



文献 CSTR:

32001.14.11-6035.csd.2023.0187.zh



文献 DOI:

10.11922/11-6035.csd.2023.0187.zh

文献分类: 信息科学

收稿日期: 2023-12-21

开放同评: 2023-12-22

录用日期: 2024-01-22

发表日期: 2024-02-23

\* 论文通信作者

胡良霖: hull@cnic.cn

## 《数据论文出版元数据》国家标准研制与实践

高瑜蔚<sup>1,2,3</sup>, 朱艳华<sup>1,3</sup>, 胡良霖<sup>1,3\*</sup>, 孔丽华<sup>1</sup>, 赵欢<sup>1,3</sup>

1. 中国科学院计算机网络信息中心, 北京 100083

2. 首都师范大学中国语言智能中心, 北京 100048

3. 国家基础学科公共科学数据中心, 北京 100083

**摘要:** 2023年8月,《GB/T 42813-2023 数据论文出版元数据》国家标准正式发布,该标准规定了数据论文出版元数据的内容与结构,包括数据集元数据、数据论文元数据及其描述方法、扩展原则与方法。适用于数据论文及其关联数据集的描述、保存、在线出版和共享服务等。该标准的发布标志着首次从国家层面对数据论文出版这一科学数据开放共享新模式进行了总体约定,解决了数据出版领域缺少标准规范指导的问题。有助于数据出版机构快速地发布数据论文和关联数据,为科学数据确权、产权保护提供支撑,对于推动科学数据开放共享生态建设具有重要意义。自标准发布以后,一些数据期刊或科学数据出版平台开始使用标准规定的元素与定义,开展较为广泛的实践。

**关键词:** 开放科学; 数据论文; 科学数据出版; 元数据; 标准规范; 科学数据集

### 1 标准研制背景

当前,我国正处于实施创新驱动发展战略和建设世界科技强国的关键时期,加强和规范对科学数据的管理是提升我国科技创新能力和保障国家安全的重要方式和手段<sup>[1]</sup>。促进科学数据的开放、流转、利用和增值,推动科学研究和科技成果产出,对于提高国家财政投入产出效率、促进我国科技创新和经济社会发展具有十分重要的意义和支撑作用。科学数据开放共享,既能避免重复实验和研究造成的浪费,缩短研究周期,打破科学垄断,解决“学术孤岛”等问题,也能促进科学数据的持续、有序积累,提升数据的质量,实现科学数据价值最大化。无论是科学数据共享实践,还是理论研究,都将科学数据出版作为推动科学数据开放共享的有效方式,逐渐受到重视,与传统的科学数据共享体系形成有效的互补关系。

近年来,以数据出版、数据论文等方式发布科学数据已成为世界范围内科学数据开放共享的新模式。2014年,国际学术期刊《自然》杂志创立数据期刊《Scientific Data》;2015年以来,在国内,《中国科学数据(中英文网络版)》《全球变化数据学报》《农业大数据学报》等数据期刊先后创办,带动了数据论文大幅增长。2018年3月,国务院办公厅发布的《科学数据管理办法》(国办发〔2018〕17号)明确提出“主管部门和法人单位应积极推动科学数据出版和传播工作,支持科研人员整理发表产权清晰、准确完整、共享价值高的科学数据”,鼓励科学数据出版,促进科技创新<sup>[2]</sup>。数据出版是指通过一定的公共机制发布科学数据集,使得公众根据一定规则发现、获取、评价和应用这些数据的过

程，其中数据论文出版是得到广泛认可的重要形式<sup>[3]</sup>。CANDELA L 等<sup>[4]</sup>认为数据论文概念至少包含 2 个具体信息对象——数据集和数据论文本身。具体来说，数据论文是通过网络在线方式出版的一种学术出版物<sup>[5]</sup>，一般由出版系统管理和发布。其关联数据集一般由独立的科学数据存储库管理和发布，并通过下载链接、标识和引用方式与数据论文实现关联。作为复杂的系统工程，其既有传统学术出版的特点，又具有数据发布且与数据论文关联等带来的新需求。

近年来，全球成立多个以数据论文方式出版科学数据的数据期刊，并取得良好成效。国内数据期刊虽陆续建立并开展大量实践，但是标准化工作尚未开展，缺少相关国家标准和行业标准规范统一指导。在此背景下，《数据论文出版元数据》（计划号 20205082-T-306）于 2020 年 12 月 28 日正式获得立项。2023 年 8 月，国家标准化委员会通过官网发布《中华人民共和国国家标准公告（2023 年第 7 号）》，《GB/T 42813-2023 数据论文出版元数据》正式发布并同时实施<sup>[6]</sup>，对我国科学数据论文出版标准化和科学数据出版工作具有重要意义。

## 2 标准研制过程

《数据论文出版元数据》作为自主研制标准，由科技部提出，全国科技平台标准化技术委员会（SAC/TC486）归口。联合 12 个国家科学数据中心及 15 家部委科学数据中心在内的 20 余家研究机构，历时五年（含立项研制的两年）完成了《数据论文出版元数据》国家标准的研制，解决了长期以来数据出版领域缺少标准规范指导的问题。标准主要起草单位包括中国科学院计算机网络信息中心、国家科技基础条件平台中心、广州物联网研究院、中国标准化研究院、中国科学院地理科学与资源研究所、中国科学院微生物研究所、中国极地研究中心、国家海洋信息中心、北京师范大学、中国农业科学院农业信息研究所、中国林业科学研究院资源信息研究所、中国科学院空间应用工程与技术中心、中国科学院空天信息创新研究院、中国科学院昆明植物研究所、中国地质调查局发展研究中心、中国科学院昆明动物所、中国医学科学院、中国科学院青藏高原研究所、中国科学院高能物理研究所、中国科学院植物研究所。

标准研制期间，标准起草组对科学数据出版要求、数据期刊平台和科学数据存储库发展现状，特别是元数据制定情况开展深入研究分析，采用文献调研、组内讨论和专家研讨等方式开展工作。围绕科学数据这一重要科研成果，在保护数据知识产权、数据来源及数据质量控制措施、实现数据溯源、推动数据期刊平台建设等方面攻克模式和技术难点，明确数据论文出版元数据要素。《数据论文出版元数据》的主要研制过程经历了标准起草、征求意见、标准审查、上报报批、发布实施等<sup>[7-9]</sup>，详见表 1。

## 3 标准主要内容

### 3.1 数据论文出版的主要特征

在科学数据相关实践中，元数据的价值主要体现在支撑整个数据生命周期中的标准化、可发现、再利用、可评价等方面。通过元数据支持数据仓储中科学数据的可发现、长期保存和共享利用。由于科学数据结构内容的复杂性以及出版模式、出版格式的多样性，构建标准体系对科学数据论文及相关数据的提交、组织、存储、检索尤为重要。科学数据出版标准化对象包括数据论文模板、数据资源元数据、数据及数据论文的开放获取方式等，籍此实现资源（数据实体、数据论文等）描述、

展示、访问、应用等方面的标准化，支持科学数据出版的规模化发展和高效应用。

表 1 国家标准《数据论文出版元数据》主要研制过程

Table 1 Key stages in developing the national standard for *Data paper publication metadata*

序号	阶段名称	主要内容
1	标准起草	2021年3月19日，国家标准《数据论文出版元数据》立项启动会召开，围绕标准框架、重点研究范围和应用场景展开讨论，明确了下一步研制工作。
		2021年4月-9月，成立标准起草组，多次召开标准编制组内部讨论会，对标准草案进行调整，编写完成标准草案。
		2021年10月，标准起草组在第七届中国科学数据大会上广泛征求意见。
		2022年1-4月，标准起草组调研了覆盖多领域的16家国内外数据期刊数据论文出版进展，内容涉及数据论文描述元素、引用格式、版本号、共享方式等重要内容，并对关联科学数据存储库进行深入研究。
		2022年5月27日，标准起草组召开专家研讨会。来自国家科技基础条件平台中心和地方平台、中国标准化研究院、高校、数据出版机构、国家科学数据中心、国家资源库的27位专家对标准草案征求意见稿提出标准规范层面、标准协调层面、数据出版和应用层面的修改建议，大家一致认为该项标准编制规范、合理，将为数据出版应用提供约束和指导作用，对提高数据出版的质量具有重要的现实意义。
2	征求意见	2022年7月19日-9月19日，标准在全国标准信息公共服务平台公开征求意见，同时面向51家国家科技资源共享服务平台（国家科学数据中心、国家生物种质和实验材料资源库）定向征求意见。
3	标准审查	2022年9月27日，召开专家研讨会，起草组汇报了标准研制进展、研究内容、面向全网公开征求意见及面向51个国家科技资源共享服务平台意见征求情况、标准应用实践情况、问题与计划等。会议提出了细节完善和未来实施建议，明确形成送审稿及后续提交全国科技平台标准化技术委员会（SAC TC486）评审的工作计划。
		2022年10月，标准进入标准审查阶段，送审稿提交至TC486开展函评，并在全国标准信息公共服务平台公示。
4	上报报批	2022年11月，标准通过TC486审查，并根据修改意见形成报批稿。
5	发布实施	2023年8月，国家标准化管理委员会通过官网发布了《中华人民共和国国家标准公告（2023年第7号）》，《GB/T 42813-2023 数据论文出版元数据》正式发布并同时实施。

研究发现，数据论文出版具有以下五个主要特征：

特征一：通过公开评议和同行评审严控数据质量。数据论文具有清晰的出版流程，通过公开评议和同行评审严控数据质量。每篇数据论文和关联数据都在编辑委员会监督下实施公开评议和同行专家评议，接受大众监督和快速验证。评审通过后进行在线发布和共享，论文和数据质量实现内容和过程管理。

特征二：强调数据观点，不关注创新和引领性。数据论文的首要目的是描述数据及其生成、收集条件，学术论文中的假设分析和结论成果则一般并不在数据论文之列。更侧重于对数据采集和处理、质量控制和评估，以及数据使用方法等方面的描述，结合数据集本身的元数据信息，最大限度地消除数据集重用过程中的阻碍。

特征三：保护知识产权，遵循通用的共享协议。他人利用公开出版的数据时，应以规范形式明确引用数据论文或数据实体（如遵循国家标准《信息技术 科学数据引用》（GB/T 35294-2017））。在数据期刊出版数据时一般遵循开放共享协议（如CC0、CC-BY等）或数据期刊独立的数据政策，以明确数据在线获取和下载使用过程中应遵循的基本规则。

特征四：全程在线出版，与数据集一同即时发布。数据论文一般经由科学数据出版系统在线出版。发布的数据论文和数据集通过唯一标识符（如DOI）实现一致性关联，经同行专家评议保障数据的高质量与可读性。要求数据集或数据产品为原创成果，并且数据格式必须为常用或规范数据格

式，如果数据格式特殊，需要配备有开放的、可处理该数据的软件工具。

特征五：数据论文出版依托一体化数据期刊出版平台和科学数据存储库。数据论文的在线出版需要依托一体化数据期刊出版平台及数据存储库，其中数据期刊出版平台是数据论文的发布与共享平台。通过科学数据存储库，完成数据的提交和长期存储、在线展示等（详见图1）。

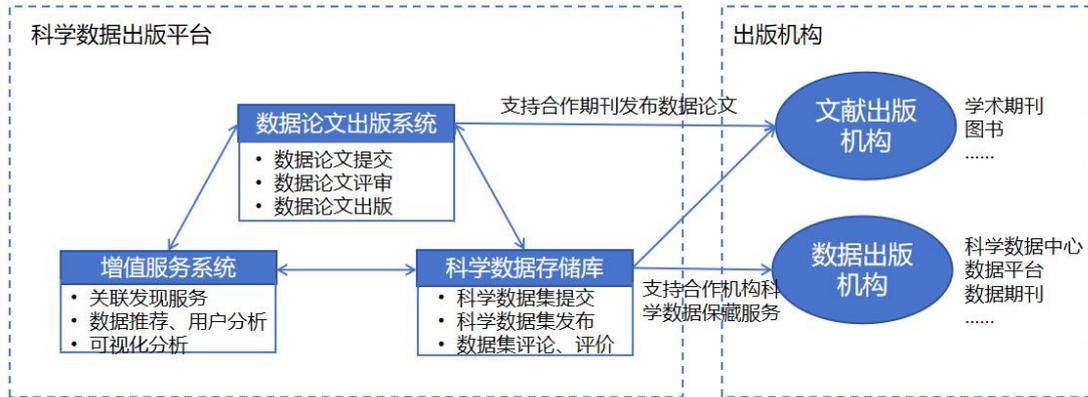


图1 数据论文出版平台架构图

Figure 1 Architecture of the data paper publishing platform

### 3.2 标准框架

基于数据论文出版的主要特征，大量的文献调研<sup>[10-13]</sup>，以及应用实践<sup>[14]</sup>，最终确定了标准的整体架构，详见图2。数据论文出版元数据由数据集元数据和数据论文元数据组成，分别描述数据集和数据论文。数据论文描述数据集内涵与特征，经过专家评审在数据期刊正式发布，其与数据集元数据具有较强一致性和关联性。

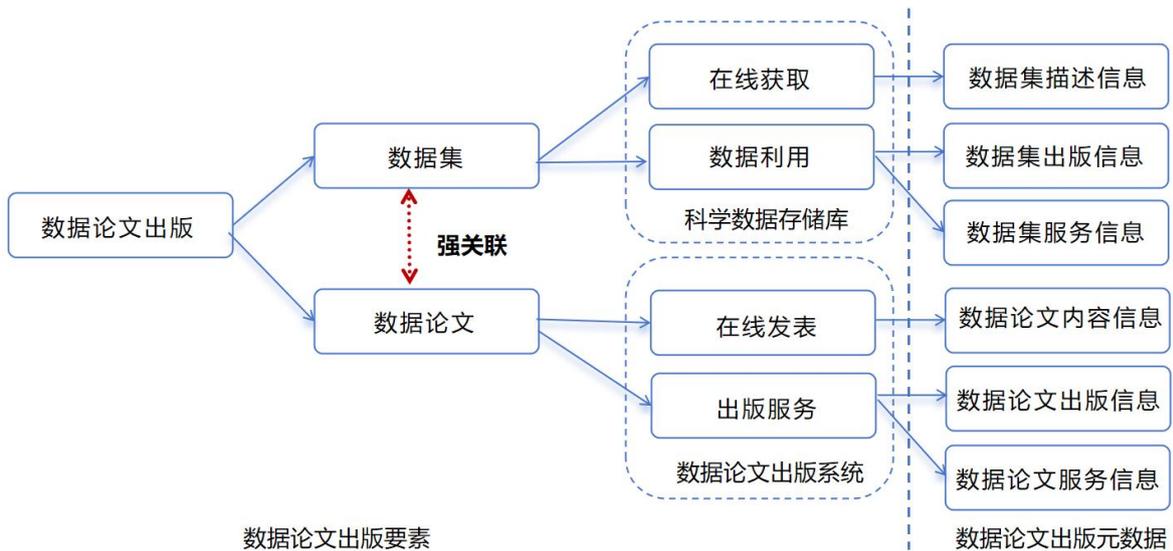


图2 国家标准《数据论文出版元数据》架构

Figure 2 National standard architecture for Data paper publication metadata

标准规定了数据论文出版元数据的内容与结构，适用于数据论文相关记录的描述、存储、在线发布和共享服务等，支持科学数据出版的规模化发展和高效应用。

《数据论文出版元数据》标准制定过程中，严格按照《GB/T 1.1-2020 标准化工作导则 第1部

分：标准的结构和编写》的规定进行编写，保证标准的编写质量。标准的主要技术内容符合现行有效国家标准和行业标准的相关规定（属于《GB/T 30522-2014 科技平台 元数据标准化基本原则与方法》的通用元数据，用于规范数据论文出版中资源的通用基本信息），并与现行相关法律法规、规章及相关标准协调一致。

《数据论文出版元数据》参考和引用的标准主要包括：《GB/T 7408 -2005 数据元和交换格式 信息交换 日期和时间表示法》《GB/T 7714-2015 信息与文献 参考文献著录规则》《GB/T 18391.1-2009 信息技术 元数据注册系统（MDR） 第1部分：框架》《GB/T 19710-2005 地理信息 元数据》《GB/T 30522-2014 科技平台 元数据标准化基本原则与方法》《GB/T 30523-2014 科技平台 资源核心元数据》《GB/T 32843-2016 科技资源标识》《GB/T 35294-2017 信息技术 科学数据引用》等。

标准中定义“数据论文是指对数据集进行规范化描述并按照学术规范在线联合出版的文章”；同时注明文章和数据集分属不同出版单元时，通过标识符呈现二者的唯一关联关系。定义“数据期刊是依法设立、公开出版数据论文及其关联数据集的刊物”，如出版传统学术论文的期刊同时出版数据论文及关联数据集，也认为是数据期刊。

标准采用GB/T 30523-2014规定的摘要表示的方式定义和描述数据论文出版元数据，包括：定义、英文名称、类型、值域、短名、注解等。

### 3.3 数据集出版元数据

数据集出版元数据包括数据集基本信息、数据集出版信息和数据集服务信息。

数据论文出版的核心是数据集，因此对数据集的描述既要与数据论文描述一致，又要能反映更多的数据描述信息，有助于用户检索和精准获取。数据集基本信息包括9个元数据元素和2个元数据实体：标识符、标题、摘要、关键词、范围（元数据实体）、语种、文件内容、基金项目、数据量、数据格式、数据集作者（元数据实体）。

数据集出版信息是数据集元数据的重要组成，主要定义数据集的发表日期、发表期刊和版本信息，可以全面呈现其出版过程关键信息。具体使用方式遵照数据出版机构的流程要求。数据集出版信息包括3个元数据元素：发布日期、出版期刊、版本信息。

数据集服务信息通过定义核心要素，为数据论文作者提供关联科学数据集服务数据。数据集服务信息包括5个元数据元素：数据集引用格式、数据集共享许可协议、数据集使用声明、数据集下载地址、数据论文访问地址。

### 3.4 数据论文出版元数据

数据论文出版元数据包括数据论文内容信息、数据论文出版信息和数据论文服务信息。

数据论文内容信息主要结合传统期刊论文结构化描述模式，对具有科学价值的某类或某个数据集进行规范化描述。数据论文内容信息包括12个元数据元素和1个元数据实体：标识符、标题、摘要、关键词、数据集基本信息、引言、数据采集和处理方法、数据样本描述、数据质量控制和评估、数据使用方法和建议、参考文献、致谢、数据论文作者（元数据实体）。数据期刊可根据实际需要定义扩展，同时，考虑数据论文作者可能会与数据集作者有差异，因此单独定义。

数据论文出版信息主要通过数据论文出版系统完成收稿日期、录用日期和出版日期等信息的记录，保留了数据论文出版继承传统出版系统优势的部分。同时也对数据论文的同评时间、版本信息进行了记录和发布。数据论文出版信息包括6个元数据元素：收稿日期、同评日期、录用日期、出

版日期、版本信息、出版期刊。

数据论文服务信息是在继承传统科技期刊要素基础上，结合数据论文的特点进行扩展和补充。数据论文服务信息包括4个元数据元素：数据论文引用格式、数据论文下载地址、数据论文共享许可协议、数据集访问地址。

具体每个元数据的属性特征，详见论文《数据论文及关联科学数据集出版元数据标准研究》<sup>[3]</sup>。此处不再赘述。

## 4 标准研制实践与应用推广

《数据论文出版元数据》国家标准早在2014年就作为项目内部标准，并在《中国科学数据（中英文网络版）》（简称《中国科学数据》，China Scientific Data, CN11-6035/N, ISSN 2096-2223）数据期刊平台进行试用，用于指导《中国科学数据》的数据论文模版、数据论文内容发布等方面。《中国科学数据》是目前中国唯一的面向多学科领域科学数据出版的学术期刊，作为国家网络连续型出版物的首批试点之一，由中国科学院主管，中国科学院计算机网络信息中心和ISC CODATA 中国全国委员会合办，国家科技基础条件平台中心、中国科学院网络安全和信息化领导小组办公室指导，是国内外公开发行的中英文季刊，也是中国科技核心期刊（CSTPCD）、中国科学引文数据库（CSCD）来源期刊。本标准规范了数据论文模版和出版描述信息等内容，将数据论文和数据集在网络平台上进行关联发布，并实现公开检索，指导《中国科学数据》出版平台的建设和使用。

同时，《数据论文出版元数据》国家标准在研制过程中，标准起草单位包括中国地质调查局发展研究中心、中国农业科学院农业信息研究所等多家期刊论文出版机构。其中，由中国地质调查局发展研究中心主办的地质学数据专刊——《地质科学数据专辑》，每期出版的数据论文及关联的实体数据，均在“地质科学数据出版中心”（[geodb.cgs.gov.cn](http://geodb.cgs.gov.cn)）与“地质科学数据出版系统”（[dcc.cgs.gov.cn](http://dcc.cgs.gov.cn)）出版。地质科学数据出版中心出版的“数据”包括元数据（中英文）、实体数据（中英文），以及通过《地质科学数据专辑》（中英文）发表的数据论文。针对数据论文出版相关描述信息，标准起草组多次交换意见。由中国农业科学院农业信息研究所主办的《农业大数据学报》（CN10-1555/G2, ISSN 2096-6369）是我国农业领域首个综合报道大数据领域相关的理论方法、技术应用、产业发展、实体数据等的专业学术期刊。其数据论文专栏分为论文部分和数据部分。针对数据论文专栏描述信息，标准起草组也进行过深度的研讨。以上这些讨论和交流，一方面为《数据论文出版元数据》国家标准的发布，提供重要参考意见；另一方面，也有助于推动期刊论文出版机构对数据论文进行规范化管理。

此外，《数据论文出版元数据》国家标准在征求意见阶段，还同时向51个国家科技资源共享服务平台（包括20个国家科学数据中心，31个国家生物种质和实验材料资源库）定向征求意见，并达成了实践上的共识。

2023年8月，《数据论文出版元数据》国家标准正式发布以后，国家政府网站与新华社同步报道<sup>[15]</sup>。中国科学院计算机网络信息中心作为该标准的第一完成单位，也积极推动标准的应用推广，在相关公众号和网站平台对其进行宣传<sup>[16]</sup>。标准起草人同步撰写学术论文《数据论文及关联科学数据集出版元数据标准研究》，通过文献调研、内容对比和归纳法分析国内外数据论文出版现状，结合科学数据出版工作实践，提炼数据论文及关联科学数据集出版要素，明确元数据规范对象、关联

关系和特点。值得说明的是,该篇论文作为《中国科技期刊研究》2023年10期封面论文发表<sup>[17]</sup>。后续,中国科学院计算机网络信息中心要继续推广《数据论文出版元数据》,为我国科学数据确权、产权保护提供支撑,推动科学数据开放共享生态建设。

## 5 结语

国内外数据出版的研究和实践表明,数据论文出版是信息技术发展到一定阶段的特有产物,具有传统学术出版的严谨性、专业性特点,同时又具有数据在线管理、实时服务和开发利用的技术需求,其标准化工作迫切,但目前我国相关标准还是空白。《数据论文出版元数据》国家标准的研制,在一定程度上解决了长期以来数据出版领域缺少标准规范指导的问题。

同时,数据论文出版作为创新数据服务模式,与传统出版物相比,对技术的要求和安全的要求更高。随着《数据出境安全评估办法(征求意见稿)》《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》《中华人民共和国网络安全法》等法律法规陆续发布,数据论文安全保藏、有序共享将面临软件技术升级、政策机制保障等现实挑战,同时开发机器可读的元数据编码方案,实现国家标准与国际标准的互操作,以及科学数据出版元数据与科学软件出版元数据的协调发展等方向,未来需要进一步研究。

## 参考文献

- [1] 苏靖. 大数据时代加强科学数据管理的思考与对策[J]. 中国软科学, 2022(9): 50–54. DOI: 10.3969/j.issn.1002-9753.2022.09.006. [SU J. Consideration and suggestion on strengthening scientific data management in the era of big data[J]. China Soft Science, 2022(9): 50–54. DOI: 10.3969/j.issn.1002-9753.2022.09.006.]
- [2] 张保钢. 国务院办公厅印发《科学数据管理办法》[J]. 北京测绘, 2018(5): 577. [ZHANG B G. Measures for the management of scientific data issued by the State Council of the People's Republic of China[J]. Beijing Surveying and Mapping, 2018(5): 577.]
- [3] 高瑜蔚, 朱艳华, 孔丽华, 等. 数据论文及关联科学数据集出版元数据标准研究[J]. 中国科技期刊研究, 2023, 34(10): 1270–1282. DOI: 10.11946/cjstp.202303050134. [GAO Y W, ZHU Y H, KONG L H, et al. Metadata standard for data paper and related scientific dataset publication[J]. Chinese Journal of Scientific and Technical Periodicals, 2023, 34(10): 1270–1282. DOI: 10.11946/cjstp.202303050134.]
- [4] CANDELA L, CASTELLI D, MANGHI P, et al. Data journals: a survey[J]. Journal of the Association for Information Science and Technology, 2015, 66(9): 1747–1762. DOI: 10.1002/asi.23358.
- [5] 邱春艳. 国内外科学数据出版理论研究述评[J]. 中国科技期刊研究, 2019, 30(3): 271–279. DOI: 10.11946/cjstp.201808020688. [QIU C Y. Review on scientific data publishing studies at home and abroad[J]. Chinese Journal of Scientific and Technical Periodicals, 2019, 30(3): 271–279. DOI: 10.11946/cjstp.201808020688.]
- [6] 国家标准《数据论文出版元数据》 [EB/OL]. [2023-06-05]. <https://std.samr.gov.cn/gb/search/gbDetailed?id=E116673EC9B3A3B7E05397BE0A0AC6BF>. [National standard *Metadata for Publication of Data Papers* [EB/OL]. [2023-06-05].

- <https://std.samr.gov.cn/gb/search/gbDetailed?id=E116673EC9B3A3B7E05397BE0A0AC6BF.>
- [7] 国家标准《数据论文出版元数据》立项启动会顺利召开 [EB/OL]. [2023-06-05]. [https://cnic.cas.cn/gzdt/202107/t20210706\\_6127431.html](https://cnic.cas.cn/gzdt/202107/t20210706_6127431.html). [National standard *Data paper publication metadata* project launch meeting successfully held [EB/OL]. [2023-06-05]. [https://cnic.cas.cn/gzdt/202107/t20210706\\_6127431.html](https://cnic.cas.cn/gzdt/202107/t20210706_6127431.html).]
- [8] 国家标准《数据论文出版元数据》专家视频研讨会顺利召开 [EB/OL]. [2023-06-05]. [http://www.cnic.cas.cn/kxyj/kyjz/202205/t20220531\\_6455980.html](http://www.cnic.cas.cn/kxyj/kyjz/202205/t20220531_6455980.html). [Expert video workshop on national standard *Data paper publication metadata* successfully held [EB/OL]. [2023-06-05]. [http://www.cnic.cas.cn/kxyj/kyjz/202205/t20220531\\_6455980.html](http://www.cnic.cas.cn/kxyj/kyjz/202205/t20220531_6455980.html).]
- [9] 国家标准《数据论文出版元数据》专家研讨会顺利召开 [EB/OL]. [2023-06-05]. [http://cnic.cas.cn/gzdt/202209/t20220929\\_6518621.html](http://cnic.cas.cn/gzdt/202209/t20220929_6518621.html). [Expert workshop on national standard *Data paper publication metadata* successfully held [EB/OL]. [2023-06-05]. [http://cnic.cas.cn/gzdt/202209/t20220929\\_6518621.html](http://cnic.cas.cn/gzdt/202209/t20220929_6518621.html).]
- [10] CHAVAN V, PENEV L. The data paper: a mechanism to incentivize data publishing in biodiversity science[J]. *BMC Bioinformatics*, 2011, 12(Suppl 15): S2. DOI: 10.1186/1471-2105-12-S15-S2.
- [11] 陈帅印, 刘桂锋, 刘琼. 基于元数据的数据论文与期刊论文的关联研究: 以全球变化科学研究数据出版系统为例[J]. *数字图书馆论坛*, 2022(8): 11–18. DOI: 10.3772/j.issn.1673-2286.2022.08.002. [CHEN S Y, LIU G F, LIU Q. A study of the association of metadata-based data papers with journal papers: take the global change science research data publishing system as an example[J]. *Digital Library Forum*, 2022(8): 11–18. DOI: 10.3772/j.issn.1673-2286.2022.08.002.]
- [12] 孔丽华, 邵明玥. 科学数据出版内容与案例分析[J]. *科研信息化技术与应用*, 2018, 9(6): 39–46. [KONG L H, SHAO M Y. Scientific data publishing: a case study of the contents and trends[J]. *e-Science technology & application*, 2018, 9(6): 39–46.]
- [13] 熊泽润, 宋立荣. 科学数据出版中同行评议的问题思考[J]. *中国科技资源导刊*, 2022, 54(5): 21–29. DOI: 10.3772/j.issn.1674-1544.2022.05.003. [XIONG Z R, SONG L R. Thinking about peer review in scientific data publishing[J]. *China Science & Technology Resources Review*, 2022, 54(5): 21–29. DOI: 10.3772/j.issn.1674-1544.2022.05.003.]
- [14] 胡良霖, 黎建辉, 高瑜蔚, 等. 科学数据出版应用实践[M]// 中国科研信息化蓝皮书 2015. 科学出版社, 2015: 307-322. [HU L L, LAI J H, GAO Y W, et al. Scientific data publishing application practices[M]// *China Scientific Research Informatization Blue Book 2015*. Science Press, 2015: 307-322.]
- [15] 支持科学数据开放共享新模式 《数据论文出版元数据》国家标准正式发布[EB/OL]. [2023-11-01]. [https://www.gov.cn/lianbo/bumen/202309/content\\_6907179.htm](https://www.gov.cn/lianbo/bumen/202309/content_6907179.htm). [Supporting a new model of open sharing of scientific data: the national standard *Data paper publication metadata* is officially released [EB/OL]. [2023-11-01]. [https://www.gov.cn/lianbo/bumen/202309/content\\_6907179.htm](https://www.gov.cn/lianbo/bumen/202309/content_6907179.htm).]
- [16] 国家政府官网与新华社同步报道国家标准《数据论文出版元数据》成果发布 [EB/OL]. [2023-12-01]. [http://cnic.cas.cn/gzdt/202312/t20231201\\_6904218.html](http://cnic.cas.cn/gzdt/202312/t20231201_6904218.html). [The official website of the national government and Xinhua News Agency synchronized report on the release of the results of the national standard *Data paper publication metadata* [EB/OL]. [2023-12-01]. [http://cnic.cas.cn/gzdt/202312/t20231201\\_6904218.html](http://cnic.cas.cn/gzdt/202312/t20231201_6904218.html).]

[17] 10期封面论文:《数据论文及关联科学数据集出版元数据标准研究》[EB/OL]. [2023-12-01]. [https://mp.weixin.qq.com/s/SSLM\\_jS2juOERsK8rDrhCQ](https://mp.weixin.qq.com/s/SSLM_jS2juOERsK8rDrhCQ). [Issue 10 cover paper: A study of publication metadata standards for data papers and linked scientific datasets [EB/OL]. [2023-12-01]. [https://mp.weixin.qq.com/s/SSLM\\_jS2juOERsK8rDrhCQ](https://mp.weixin.qq.com/s/SSLM_jS2juOERsK8rDrhCQ).]

## 作者分工职责

高瑜蔚, 博士, 工程师, 主要从事科学大数据技术、标准政策和智能服务研究工作。在本文中承担论文撰写、统稿修改工作。

朱艳华, 高级工程师, 主要从事科学大数据管理与共享、标准制订和数据应用服务研究工作。本文中负责素材分析、提出修改意见。

胡良霖, 教授级高级工程师, 主要从事科学大数据管理、开放共享、应用服务等方面研究工作。本文中负责确定文章总体思路, 提出修改意见。

孔丽华, 博士, 高级工程师, 研究方向为科学数据管理政策、数据开放共享政策与实践, 以及数据出版等方面的研究。本文中负责对文章提出思路建议, 对文章内容提出修改意见。

赵欢, 工程师, 主要从事科学大数据管理与共享、数据共享服务平台研究工作。本文中负责文献调研和文章修改。

## 论文引用格式

高瑜蔚, 朱艳华, 胡良霖, 等. 《数据论文出版元数据》国家标准研制与实践[J/OL]. 中国科学数据, 2024, 9(1). (2024-02-22). DOI: 10.11922/11-6035.csd.2023.0187.zh.

## Development and practices of the national standard for *Data paper publication metadata*

GAO Yuwei<sup>1,2,3</sup>, ZHU Yanhua<sup>1,3</sup>, HU Lianglin<sup>1,3\*</sup>, KONG Lihua<sup>1</sup>, ZHAO Huan<sup>1,3</sup>

1. Computer Network Information Center (CNIC), , Chinese Academy of Sciences, Beijing 100083, P.R. China

2. Research Center for Language Intelligence of China, Capital Normal University, Beijing 100048, P.R. China

3. The National Basic Science Data Center (NBSDC), Beijing 100083, P.R. China

\*Email: hull@cnic.cn (HU Lianglin)

**Abstract:** In August 2023, the national standard for *GB/T 42813-2023 Data paper publication metadata* was officially released. This standard specifies the content and framework of metadata required for data paper publication, including dataset metadata, data paper metadata, as well as their description methods, extension principles, and methods. It is applicable to the description, preservation, online publishing, and sharing services associated with data papers and their corresponding datasets. The release of this standard marks the first comprehensive agreement at the national level regarding the new mode of data paper

publishing, namely the open sharing of scientific data, which has addressed the prior absence of standardized guidance in the field of data publishing. It facilitates data publishing institutions in promptly releasing data papers and their associated data, providing support for the rights and protection of scientific data. The standard holds significant importance in promoting the establishment of an open and shared ecosystem for scientific data. Since its publication, certain data journals or scientific data publishing platforms have begun adopting the elements and definitions specified in the standard, thus engaging in a more extensive practice.

**Keywords:** open science; data paper; scientific data publication; metadata; standard; scientific dataset