位置的研究活动评估和基于位置的研究发现。这些场景可基于诸如以下问题：（1）所研究的地理区域；（2）基于时间段以及超越常用关键词和全文检索的关注领域，在研究成果之间建立新的跨学科联系。

4 讨论

上述 The OPTIMETA Way 提出了三项重要贡献，我们在此通过所展示的插件将其作为示例加以实现。第一项贡献是**扩大学术元数据公地**，纳入出版流程中生成的元数据。内置的机制可用于查找现有元数据，并在随后导入 ORCID 标识符等持久标识符，从而创建关联紧密的研究信息，并将其导出至现有的数据接收端。尽管非英语出版物在学术界发挥着重要作用（Kulczycki 等，2020年；Liu，2017年；Nazarovets 和 Mryglod，2021年），但目前它们的元数据在主要引文数据库中的收录比例并不相称（Tennant，2020年；Vera-Baceta 等，2019年）。

第二项贡献是让更多**开放获取期刊能够进入引文数据库**，并让更多基于学术元数据公地构建的其他服务得以发展。目前，OJS 已被来自 155 个国家的 25,000 种期刊使用（其中大多数来自全球南方），这些期刊以 56 种语言出版（Khanna 和 Willinsky，2022）。使用我们的插件将大幅降低独立期刊参与开放书目元数据共享的门槛，不过目前仅适用于使用 OJS 的期刊。这是朝着解决方案迈出的一大步，但我们尚无法彻底消除这一问题。因此，我们希望未来能在其他出版平台上实施 The OPTIMETA Way。

第三项贡献是**通过纳入时空元数据来拓展学术元数据公地**，从而推动开放元数据用于新的应用场景。正如 Niers 和 Nüst（2020）所述，时空元数据可用于发现研究地理覆盖范围中的偏差。例如，某一领域的研究可能过度聚焦于一个地区，而忽略了其他同样值得关注的地区。时空元数据还有助于识别不同研究成果之间的联系，并加深我们对某一研究领域内地理关系和时间关系的理解。此外，可视化，尤其是地图形式的可视化，有助于传播研究内容，也能彰显研究整体的必要性。迄今为止，时空元数据的可用性仍然有限；随着 geoplugin 的发布，我们将为开放学术出版生态系统增添一个新组件。此外，geoplugin 还将为基于位置数据的研究活动评估开辟新的应用场景：（1）回答有关研究区域的问题，例如展示研究位置的特定覆盖范围或分布情况；（2）根据时间段和关注区域，发现研究成果之间可能有价值的跨学科联系，而这些联系可能无法通过常用关键词和全文检索发现。例如，将有关中欧社会问题的历史研究与当前的健康研究联系起来。未来，这些元数据可用于构建平台，依据用户定义的时空兴趣及时推送出版物通知；也就是说，当有新出版物涉及某评估方案特别关注的区域时，系统或用户即可收到通知。刻意降低精度的坐标可用于保护人类受试者的隐私，或隐藏受保护实体的位置。

研究论文之间**空间关系的整合作用**已得到其他人的认可。然而，现有解决方案无一遵循 The OPTIMETA Way，因此要么过于复杂，要么未融入出版工作流程，要么无助于开放的学术元数据公地。例如，JournalMap（https://www.journalmap.org/；Karl等，2013年）展示研究论文的位置和出版元数据（标题、摘要等），以支持基于地图的发现。然而，JournalMap 仅限于点几何图形；尽管它提供 API，并与一些出版商开展合作（https://www.journalmap.org/publishers；https://web.archive.org/web/20161016000907/https://newsroom.taylorandfrancisgroup.com/news/press-release/taylor-francis-journal-map-partnership#.WALFJmF_o88），其数据并未完全开放。自网站创立以来，网站上就一直公告称数据下载选项“即将推出”（参见 https://web.archive.org/web/20130615020154/https://www.journalmap.org/downloads）；其许可协议定义为知识共享署名—相同方式共享（CC-BY-SA，https://www.journalmap.org/developer/documentation/1-0），但使用条款大幅限制了许可协议的条款，并禁止将数据用于商业用途（https://www.journalmap.org/terms-of-use）。网站确实提供了一些高级筛选选项，包括额外的主题筛选选项。然而，网站上宣传的商业选项与我们对知识进步的理解背道而驰。其次，Kmoch 等（2018年）分析了地球科学期刊中的论文，以便从论文书目元数据中的非结构化信息自动提取空间元数据。提取的数据随后发布到了一个公共地理空间目录服务中。然而，这种方法需要大量技术知识，而且缺少人工质量检查，因为并非所有数据都由最合适的专家检查。因此，尽管这是一种应对过去未收集时空

元数据这一现状的有效方法，但它既不是完整的解决方案，也不符合 The OPTIMETA Way。Garzón 和 Nüst（2021b）采用了类似方法，使用 geoextent 工具为通用研究数据仓储创建了一个基于地理空间元数据的发现索引。他们采用蛮力方法，尽可能从多种地理空间文件格式中提取空间范围。如果论文恰当地引用了所用数据，这或许可以作为丰富文章元数据的中间方法；不过，由于数据集可能因各种原因被引用，可能仍需人工核验。目前，一个用于从多种期刊收集时空元数据的搜索门户（参见图 2）——OPTIMAP——正在开发中（参见https://optimap.science/和https://github.com/ifgi/optimetaPortal）。

开放引文信息的可用性所带来的**优势**毋庸置疑。Peroni 和 Shotton（2020年）列出了广泛受益的利益相关者：不属于“那些有能力订阅商业引文索引 WoS 和 Scopus 的研究型大学精英圈子”的研究人员、希望提供自身研究数据的文献计量学家、图书馆员、资助方、科研管理人员等。引文插件的一项核心价值在于，它让大量开放获取期刊能够在开放科研公地中共享作者的出版物，不受语言或学科领域限制，也无论期刊是由出版者主导还是由独立学者主导。因此，长期以来常常遭到可耻忽视的学术研究长尾将得以显现，并有机会被妥善整合，尤其是常被忽视的非英语文献（Lazarev 和 Nazarovets，2018年）。

关于**科学传播与评估**，Krüger（2020年）阐述了如何通过对研究成果及相关元数据和指标进行衡量绩效的量化，来应对社会对科学的不信任。她令人信服地指出，文献计量基础设施和应用程序各自都有一套关于如何通过自身使用来理解研究的观念。通过拓展我们所拥有的元数据的范围和类别，The OPTIMETA Way 虽无法完全避免这一问题，但可以限制人们通过元数据中的数字表示来观察、衡量、评估和视为知识的内容受到扭曲的程度，从而应对问题的规模。这样一来，时空元数据和引文元数据的可用性得到提升，研究评估也就能够更迅速、轻松、透明且负责任地开展。**元数据方面的理想要求**可以包括对研究问题、方法、与资助编号或资助方等外部实体的关联、实物样本标识符或仪器标识符进行机器可操作的描述。

我们估计，插件被**不道德地使用**的风险较低，也看不出时空数据遭到滥用的特别可能性。即使在最坏的情况下，蓄意提供有缺陷的元数据会降低内容的可发现性，但不会影响识别出版物的其他方式。诚然，就引文元数据而言，当恶意方关注引文计数等不负责任的研究评估方法时，这些插件确实降低了提交伪造引文信息的门槛。然而，只有在作者和责任编辑都怀有恶意时，引文元数据才可能遭到滥用；同时，我们也鼓励责任编辑在出版前核查引文信息。此外，向公共数据库提交数据并非匿名操作；一旦发现滥用行为，便可撤销相关数据，并禁止相关账户继续上传数据。

展望**未来**，此处介绍的 The OPTIMETA Way 实现方案可根据不同学科期刊所报告的经验，加入新功能并提升易用性。在技术方面，我们设想可以将更先进的方法（例如机器学习方法，或从 PDF 文件和数据文件中提取信息）集成到 OPTIMETA插件中，以提高元数据的易用性并扩充其范围。例如，可以用类似的半自动方式收集致谢（如资助机构和资助项目编号）、作者贡献（例如根据采用 CRediT声明的致谢部分）或学科分类，并按照 The OPTIMETA Way 将其公开存入数据库。不过，为了保障数据质量，不应省略人类专家的验证。元数据增强的适用范围可扩大至预印本、专著和编著集，但需要在元数据字段中加入语义明确的属性，以区分未经评审和经过评审的研究成果。为了支持预印本和图书，除了其他图书出版平台外，还可以将引文插件和地理插件移植到 PKP 的预印本平台。

未来，为应对更开放、也更紧迫的社会研究**挑战**，研究的粒度和速度预计都会发生变化。为了有效沟通，我们将更加需要及时且经过验证的研究元数据。研究成果正越来越多地分阶段发表（例如 Octopus，https://www.octopus.ac/about），并以独立构件的形式发布（例如：想法/文本/阐释、代~~码/软件、数据），而非等到~~期刊期次可能已经解决相关问题数年后，才以润色完成的文本成果形式发表。研究人员将更经常地通过免费基础设施，以便于获取的方式分享成果；同行评审实践也将随之调整，例如采用叠加期刊（Brown 2010年；Rousi 和 Laakso 2022年）。然而，与确保我们开发的插件得到长期维护这一组织层面的挑战相比，这些技术挑战微不足道。当前的资助促成了稳定插件的开发，并支持将其发送给选定的合作伙伴进行评估；与此同时，越来越多资助核心研究软件的项目也陆续设立（例如，https://chanzuckerberg.com/eoss/），但我们仍面临先有鸡还是先有蛋的问题。期刊若要广泛采用这些插件，就要求承诺长期维护软件；~~而只有在能够证明其得到广泛~~使用时，才更容易获得用于维护和改进插件的资助。

向开放基础设施中的开放获取和 FAIR 研究信息转变的文化进程（Hauschke 等，2021a；Hendricks 等，2021年）将在不同国家和学科以不同速度发生，并促使各种平台并存。与当今的集中式系统及大型出版商对其拥有的控制权相比，这是一项优势；但同时也是一项挑战，因为这些服务必须能够互联并交换元数据。因此，我们希望 The OPTIMETA Way 能够推广至学术开放基础设施中的其他组成部分，从而以尽可能减少对现有工作流程影响的方式，由最了解情况的一方收集有针对性、新颖甚至小规模的元数据属性，并广泛、开放且迅速地共享这些属性，以推动知识发展。

5 利益冲突

作者声明，本研究是在不存在任何可能被视为潜在利益冲突的商业或财务关系的情况下开展的。

6 作者贡献

作者对本研究的贡献均等。所述插件由 Daniel Nüst（时空元数据插件）和 Gazi Yücel（引文插件）开发。

7 资助

作者感谢德国联邦教育与研究部（BMBF）在 OPTIMETA 项目框架内提供的资金支持（资助编号：16TOA028A 和 16TOA028B）。

8 致谢

作者谨向项目合作伙伴致谢，感谢大家持续讨论如何改进 OPTIMETA Way。特别感谢 Open Citations 和 PKP 积极参与技术讨论并提供支持。感谢 Tom Niers 开发时空元数据插件的首个原型，也感谢 Svantje Lilienthal 对引文插件早期概念讨论所作的贡献。感谢明斯特大学学术编辑服务的 Julie Davies 协助修订稿件。

9 参考文献

Bijsterbosch, Magchiel/Dunning, Alastair/Jansen, Darco/Haring, Max/Rijcke, Sarah de/Vanderfeesten, Maurice（2022年）。开放研究信息的七项指导原则。https://doi.org/10.5281/zenodo.6074943。Brembs, Björn（2021年）。关于爱思唯尔出版成本的推文。在线查阅：https://twitter.com/brembs/status/1440942528023470081。Brown, Josh（2010年）。叠加期刊简介。在线查阅：https://discovery.ucl.ac.uk/id/eprint/19081/。Cagan, Ross（2013年）。旧金山研究评估宣言。《Disease models & mechanisms》6（4），869–870。https://doi.org/10.1242/dmm.012955。Conlon, Michael/Woods, Andrew/Triggs, Graham/O'Flinn, Ralph/Javed, Muhammad/Blake, Jim/Gross, Benjamin/Ahmad, Qazi/Ali, Sabih/Barber, Martin/Elsborg, Don/Fofack, Kitio/Hauschke, Christian/Ilik, Violeta/Khan, Huda/Lawless, Ted/Levernier, Jacob/Lowe, Brian/Martin, Jose/McKay, Steve/Porter, Simon/Walther, Tatiana/White, Marijane/Wolff, Stefan/Younes, Rebecca（2019年）。VIVO：一种用于研究发现的系统。《Journal of Open Source Software》4（39），1182。https://doi.org/10.21105/joss.01182。euroCRIS（2018年）。第14届国际当前研究信息系统会议。https://doi.org/10.25798/2y3h-k877。联邦地理数据委员会。ISO 191** 地理空间元数据标准套件。在线查阅：https://www.fgdc.gov/metadata/iso-suite-of-geospatial-metadata-standards。Franceschi-Bicchierai, Lorenzo（2022年）。学术期刊声称对 PDF 文件进行指纹识别是为了防范“勒索软件”，而非监控。在线查阅：https://www.vice.com/en/article/4aw48g/academic-journal-claims-it-fingerprints-pdfs-for-ransomware-not-surveillance。

Garfield, Eugene（1952年）。《化学文献危机》。在美国化学会马里兰分会的演讲，1952年9月22日。在线获取：http://www.garfield.library.upenn.edu/papers/crisisinchemlit1952.html。Garzón, Sebastian/Nüst, Daniel（2021a）。使用 geoextent 探索研究数据仓储。在线获取：https://earthcube2021.github.io/ec21_book/notebooks/ec21_garzon_etal/showcase/SG_01_Exploring_Research_Data_Repositories_with_geoextent.html。Garzón, Sebastian/Nüst, Daniel（2021b）。使用 geoextent 探索研究数据仓储。Zenodo。https://doi.org/10.5281/zenodo.5496311。Gonzales, Brighid M.（2014）。将图书馆连接到 Web：关联数据与书目记录的未来。《Information Technology and Libraries》（在线版）33（4），10–22。https://doi.org/10.6017/ital.v33i4.5631。Hauschke, Christian/Nazarovets, Serhii/Altemeier, Franziska/Kaliuzhna, Nataliia（2021a）。开放基础设施中的 FAIR 研究信息路线图。《Journal of Library Metadata》21（1-2），45–61。https://doi.org/10.1080/19386389.2021.1999156。Hauschke, Christian/Nüst, Daniel/Cordts, Anette/Lilienthal, Svantje（2021b）。OPTIMETA——通过开放引文和时空元数据强化开放获取出版系统。《Research Ideas and Outcomes》7。https://doi.org/10.3897/rio.7.e66264。Hendricks, Ginny/Kramer, Bianca/Maccallum, Catriona J./Manghi, Paolo/Neylon, Cameron（2021）。现在正是携手建设学术元数据开放基础设施的时候。在线获取：http://eprints.lse.ac.uk/id/eprint/112845。Hess, Charlotte/Ostrom, Elinor（2006）。《理解作为公地的知识》。麻省理工学院出版社。ISBN 9780262516037。Hey, Tony（2015）。数据策展——过去、现在与未来。伦敦，第十届国际数字策展会议，2015年2月9日。在线获取：https://www.dcc.ac.uk/sites/default/files/documents/IDCC15/Presentations%20Day%201/(Feb15)%20TonyHey%2 0-%20Data%20Curation%20talk%20(final).pdf。Hicks, Diana/Wouters, Paul/Waltman, Ludo/Rijcke, Sarah de/Rafols, Ismael（2015）。文献计量学：《莱顿宣言》与科研评价指标。《自然》520（7548），429–431。https://doi.org/10.1038/520429a。Jaradeh, Mohamad Yaser/Oelen, Allard/Farfar, Kheir Eddine/Prinz, Manuel/D'Souza, Jennifer/Kismihók, Gábor/Stocker, Markus/Auer, Sören（2019）。开放研究知识图谱：新一代语义学术知识基础设施。载于：Mayank Kejriwal/Pedro Szekely/Raphaël Troncy（编）。第十届国际知识捕获会议论文集，K-CAP '19：知识捕获会议，美国加利福尼亚州玛丽娜德尔雷，2019年11月19日至21日。美国纽约州纽约市，计算机协会，243–246。https://doi.org/10.1145/3360901.3364435。Kaliuzhna, Nataliia/Altemeier, Franziska（2021）。迈向研究信息的 FAIR 原则：系列研讨会报告。《Ukrainian Journal on Library and Information Science》（7），128–132。在线获取：http://librinfosciences.knukim.edu.ua/article/view/233322。Karl, Jason W./Herrick, Jeffrey E./Unnasch, Robert S./Gillan, Jeffrey K./Ellis, Erle C./Lutters, Wayne G./Martin, Laura J.（2013）。通过地理语义检索，从已发表的研究中发现生态学相关知识。《BioScience》63（8），674–682。https://doi.org/10.1525/bio.2013.63.8.10。Khanna, Saurabh/Willinsky, John（2022）。学术传播开放基础设施的负责人可以采取哪些措施，应对可能存在的掠夺性做法。https://doi.org/10.1590/SciELOPreprints.3474。Kmoch, Alexander/Uuemaa, Evelyn/Klug, Hermann/Cameron, Stewart（2018）。增强位置相关的水文地质知识。《ISPRS International Journal of Geo-Information》7（4），132。https://doi.org/10.3390/ijgi7040132。Krüger, Anne K.（2020）。量化 2.0？学术评价中的文献计量基础设施。《Politics and Governance》8（2），58–67。https://doi.org/10.17645/pag.v8i2.2575。Kuhn, Werner（2012）。跨学科研究的空间信息核心概念。《International Journal of Geographical Information Science》26（12），2267–2276。https://doi.org/10.1080/13658816.2012.722637。Kulczycki, Emanuel/Guns, Raf/Pölönen, Janne/Engels, Tim C. E./Rozkosz, Ewa A./Zuccala, Alesia A./Bruun, Kasper/Eskola, Olli/Starčič, Andreja Istenič/Petr, Michal/Sivertsen, Gunnar（2020）。多语种出版在

社会科学与人文学科：一项涵盖七个欧洲国家的研究。《信息科学与技术协会期刊》71 (11), 1371–1385. https://doi.org/10.1002/asi.24336. Lausch~~er, Anne, Kai Eckert, Lukas~~ Galke, Ansgar Scherp, S yed Tahseen Raza Rizvi, Sheraz Ahmed, Andreas Dengel, Philipp Zumstein, and Annette Klein (2018年). 开放引文数据库：助力图书馆为开放且互联的引文图谱贡献力量。载于第 18 届 ACM/IEEE 数字图书馆联合会议（JCDL '18）论文集。计算机协会，纽约，NY，美国，109–118. https://doi.org/10.1145/3197026.3197050. Lazarev, Vladimir S./Nazarovets, Serhii A. (2018年). 不要忽视对非英语期刊的引用。《自然》556 (7700), 174. https://doi.org/10.1038/d41586-018-04169-2. Liu, Weishu (2017年). 非英语论文在学术传播中不断变化的角色：来自科学网三个期刊引文索引的证据。《学术出版》30 (2), 115–123. https://doi.org/10.1002/leap.1089. Manghi, Paolo/Mannocci, Andrea/Osborne, Francesco/Sacharidis, Dimitris/Salatino, Angelo/Vergoulis, Thanasis (2021年). 科学知识图谱与研究影响力评估的新趋势。《定量科学研究》2 (4), 1296–1300. https://doi.org/10.1162/qss_e_00160. Mansell, Robin (2013年). 利用数字众包信息资源：管理新兴的信息公地。《公地国际期刊》7 (2), 255. https://doi.org/10.18352/ijc.395. Massari, Arcangelo/Heibi, Ivan (2022年). 如何在OpenCitations认可格式中组织引文数据和书目元数据。https://doi.org/10.48550/arXiv.2206.03971. Nazarovets, Serhii (2019年). 乌克兰开放引文索引（OUCI）：构想、工作原理与发展前景。《乌克兰的知识产权》, 3, 10-13. Nazarovets, Serhii/Mryglod, Olesya (2021年). 乌克兰人文学科的多语性：全球视角。载于：第18届计量学与信息计量学国际会议，鲁汶大学，2021年，1521–1522. http://eprints.rclis.org/42221/. Nielsen, Finn Årup, Mietchen, Daniel, and Willighagen, Egon (2018年). 地理空间数据与Scholia。GeoLD2018与QuWeDa2018联合论文集：第3届地理空间关联数据国际研讨会／第2届数据网络查询研讨会（GeoLD-QuWeDa-2018）。CEUR研讨会论文集。ISSN 1613-0073. http://ceur-ws.org/Vol-2110/paper3.pdf Niers, Tom/Nüst, Daniel (2020年). 学术出版中的地理空间元数据及其发现。Septentrio会议系列 (4). https://doi.org/10.7557/5.5590. Peroni, Silvio/Shotton, David (2012年). FaBiO 和 CiTO：用于描述书目资源与引文的本体。《网络语义学期刊》 17, 33–43. https://doi.org/10.1016/j.websem.2012.08.001. Peroni, Silvio/Shotton, David (2020年). OpenCitations：支持开放学术的基础设施组织。《定量科学研究》1 (1), 428–444. https://doi.org/10.1162/qss_a_00023. Pooley, Jeff (2022年). 监控式出版。《电子出版期刊》25 (1). https://doi.org/10.3998/jep.1874. Priem, Jason/Piwowar, Heather/Orr, Richard (2022年). OpenAlex：一个完全开放的学术作品、作者、出版渠道、机构和概念索引。https://doi.org/10.48550/arXiv.2205.01833. Rasberry, L., Willighagen, E., Nielsen, F., Mietchen, D. (2019年). 强化Scholia：借助维基数据推动知识发现与研究评估。Research Ideas and Outcomes 5: e35820. https://doi.org/10.3897/rio.5.e35820. Rousi, Antti Mikael/Laakso, Mikael (2022年). 叠加期刊：当前格局研究。https://doi.org/10.48550/arXiv.2204.03383. Smith, Arfon M./Katz, Daniel S./Niemeyer, Kyle E. (2016年). 软件引用原则。《PeerJ计算机科学》2, e86 https://doi.org/10.7717/peerj-cs.86。 Smith, MacKenzie/Barton, Mary/Branschofsky, Margret/McClellan, Greg/Walker, Julie Harford/Bass, Mick/Stuve, Dave/Tansley, Robert（2003）。DSpace。《D-Lib Magazine》9 (1)。https://doi.org/10.1045/january2003-smith。 Starr, Joan/Gastl, Angela（2011）。被引用：面向 DataCite 的元数据方案。https://doi.org/10.3929/ethz-a-006687003 .

Tennant, Jonathan（2020）。科学网和 Scopus 并非全球知识数据库。《美国细胞生物学学会》（2012）。2012年 ASCB 年会。ASCB。https://doi.org/10.25798/najp-rd56。Turki, Houcemeddine/Hadj Taieb, Mohamed Ali/Ben Aouicha, Mohamed/Fraumann, Grischa/Christian Hauschke/Heller Lambert（2021）。利用书目元数据增强学术出版物的知识图谱提取与验证。《研究计量与分析前沿》6，694307。https://doi.org/10.3389/frma.2021.694307。Vera-Baceta, Miguel-Angel/Thelwall, Michael/Kousha, Kayvan（2019）。科学网和 Scopus 的语言覆盖范围。《科学计量学》121（3），1803–1813。 https://doi.org/10.1007/s11192-019-03264-z。维基百科（2022）。地理空间元数据。在线获取：https://en.wikipedia.org/w/index.php?title=地理空间_元数据&oldid=1063813268。

Wilkinson, Mark D./Dumontier, Michel/Aalbersberg, I. Jsbrand Jan/Appleton, Gabrielle/Axton, Myles/Baak, Arie/Blomberg, Niklas/Boiten, Jan-Willem/Da Silva Santos, Luiz Bonino/Bourne, Philip E./Bouwman, Jildau/Brookes, Anthony J./Clark, Tim/Crosas, Mercè/Dillo, Ingrid/Dumon, Olivier/Edmunds, Scott/Evelo, Chris T./Finkers, Richard/Gonzalez-Beltran, Alejandra/Gray, Alasdair J. G./Groth, Paul/Goble, Carole/Grethe, Jeffrey S./Heringa, Jaap/Hoen, Peter A. C. 't/Hooft, Rob/Kuhn, Tobias/Kok, Ruben/Kok, Joost/Lusher, Scott J./Martone, Maryann E./Mons, Albert/Packer, Abel L./Persson, Bengt/Rocca-Serra, Philippe/Roos, Marco/van Schaik, Rene/Sansone, Susanna-Assunta/Schultes, Erik/Sengstag, Thierry/Slater, Ted/Strawn, George/Swertz, Morris A./Thompson, Mark/van der Lei, Johan/van Mulligen, Erik/Velterop, Jan/Waagmeester, Andra/Wittenburg, Peter/Wolstencroft, Katherine/Zhao, Jun/Mons, Barend（2016年）。科学数据管理与管理服务的FAIR指导原则。《科学数据》3，160018。 https://doi.org/10.1038/sdata.2016.18。

John Willinsky（2005年）。开放期刊系统。Library Hi Tech 23 (4), 504–519。https://doi.org/10.1108/07378830510636300。