

中华人民共和国国家标准

GB/T 7713.4—2025

信息与文献 编写规则 第4部分:数据论文

Information and documentation—Presentation of documentation— Part 4: Data papers

2025-08-29 发布 2026-03-01 实施

目 次

			······
引	言		IV
1	范	這围	
2	规	l范性引用文	件
3	术	语和定义、缩	i略语·······
	3.1	术语和定义	χ 2
	3.2	缩略语 …	
4	组	[成部分	
	4.1	一般要求	
	4.2	前置部分	
	4.3	正文部分	
	4.4	后置部分	
	4.5	附属部分	
5	编	排格式	<u>571</u> G
附	·录	A (规范性)	数据论文的构成要素
附	录	B(资料性)	数据论文作者学术贡献类型示例 10
参	考]	文献	



前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分:标准化文件的结构和起草规则》的规定起草。

本文件是 GB/T 7713 的第 4 部分。GB/T 7713 已经发布了以下部分:

- ——信息与文献 编写规则 第1部分:学位论文;
- ——学术论文编写规则;
- ——科技报告编制规则;
- ——信息与文献 编写规则 第 4 部分:数据论文。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国信息与文献标准化技术委员会(SAC/TC 4)提出并归口。

本文件起草单位:中国科学院文献情报中心、中国农业科学院农业信息研究所、中国科学院计算机网络信息中心、中国国家图书馆、中国科学技术信息研究所、南京大学、中国地质调查局、北京大学、中国科学技术期刊编辑学会、北京卓众出版有限公司、中国高校科技期刊研究会、《中国科学》杂志社有限责任公司、《中华医学杂志》社有限责任公司、中国医学科学院医学信息研究所、中国科学院地理科学与资源研究所、中国农业大学、北方民族大学、福建农林大学、北京万方数据股份有限公司、《中国学术期刊(光盘版)》电子杂志社有限公司。



本文件主要起草人:马峥、王健、孔丽华、刘筱敏、张品纯、顾犇、潘云涛、沈玉兰、李勃慧、张贵兰、杨海平、涂骏、杜建、陈莉玥、谢艳丽、魏均民、张铁明、张昕、任胜利、侯丽、王卷乐、高秉博、刘建平、黄晓磊、梁冰、潘薇、齐萍、伍军红、赵巍、刘荣梅、刘婷婷。

引 言

数据论文是以数据可复现性论证和可复用性说明为主要内容,全面、规范、深入描述特定研究数据集的新兴学术论文体裁,其作用是促成数据的规范表达、准确理解、价值发现、重复利用和广泛传播。数据论文特定的描述对象、传播特点和传播目的,决定了其需要不同于知识类研究论文的撰写和编排规范。

GB/T 7713—1987《科学技术报告、学位论文和学术论文的编写格式》对科学技术报告、学位论文和学术论文的撰写要求及编排格式作了统一规定。鉴于各类型文献之间撰写要求及编排格式的特异性逐渐加大,在其后的修订工作中,将其分成多个部分分别修订。GB/T 7713 拟由 4 个部分组成。

- ——第1部分:学位论文。目的在于规定学位论文的撰写格式和要求。
- ——第2部分:学术论文。目的在于规定学术论文的撰写格式和编排要求。
- ——第3部分:科技报告。目的在于规定科技报告的编写、组织、编排要求。
- ——第4部分:数据论文。目的在于规定数据论文的组成部分、撰写格式和编排要求。

鉴于数据论文已成为开放科学背景下日趋重要的全新文献类型,本文件的目的在于规定数据论文的组成部分、撰写格式和编排要求,主要包括数据论文的内容元素和格式规范,涵盖但不限于表达内容完整性、结构组织清晰性、格式体例规范性、语言文字和符号规范性,以及技术和计量单位等方面的标准描述。

本文件发挥的作用将便于数据论文的撰写、组织、检索、传播、引用及数据的重复利用,更好地实现研究数据共享与应用,推动开放科学发展。

本文件的适用范围考虑了各学科领域的特殊性,不同学科领域研究在研究设计、数据方法、数据质量控制、数据内容、数据共享、数据安全与伦理保护等方面的差异,采用本文件编制数据论文时宜兼顾严谨性和灵活性。

信息与文献 编写规则 第4部分:数据论文

1 范围

本文件规定了数据论文的组成部分和编排格式。本文件适用于印刷版、缩微版、电子版等所有传播形式的数据论文。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

- GB/T 3101 有关量、单位和符号的一般规则
- GB/T 3102.1 空间和时间的量和单位
- GB/T 3102.2 周期及其有关现象的量和单位
- GB/T 3102.3 力学的量和单位
- GB/T 3102.4 热学的量和单位
- GB/T 3102.5 电学和磁学的量和单位
- GB/T 3102.6 光及有关电磁辐射的量和单位
- GB/T 3102.7 声学的量和单位
- GB/T 3102.8 物理化学和分子物理学的量和单位
- GB/T 3102.9 原子物理学和核物理学的量和单位
- GB/T 3102.10 核反应和电离辐射的量和单位
- GB/T 3102.11 物理科学和技术中使用的数学符号
- GB/T 3102.12 特征数
- GB/T 3102.13 固体物理学的量和单位
- GB/T 6447 文摘摘要编写规则
- GB/T 7408.1 日期和时间 信息交换表示法 第1部分:基本原则
- GB/T 7713.2 学术论文编写规则
- GB/T 7714 信息与文献 参考文献著录规则
- GB/T 13745 学科分类与代码
- GB/T 14559 变化量的符号和单位
- GB/T 19710.1 地理信息 元数据 第1部分:基础
- GB/T 32843 科技资源标识
- GB/T 35294 信息技术 科学数据引用
- GB/T 42813 数据论文出版元数据
- CY/T 173 学术出版规范 关键词编写规则
- ISO 26324 信息与文献 数字对象标识系统(Information and documentation—Digital object identifier system)
 - 注: GB/T 36369-2018 信息与文献 数字对象唯一标识符系统(ISO 26324;2012,IDT)

3 术语和定义、缩略语

3.1 术语和定义

下列术语和定义适用于本文件。

3.1.1

科学数据 scientific data; research data

遵循学术共同体共识及必要规范,对目标对象或现象进行测量或记录所形成结果的符号化记录,包括但不限于数值、图文、音视频等形态。

3.1.2

数据论文 data paper

对一个或多个具有逻辑关系的数据集(3.1.4)进行规范化描述,并按照学术规范出版的文章。 「来源;GB/T 42813—2023,3.2,有修改]

3.1.3

数据文件 data file

物理上独立的电子文件。

注:每个数据文件均具有名称和特定的技术格式。一个或多个存在联系的数据文件组成一个数据集。

3.1.4

数据集 dataset

一个或多个数据文件组成的反映同一主题的数据文件集合。

注:数据库是数据集的形式之一。

[来源:GB/T 31075—2014, A.2.16, 有修改]

3.1.5

数据对象 data object

数据描述的个体、群体或现象。

3.2 缩略语

下列缩略语适用于本文件。

CCO:作者放弃作品所有权利的一种共享协议(Creative Commons Zero)

CRediT:贡献者角色分类法(Contributor Roles Taxonomy)

CSTR:科技资源标识(Common Science and Technology Resource Identification)

DOI:数字对象唯一标识符(Digital Object Unique Identifier)

ISO:国际标准化组织(International Organization for Standardization)

xls:Excel 电子表格文件格式(Excel Spreadsheet)

4 组成部分

4.1 一般要求

论文宜包括以下 4 个组成部分:

- a) 前置部分概述数据论文的核心属性,前置部分各个项目宜同时使用内容对应的中文和外文(多用英文)表达;
- b) 正文部分描述数据集的主体内容;

- c) 后置部分描述论文与其他资源的关联:
- d) 附属部分描述论文附加信息。

数据论文各组成部分及相关的元数据信息应按照附录 A 进行。根据期刊制作印刷版或电子版呈现内容的实际需要,可调整各要素内容的编排位置和前后顺序。

4.2 前置部分

4.2.1 题名(title)

题名是对数据论文描述数据集的准确概括,是数据集内容、时空特征、技术特征、学术特征等的逻辑组合。

题名应简明。题名中的词语需有助于选定关键词和编制题录、索引等二次文献所需的使用信息,应使用标准术语、学名全称、药物和化学品通用名称,不应使用广义术语、夸张词语等。

下列情况可使用副题名:题名语义未尽,用副题名补充说明论文中的特定内容;研究成果分部分或 分阶段报道,可用不同副题名加以区别;其他有必要是使用副题名作为引申或说明的情况。

题名在数据论文中不同地方出现时应保持一致。

本项应为必备项。

4.2.2 作者信息(author)

数据论文作者署名方式遵守《中华人民共和国著作权法》的要求。

数据论文署名为个人作者的情况,应标注作者的姓名、机构、学术头衔等个人信息,宜标注作者的唯一标识、联系方式等信息。单一作者应标注联系信息;多作者应指定通信作者并标注联系信息。多作者的排序应依据个人对数据论文的实质性贡献,从大到小依次排列。

数据论文署名为团体作者的情况,应使用团体的标准规范全称,不应单独使用简称或缩写形式。应标注通信作者的联系信息。宜列出全部作者或主要作者信息。

本项应为必备项。

4.2.3 作者贡献(contribution)

数据论文应明确描述每个作者的贡献,同一作者可有多个类型的贡献。具体贡献类型可由期刊出版单位自行制定,宜参见附录B示例的作者贡献类型结构,或参考CRediT等类型结构。

可逐一描述每一位作者参与的工作和贡献,也可逐一描述每一项工作的参与作者。本项应为必备项。

4.2.4 摘要(abstract)

摘要是对数据论文内容不加注释和评论的简短陈述,应具有独立性和自明性。摘要宜采用报道性 结构式摘要。

摘要的撰写形式应符合 GB/T 6447 的规定。

本项应为必备项。

4.2.5 关键词(keyword)

关键词宜精要概括数据集的主题、内容、数据对象、数据方法、数据特点或数据价值等。关键词的设置应以有助读者推断数据内容和提升数据集检出率为原则。

GB/T 7713.4-2025

关键词的撰写按照 CY/T 173 相关要求与规定。 本项应为必备项。

4.2.6 数据概括(data summary)

4.2.6.1 形式和结构

数据概括是数据集基本信息的汇总,反映当前数据集最重要或读者最关注的描述信息。数据概括 应以表格或其他结构化形式呈现。

数据概括的结构包括内容属性、技术属性、关联属性,各属性下设对应的相关要素。其中部分要素 应为必备项、部分要素为可选项或有则必备项。数据概括依据 GB/T 42813 提供数据论文出版元数据。

4.2.6.2 内容属性

数据集的内容属性包括描述数据集内容特征的各相关要素。

- a) 数据集名称(dataset title) 数据集的名称,官含有数据对象,应与对应的数据存储名称保持一致。本项应为必备项。
- b) 数据集版本(dataset version) 数据集的版本编号。本项应为必备项。
- c) 数据对象(data object) 科学数据描述的实体对象(个体或群体)或现象。本项应为必备项。
- d) 数据内容(data content) 按照 4.3.3 概述数据集的主要内容,可采用文本描述,也可列出数据元或关键测量指标。本项 应为必备项。
- e) 学科代码(subject code) 数据集所属研究的学科。按照 GB/T 13745 的规则或参考《中国图书馆分类法》第三级学科表示学科和对应代码。本项应为必备项。
- f) 主题(research area) 数据集的主题是数据集所属研究的更详细划分,大部分情况下指向更详细的子学科、子领域或 科学问题,部分情况下可指向大型的科学行动。主题词宜使用叙词,也可使用自由词。本项为 可选项。
- 时间范围(temporal range) 数据对象被测量的时间范围,是数据采集开始时点直至完成全部采集活动的时段。应按照 GB/T 7408.1的规则表示时间范围。本项为可选项。
- h) 空间范围(spatial range) 数据对象被测量时的空间范围(如地理位置与范围、海拔高度等)。应按照 GB/T 19710.1 的 规则表示空间范围。本项为可选项。
- i) 主要物理量的值(physical quantities value) 数据对象被测量时主要物理量(如温度、压强、密度、速度、电流等)的数值或数值范围。物理量的选择依据学科特点和需求。按照 GB/T 3101、GB/T 3102.1、GB/T 3102.2、GB/T 3102.3、GB/T 3102.4、GB/T 3102.5、GB/T 3102.6、GB/T 3102.7、GB/T 3102.8、GB/T 3102.9、GB/T 3102.10、GB/T 3102.11、GB/T 3102.12、GB/T 3102.13、GB/T 14559的规则表示主要物理量。本项为可选项。

4.2.6.3 技术属性

数据集的技术属性包括描述数据集技术特征的各相关要素。

a) 数据方法(data method)

按照 4.3.2 概要描述数据采集和处理方案设计、执行过程。本项应为必备项。

b) 数据的技术格式(data format)

按照 4.3.2 分类型列出当前数据集所属数据文件的技术格式,包括格式名称和对应的数据文件后缀名。本项应为必备项。

c) 数据集规模(dataset volume)

按照 4.3.2 给出数据集字节数值的总和。本项为可选项。

d) 数据集结构(dataset structure)

按照 4.3.2 给出简要描述数据集的主体结构。本项为可选项。

e) 数据处理工具和代码可用性(tool and code availability)

按照 4.3.8 列出数据处理或分析工具名称和获取信息、数据处理代码的获取信息。本项为可选项,有则必备。

4.2.6.4 关联属性

数据集的关联属性包括描述数据集与其他资源关联关系的各相关要素。

a) 数据存储与访问地址(dataset accessibility)

按照 4.3.7 列出数据集(及其所属数据文件)所在的数据仓储库名称,提供至少一个长期有效的互联网访问地址。本项应为必备项。

b) 数据集的唯一标识(dataset unique identifier)

按照 4.3.7 给出数据集的唯一标识。长期有效互联网访问地址可看作是唯一标识。本项应为必备项。

c) 数据集共享协议(dataset sharing protocol)

数据集采用标准化共享协议的情况下应列出协议名称和公开利用的时间。数据集采用非标准化共享协议的情况下应概述主要协议条款。本项应为必备项。

d) 数据集相关的研究成果(research products)

按照 4.4.4 列出数据集相关研究成果信息,宜列出研究成果的唯一标识。本项为可选项,有则必备。

e) 相关的数据集或历史版本数据集(related dataset or historical versions of the dataset) 按照 4.4.5 列出相关数据集或历史版本数据集的唯一标识。本项为可选项,有则必备。

4.2.7 其他部分

论文前置部分要求、建议或允许标注的其他项目。

a) 基金项目(funding)

基金资助项目产出的数据论文,应标注基金来源、项目名称及编号。本项为可选项,有则必备。

b) 投稿日期(date of submission)

宜标注投稿日期。可同时标注修回日期(date of revised)。本项为可选项。

c) 引用格式(cites)

可标注引用本论文的参考文献格式。本项为可选项。

d) 其他

本项为可选项。

4.3 正文部分

4.3.1 引言

引言用以指明数据产生的背景、数据的重要性或其对所属项目、主题和学科领的意义,用以说明数据生产和传播的必要性。

本项应为必备项。

4.3.2 数据方法与技术格式

数据方法是数据采集、处理方法及相关设计、材料及技术的统称。在涉及材料制备的研究可命名为"材料与方法"或更准确的名称。

对数据方法的描述是数据可复现性判断的主要依据,是数据论文的核心内容。数据方法宜完整、准确、清晰、规范地说明数据生产的全部方法与完整过程。

数据方法包括但不限于数据采集方案(或者,试验设计,包括主要采集设备、采集标准、流程或规程)、采集方案的具体实施、原始数据的筛选/清洗/合并/插值等处理操作、数据质量控制方案及其实施效果、数据文件的形成以及集结为数据集的规则与过程等。

对数据方法的描述详尽程度官以读者可重复(或理论上可重复)该过程并复现数据集为标准。

数据的技术格式是数据编码和组织的格式。不同格式对应不同的数据处理算法、软件或工具(以下简称"数据处理工具")。 宜选择通用且具备互操作性的技术格式,不宜选择依赖专有数据处理工具的技术格式。宜分类型列出数据文件的技术格式,包括格式名称和对应的数据文件后缀名(如:*.xls)。

本项应为必备项。

4.3.3 数据内容

数据内容应说明数据集的范畴、结构、规模和特征,可添加数据样例。

数据范畴宜全面、详细、准确且有逻辑。

数据集的结构是数据集包含的数据文件及其相互关系。

数据集的规模是数据集的字节数值,是所有数据文件字节数值的总和。

数据特征说明官采用规范术语或共识词汇,且需与数据集名称用词保持一致。

本项应为必备项。

4.3.4 数据质量控制与评估

数据质量控制与评估包括数据生产与加工处理过程中的质量控制方案(包括但不限于质量控制方法、策略、原则以及所依据的标准和使用的技术等)、方案的执行以及最终数据质量的评价结果。数据质量控制与评估宜采用量化方法与结果,部分无法量化控制与评估的数据可采用定性方法。

数据质量控制与评估是数据可复现性和数据质量判断的重要依据,其描述应准确、全面与规范,应以可有效证明数据的可靠性为标准。

本项应为必备项。

4.3.5 数据使用

在具备条件的情况下宜指明当前数据集的使用情况,以及扩展说明更广泛的使用场景。数据使用 说明可为数据的重复利用给出示例性信息。

本项为可选项。

4.3.6 数据价值

数据价值宜通过同类数据对比的方式,从数据的学术意义、潜在用途、稀缺性和难以复现等方面予以说明。数据价值可指数据集的学术价值,也可指社会经济等衍生价值。

本项为可选项。

4.3.7 数据可用性

数据可用性信息包括数据集所属仓储库名称、数据集的唯一标识、数据集的互联网访问地址等。数据集的唯一标识可单独或同时按照 GB/T 32843 规则采用 CSTR,按照 ISO 26324 规则采用 DOI,或采用其他具有明确规范、可解析的标识。当数据集的唯一标识无法映射至互联网访问地址时,应明确给出有效的互联网访问地址。

本项应为必备项。

4.3.8 数据处理工具或代码可用性

对于数据集整体或其中部分数据文件,如使用通用格式,可不说明数据处理工具;如使用非通用格式需要专门数据处理工具或代码,应给出数据处理工具的名称和获取方式、代码的获取方式。

宜给出获取和使用数据处理工具或代码的必要信息,包括但不限于工具的构成、下载、安装、运行或使用等方面信息。以读者能够使用该工具处理与分析数据为准则。

本项为可选项,有则必备。

4.3.9 同行评议信息

可采用匿名或非匿名的形式给出数据论文的同行评审过程、评审意见与修回情况。当存在独立的数据集评审时,官给出相关评审信息。

本项为可选项。

4.4 后置部分

4.4.1 共享约定

数据论文的共享约定规定了作者向使用者和传播者让渡的权益、方式与条件,宜综合考虑隐私保护、学术优先权等因素。

宜优先采用标准化的数据共享协议格式(如 CCO 协议),同时应给出协议名称表示并附加协议源链接,也可自定义数据共享协议,同时应给出自定义协议全文的获取方式,并应概括列出对数据传播、复制、编辑、使用的要求。

本项应为必备项。

4.4.2 伦理与利益冲突声明

当数据集涉及人类参与者、动物福利、生物多样性等研究伦理事项时,应声明数据集的伦理符合程度和取得必要伦理授权的情况。应给出有效的伦理审查批号(经过伦理审查委员会批准、符合研究道德规范的论文所获得的编号)。

应说明作者及其所在机构与该数据集、数据论文相关研发应用活动是否存在利益关联和冲突。 本项应为必备项。

4.4.3 参考文献

按照 GB/T 7714 和 GB/T 35294 的规则编写。各项参考文献的唯一标识可单独或同时按照



GB/T 7713.4—2025

GB/T 32843 的规则采用 CSTR,按照 ISO 26324 的规则采用 DOI,或采用其他具有明确规范、可解析的标识。

本项为可选项,有则必备。

4.4.4 数据相关研究成果

基于数据集开展研究工作而形成的各类学术成果。研究成果形式包括但不限于期刊论文、会议论文、专利、预印本论文、学位论文、著作、研究报告等。

按照 GB/T 7714 的规则格式编写。研究成果的唯一标识可单独或同时按照 GB/T 32843 规则采用 CSTR,按照 ISO 26324 规则采用 DOI,或采用其他具有明确规范、可解析的标识。

本项为可选项,有则必备。

4.4.5 相关数据集或历史版本数据集

相关数据集或历史版本数据集指在内容或工作等方面相关的其他数据集或同一数据集的历史版本。标注内容应包括相关数据集或历史版本数据集的名称、版本、发布时间及数据集唯一标识与访问地址。数据集唯一标识可单独或同时按照 GB/T 32843 规则采用 CSTR,按照 ISO 26324 规则采用 DOI,或采用其他具有明确规范、可解析的标识。

本项为可选项,有则必备。

4.5 附属部分

4.5.1 致谢

致谢是对在数据集及数据论文的生成过程(包括从数据采集到数据集生成以及论文撰写全部过程)中,作者以外提供协助的组织或个人予以感谢的文字记录。致谢内容应客观、真实地表述被致谢者及具体协助内容。致谢范围和方式按照 GB/T 7713.2 的规则编写。

本项为可选项,有则必备。

4.5.2 附录

按照 GB/T 7713.2 的规则编写。 本项为可选项。

4.5.3 其他信息

作者或出版者认为重要或需要公开的其他相关信息或声明。 本项为可选项。

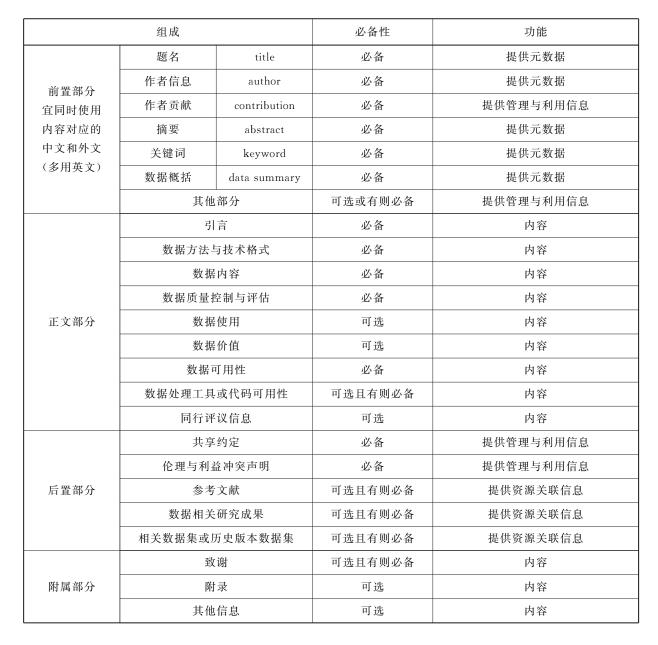
5 编排格式

按照 GB/T 7713.2 的规则编写。

附 录 A (规范性) 数据论文的构成要素

表 A.1 规定了数据论文的构成元素。

表 A.1 数据论文的构成要素



附录 B (资料性)

数据论文作者学术贡献类型示例

表 B.1 给出了数据论文作者学术贡献类型示例。

表 B.1 数据论文作者学术贡献类型示例

编号	贡献类型	定义
1	方案设计(methodology)	提出创意和想法。制定或完善总体研究目标,开发或设计数据相关的方案、方法设计
2	项目管理(project administration)	管理、协调数据相关的研究活动,包括制定计划与具体执行。监管和领导数据相关研究活动的计划和执行,特别是数据共享策略与计划的制定
3	数据管理 (data curation)	数据集生产与管理,包括数据注释(产生元数据)、数据清理、数据(包括数据所需的软件)管理(含持久存档、共享和后续利用等)等
4	资源准备 (resources)	提供或准备数据相关的研究资源,包括但不限于研究材料、试剂、患者、实验室样本、动物、仪器、计算资源或其他分析工具;与数据相关的软件开发、测试、部署与运行
5	数据获取 (investigation)	开展数据相关的研究和调查活动,主要是进行试验或数据/证据的收集
6	数据验证 (validation)	数据验证、数据质量控制和其他与数据再现性有关的事项
7	数据分析 (formal analysis)	运用统计、数学、计算等方法或其他技术手段对研究数据 进行集成与描述性分析
8	成果可视化(visualization)	可视化设计、创作和制作,用于实现数据成果及分析过程 与结果的展示和描述等
9	论文撰写 (writing and reviewing)	数据论文初稿撰写、评审或重要修改

参 考 文 献

- [1] GB/T 31075-2014 科技平台 通用术语
- [2] 中华人民共和国著作权法(中华人民共和国主席令第62号)
- [3] CRediT [EB/OL].[2019-12-01].https://casrai.org/credit/
- [4] 中国图书馆分类法

5/10



