



图书馆杂志
Library Journal
ISSN 1000-4254, CN 31-1108/G2

《图书馆杂志》网络首发论文

题目： 我国《科学数据管理办法》中概念视角数据域范畴与管理边界研究
作者： 都平平，李雨珂，张雪媛
网络首发日期： 2022-02-28
引用格式： 都平平，李雨珂，张雪媛. 我国《科学数据管理办法》中概念视角数据域范畴与管理边界研究[J/OL]. 图书馆杂志.
<https://kns.cnki.net/kcms/detail/31.1108.G2.20220225.0930.002.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

我国《科学数据管理办法》中概念视角数据域范畴与管理边界研究*

都平平¹ 李雨珂² 张雪媛¹ (1 中国矿业大学图书馆 2 上海出版印刷高等专科学校文化管理系)

摘要 从概念域视角研究数据范畴及数据域边界,明确其多维度的类型和属性划分,为分类规范数据管理,明确不同属性数据保障策略,具体实施《科学数据管理办法》服务。①从相关文献中析出若干关于数据和科学数据的相关概念、定义表述,按照其产生过程、发布形态、功能作用、学科属性、责任与权益归属等特征进行了梳理和解读,形成数据域概念视角的数据类型和数据集属性边界。②对照“办法”中的相关条文,具体提出不同类别不同属性的数据域边界及管理细则,为《办法》的全面实施提供细化策略。③尤其是自然科学和社会科学数据的边界界定,有利于科学数据的登记、汇集、管理、实施,也有利用不同属性数据域权限归属的合理复用、共享。

关键词 科学数据 概念域 边界 数据域管理

Research on Conceptual Domain and Management Boundary of Data Set in China's Measures for Scientific Data Management

Du Pingping¹ Li Yuke² Zhang Xueyuan¹ (1 Library of China University of Mining and Technology; 2 Department of cultural management, Shanghai Publishing and Printing College)

Abstract Study the data category and data domain boundary from the perspective of concept domain, clarify its multi-dimensional types and attribute division, and serve for classifying and standardizing data management, clarifying the data guarantee strategies of different attributes, and specifically implementing the measures for scientific data management. ① Some related concepts, definitions and expressions about data and scientific data are extracted from relevant literature, sorted and interpreted according to their production process, release form, function, discipline attribute, responsibility and equity ownership, so as to form the data type and dataset attribute boundary from the perspective of data domain concept. ② According to the relevant provisions in the "measures", specific data domain boundaries and management rules of different categories and attributes are put forward to provide detailed strategies for the full implementation of the "measures".③ In particular, the boundary definition of natural science data and social science data is conducive to the collection, management and implementation of scientific research data, as well as the rational reuse and sharing of the authority ownership of data fields with different attributes.

Keywords Scientific Data (Research Data), Conceptual Domain, Boundary, Data domain management

0 引言

《科学数据管理办法》(简称《办法》)的出台,为其数据域管理提供了最权威的国家层面纲领性的文件,昭示了学科数据从内涵、范围、功效、实施的科学化管理水平。对于指导

*本文系国家社会科学基金项目“机构知识资源领域研究数据的关联组织和开放复用研究”(项目编号:19BTQ073)研究成果之一。

通信作者:李雨珂, E-mail:119448932@qq.com

我国各级各类科学数据产生、管理、传播有着重要指导作用，但是《办法》是一个宏观文件，对于目前复杂的数据现象和快速发展的数据行业现状，不可能面面俱到。由于数据的类型和范畴属性相当复杂，且数据科学领域发展变化太快，落实科学数据管理问题，关键在于根据数据类型、数据属性及数据边界，依据它们的特征，制定适合区域、领域、行业针对性管理措施和具体策略。本文基于概念视角从数据域的角度讨论“数据”与“数据集”的不同属性与类型，并根据不同属性的“数据集”里不同类型“数据”试拟相应的管理策略要旨。

1 我国科学数据概念的发展基础

数据域是数据的边界和范围，即对数据进行分类及分类的标准。而“数据（data）是事实或观察的结果，是对客观事物的逻辑归纳，是用于表示客观事物的未经加工的原始素材”^[1]。《牛津英语词典》和刘建平给数据的解释可以概括为：“数据是反映客观事实、用于事实与统计汇集分析和参考、证据性材料的符号化表达”。“科学数据是数据的子范畴，科学数据作为数据概念分支的子类，‘科学’作为‘数据’的分类的限制词”^[2]，为特定数据定义提供了限定的场景。而其数据域是场景的范围和边界。

科学数据的国家层面文件界定及发展，早在 2005 年国家科技基础平台的《国家科学数据共享工程技术标准》中“特指以公益性和基础性为主体的，具有科学研究应用价值的的数据资源，包括在观测、监测、调查、实验以及研究等科学技术活动过程中产生的原始性数据，以及按照不同科技活动需求进行系统加工整理的各类数据”^[3]。2006 年科技部就发布了《科学数据共享工程技术标准（数据分类与编码方案）（SDS/T2122—2004）》，中的定义为“人类在认识世界、改造世界的科技活动中所产生的原始性、基础性的数据，以及按照不同需求系统加工的数据产品和相关信息”^[3]。科技部“科学数据共享调研组”定义为“科学数据是指人类社会科技活动所产生的基本数据，以及按照不同需求而系统加工的数据产品和相关信息”^[3]。在《中国数据要素市场发展报告》（2020-2021）中提到，数据资源是载荷和记录信息的按一定规则排列组合的物理符号的集合可以是数字、文字、图像，也可以是计算机代码的集合。2018 年出台的《科学数据管理办法》未对“科学数据”进行定义，只是在总则中解释为“在自然科学、工程技术科学等领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据”^[4]。而这一解释与国际上诸多对科学数据的定义内涵更为接近。“目前，我国对科学数据的定义变得更注重数据本身，取代了以往将科学数据混淆为数据共享对象及数据范围集合的定义方式”^[5]。

由此可见，科学数据主要包括纵观相关科学活动及学者在研究中根据不同的功能、场景、管理、用途、需求等进行概念的解读、总结、对比、分析、演绎，诠释了科学活动及科学研究环境下产生的数据及数据专属特征^[2]。

2 科学数据概念的数据范畴及多视角定义描述

2.1 《办法》中科学数据的含义的理解

《办法》中对科学数据未做定义，但外延进行了从界定，简言之，科学数据是科学研究中产生的数据，其包含了两个方面的数据：

① “基础研究、应用研究、试验开发等产生的自然科学、工程技术科学等领域数据”^[4]，主要强调了自然科学、工程技术科学。但没有提到社会科学数据。

② “用于科学研究活动的原始数据及其衍生数据（观测监测、考察调查、检验检测等）”^[4]。但没有提到社会普查、统计等数据。

另外，《办法》中通用词为“科学数据”未提及“研究数据”、“科研数据”等。《办法》中未明确提及：社会科学数据，人文社会科学数据等。

科学数据来源于科研活动过程中产生的数据（自然和工程技术领域），在现实中，人文社会科学研究活动过程也会产生研究数据（社会科学领域），所以广义的科研数据范畴也包含人文社科领域的科研数据。它的管理和应用主要的表现为社会领域调查、统计数据等，如：社会科学普查数据、社会科学调查数据、社会科学统计数据等。

2.2 科学数据的内涵多视角概念及描述

对相关文献的查找，我国众多学者对科学数据的内涵从不同视角给出“数据”概念描述，见表1：

表1 国内学者对科学数据概念汇总及分析

序号	描述视角	学者	科学数据描述内容及采用专属词	概念特点说明
1	基于功能的定义和基于情景过程的定义	刘建平 (2020)	“科学活动及科学研究环境下产生的数据及数据 专属特征 （是证据性的操作、符号化的表达）” ^[2]	是从数据产生的活动范围进行定义，对具体数据类型没有进行明确的说明，对有价值数据和无价值没有区分。
2	从科学数据本质进行界定研究角度出发	秦小燕、 初景利 (2019)	“①是指对研究成果进行 验证 的实际记录材料；②是将与科学研究相关的实验数据、实验笔记、图像、音视频、模拟系统等 半成品 作为科学数据范畴；③广义的理解，即包括科学研究全过程所产生的过程数据、半成品以及 研究成果 等” ^[6]	是从科学活动本质进行了区分，提到了“验证”、“半成品”数据。明确说明不包括科学研究过程中产生的笔记、初步的实验分析结果等未经过验证的数据、资料。
3	从数据产生根源进行描述	张计龙、 殷沈琴等 (2015)	沿用了(SDS/T2122—2004)概念，同时增加了“社会公益性事业部门开展的大规模观测、探测、调查、实验和综合分析所获得的长期积累与整编的海量数据，也包括国家科技计划项目实施与科技工作者长年累月科学实践所产生的大量数据” ^[7]	是从原始性、基础性数据、加工的数据进行定义。定义中包含了社会科学数据
4	在总结实践中	何思源 (2020)	“①广义的科学数据泛指科研活动中产生的所有数据，即被收集起来且被视为推理、讨论或计算基础的所有信息，特别是事实或数字。②狭义是指探测器等数据采集设备自动采集的数据，如观测数据；（包括通过处理、重建、分析等方式获得的数据和数据产品，还包括整个科研活动产生的设计文档、进展文档、成果、专利等。还涉及科研管理活动中形成的各类文档	①将科学数据理解为狭义和广义两种；②谈到了科研活动中的数据；③认同管理活动中的科研档案数据

			材料。” ^[8]	
5	从数据来源上的描述	司莉等 (2013)	“指在科技活动中或通过其他方式所获取的反映客观世界的特征、本质、变化规律等原始基本数据, 以及根据不同科技活动需要, 进行系统加工整理的各类数据集” ^[9]	提到加工的数据和原始数据。并说明是数据集。
6	对科学数据多元性描述	严春来 (2020)	“是指通过收集、观察、分析和创造各类实验、个人观测、互联网、统计、仿真等数据, 并以表格、数字、图像、新媒体等方式呈现出来的数据” ^[10]	把实验、个人观测、互联网、统计、仿真等数据认同为科学数据
7	从数据类型属性上认为	章琰等 (2021)	“是指来源于科学研究的非文本的事实记录, 包括地图集、基因组、化学化合物、数学公式、医学数据和临床试验等非文本材料” ^[11]	强调了非文本, 及列举了数据类型
8	从名称上的描述	唐素琴、赵宇 (2021)	“科学数据”(Scientific Data), 亦称为“科研数据”“研究数据”“技术数据” ^[12]	认同科学数据的多种概念词, 但没有给出具体的定义
9	从数据衍生上的描述	徐天雪 (2019)	“通过实验、观测、探查、分析等手段不断认识各种事物的科学研究过程中所产生政府部门或各类型科研机构等的的数据, 同时该数据加工所得的副产品也可看作特定类型的科学数据” ^[9]	提到政府、机构、二次元等数据
10	从数据来源上的描述	闫鑫, 黄国彬	“在科技活动中(模拟、仿真、计算等)或通过其他活动得到的反映客观世界本质特征变化规律等的原始基本数据, 以及根据不同科技活动需要, 进行系统加工整理的各类数据集” ^[13]	并说明是数据集。认同原始数据和加工的数据。
11	从数据分类上的描述	杨林、潘教峰、张晓林	“经过观察、感知、计算、实验、仿真、模拟等产生的自然、社会科学的数据, 包括观测型数据(Observational data)、计算型数据(Computational data)、实验型数据(Experimental data)、记录型数据(Records)4种类型” ^[14, 15]	认同自然和社会数据

结合前文提到的国家政策文件中科学数据的内涵范畴, 及表 1 国内学者对科学数据的含义概括和定义, 说明其具有领域的特殊属性、数据的基本属性及内涵和外延, 与科研活动和环境密不可分。由于视角的不同存在学科领域的类型区别、科学活动形态和获取方式的数据分类不同, 以及多维度特性概念域的科学数据集属性因素等概念的落脚点有所不同。

从表 1 看出, 对科研数据定义或描述侧重点各有不同, 有些定义是在《办法》出台之前的研究成果, 大部分是《办法》出台之后的定义, 普遍认同了①为了科学研究的目標, 在科研活动过程中产生的数据; ②社会活动及公益性事业中产生的科研数据(即, 社会科学数据); ③多样性称谓的科学数据, “研究数据”、“科研数据”、“技术数据”等的认同; ④范围的边界, 认为实验、观测、勘查、调研、仿真、互联网、衍生(加工过的数据、二次数据)、半成品、科技档案等数据属于研究数据范畴; ⑤管理数据是否是科学数据, 意见不一致; ⑥未经过验证的数据、资料、文本数据等不属于科学数据。

表 2 国外对科学数据概念描述

序号	概念来源	视角分类	定义内容	概念特点说明
1	Borgman (2015)	功能性定义	“(科学数据意味着) 科研活动中广泛多样的物质产出, 是以研究和学术为目的并可作为现象的证据	理解性广泛的定义

			实体” ^[16]	
2	scientific data Journal	功能性定义	“科研数据是具有 科学价值 的数据集” ^[14]	强调了有用的数据
3	爱丁堡大学	功能性定义	“出于产生原始性研究成果而搜集、观测以及生产的 信息 ” ^[17]	扩大到信息
4	帝国理工学院 (Imperial College London, UK)	功能性定义	《科学数据管理政策》中指出：科学数据“是以 收益性 研究、政府资助研究或者其他来源收集、生产、观察、制造从而取得的数据并应用于分析之后而产生的原始科研成果” ^[14]	价值属性
5	加拿大社会科学与人文学研究委员会 (SSHRC)	功能性定义	“为产生 知识 而使用特定方法结构化的数字化信息” ^[2]	Social Sciences and Humanities Research Council 的定义扩大到信息，强调了有用的数据，
6	美国国立卫生研究院	情景过程定义	“任何介质上记录的信息，主要有胶卷、书面材料、声音制品、制图、图片复制品、设计或其他图解显示、程序手册、工作流程图、表格、数据文件、框图、设备描述、数据记录或计算机程序（软件）、统计记录以及其他研究数据” ^[2]	广义的定义
7	美国新罕布什尔大学	情景过程定义	“为证实研究结果而被科学界普遍接受的、记录的真实材料。在涵盖原始实验结果和仪器输出，用于收集和重构数据的相关协议、数字、图形等，还包含数字、实地记录或通过观察、数据分析等方式采访、调查获得的数据、计算机文件、数据库、研究记录、幻灯片、音频/视频记录和照片等” ^[18-19]	广义的数据
8	英国工程和物理研究委员会	情景过程定义	“记录的事实材料，通常被科学界重新处理 (retrained)，作为验证研究结果所必需的材料” ^[20]	数据的二次加工
9	世界经济合作与发展组织 (OECD) (2007)	管理分类定义	“指 科研团体或科研工作者 从科学研究事实记录中得到的，他们认为对研究结果有用的数据，涵盖实验数值以及图像等，这是在《世界经济合作与发展组织关于公共资助科学数据获取的原则和方针》的定义” ^[21]	强调了数据的产生者

表 2 国外机构对科学数据的概念的概括，相比国内机构、学者给出的概念更为具体，对于不同数据类型针对性更强。每一个概念均有相应的政策文件支持，说明国外对于科学数据的研究比国内发展更早，也更全面。从内容的视角来看，每一个定义各具特色。同时。科学数据具有数据的基本属性及领域的特殊属性，与科研活动情景和环境密不可分。这一点在国外科学数据的管理和使用中更加融为一体。

3 基于不同概念的**科学数据“域”**的类型与属性

3.1 科学数据范畴及数据域边界属性

数据域类似数据区域划分及管理，对数据进行一种数据的分类及分类的标准，是对数据的边界和范围的筛选和圈定，使其不重不漏。具体分析《办法》阐述及学者的定义，数据域的边界特点有：①数据来源：科学数据来源主要有“研究者（学者本身）数据”和“研究者研究过程产生的数据”，研究者个人或项目组集体研究产生的数据，研究者个人和项目组成员研究产生数据（实验室实验、测试的数据）。其次是根据研究需要采集而来的各类数据。

另外根据研究需要二次加工或计算形成。②数据价值：“能保障和支撑科学研究和科研活动，归口科学数据”^[8]，或对后续的科研和相关业务活动有可持续生产要素的价值，也可以构成资本性科学数据。③数据范畴：《办法》重点阐述自然和工程技术领域的科学数据，但未明确提到社会科学数据，科研管理活动数据等。④数据属性：“科学数据在科研活动中的角色和作用已不再局限于科学研究的验证和补充，而被视为一种独立的科研成果”^[22]，数据资源成为生产要素的基础，具有资产的属性，数据的资产化形成资本，数据的价值属性使数据具有商品的交换属性^[23]。⑤数据利用：遵循规则基础上的数据共享、数据复用、协议使用等。由此可见，科学数据从数据的来源、价值、范畴、属性、利用等进行了数据域边界描述。

科研数据的数据域的边界和范围，在其范畴分类的标准划分之下，对其数据区域划分并进行管理。从领域科学数据视角分析，《办法》“自然科学领域、工程技术等领域的数据提到”^[4]，综合国内外机构和学者的阐述，在现实中科学数据的范围广泛地从自然科学数据、人文社会科学数据及其他类型的数据（含交叉科学、综合类等数据）。“从管理的数据域来看包含科研档案、科技管理、研究统计等数据”^[14]。“科学数据在实施管理过程中要进行细化，随着科学研究进入数据密集型科学范式，全领域、多类型、高复杂度的海量数据成为科学研究的中心”^[8]，从分类的多元化来看，科学数据集（Scientific Datasets）成为科学数据域管理的基本单位。

3.2 基于领域数据域的“科学数据”概念类型

对多元范畴中科学数据的内涵的界定，及科学数据的学科属性分析，在国外文献中“科学数据”一般用英文表达为（Research Data），国内学者通常称为“科学数据”。从数据域的角度，科学数据一般指数据学和数据科学（百度），对数据的属性归属可以从领域分类、学科分类、学术分类、理论分类、实践分类、过程分类、管理分类等多元范畴进行思考。从多种概念的综合描述，可概括为：高校科学（科研）数据即在科学研究过程中产生的有价值的数据集，在国内 Research Data 有时也翻译成“研究数据”或“科研数据”。也有国内学者认为是经由“科技数据”演化为“科学数据”。科学数据常用的同类或近义的广义词有数据科学，泛在的词有科研、实验室、统计、科技、科研、研究、档案等数据。

科学数据基于数据域的多元分类及数据域名称大概可以从以下视角进行解读：见表 3。

表 3 科学数据基于数据域的多元分类及名称

领域分类 (数据)	学科分类 (数据)	学术分类 (数据)	理论分类 (数据)	实践分类 (数据)	过程分类 (数据)	管理分类 (数据)
自然科学	能源矿业工程科学	科学	单元数据	事实统计	实验室	科研管理
人文社会科学	农业科学	研究	数据集	勘查	管理过程	档案管理
交叉科学	地球科学	科技	数据域	验证	档案	
综合性科学	深地深海科学	科研	数据子集	观测		

查阅国内外资料，英文直译表达主要有：“scientific data（科学数据），Science Data（科学数据），research data（研究数据），science and research data（科研数据）”^[24]。

^{25]}，在国外三种表达基本认为等同概念。分析《办法》中所说的科学数据，从众多异词概念来看，相近的概念词有：“研究数据”“科研数据”“科技数据”“科学数据”^[26]，四者含义既有共同点，也有细微的各自的特点及差异。研究数据主要指研究过程产生的数据（有价值+无价值）；科研数据主要指研究全过程的各类数据（有价值+无价值）；学术数据就是对数据集的一种概括性的名称（有价值）；科技数据指研究的成果的技能数据（有价值）。如果加上学科数据域、概念域、数据集，科学数据（自然科学、社会科学、交叉学科）、档案管理等数据概念及扩展称呼，每个概念之间并不是完全的包含或下位类的关系，也存在一定交叉与重叠。主要概念词及边界域见表4。

表4 基于科学数据的数据域及类属

名称	边界归属（数据域）
科学数据	自然科学
研究数据	自然科学、社会科学
科研数据	自然科学、社会科学
科技数据	自然科学、工程
学术数据	自然科学及社会科学
自然科学数据	自然科学
社会科学数据	社会科学
交叉科学数据	自然科学、社会科学
科技档案数据 ^[5]	提取保存级数据（包含自然和社会科学数据）

4 科学数据范畴、属性与管理政策

数据科学与科学数据的重要性的体现，为了管理和规范国家出台了许多关于数据管理与政策相关的文件，汇总如下：

表5 针对数据科学的管理实施的相关文件

出台时间	管理文件	机构名称	文件主要规定与职责
2006	《知识产权法》 (数据部分)		法律文件：以法律的形式确认了数据权保护
2013	《“十二五”国家自主创新能力建设规划》	国务院办公厅	管理规划：构建科技资源的集中与分散数据中心群，从采集、存储、管理、挖掘数据的完整信息到开放共享的有效服务链，形成国家科学数据分级分类共享服务体系。
2015	《数据治理白皮书》	国务院办公厅	管理规划：加强和规范数据管理，
2016	《中华人民共和国国民经济和社会发展第十三个五年规划纲要》		指导文件：数据作为基础性战略资源，实施推动数据资源共享开放和开发应用。
2015	《促进大数据发展行动纲要》	国务院办公厅	实施指导：指导数据管理
2016-2018	《国家科学数据资源发展报告》，		年报：科学数据资源现状及发展预测
2018	《科学数据管理办法》	国务院办公厅	管理文件：科学数据进行规范管理，重视科学数据安全，发挥开放共享水平和效果。
2019	《中国科学院科学数据管理与开放共享办法(试行)》	中科院	管理规则：细化了科学数据管理，提高开放共享水平和保障科学数据安全提供了制度规范。
2019	《科研数据北京宣言》	科技部、中国科学院	管理文件：科学数据管理，提高开放共享水平，重视数据要素、进行数据分类分级安全保护。

		国际科学理事会数据委员会 (CODATA)	
2020.5	《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》含数据要素部分	国务院办公厅	资产管理：数据作为生产要素，进行市场化运作，作为数据资产的指导文件。
2020.7	《数据安全法》(简称《数安法》) 《中华人民共和国数据安全法(草案)》	国务院办公厅	规范数据安全全：“对数据活动进行依据法律的规范处理，数据安全、个人、组织的合法权益得到保护，对国家主权、安全和发展利益进行捍卫。
2020.9	《科学技术研究档案管理规定》	国家档案局局长办公会议、科学技术部常务会议修订通过	旨在加强科学技术研究档案管理，有效保护和利用科研档案
2021.8	《中华人民共和国个人信息保护法》	全国人大	法律文件：包含保护个人数据信息的法律条款。对拥有的科学数据集进行审核等
2021.9	《知识产权强国建设纲要(2021—2035年)》	中共中央、国务院	有关数据要素的保护，数据版权
2021.11	《数据出境安全评估办法》(征求意见稿)		数据安全管理和鉴定

以上这些是根据政策、法律、规则等制定的相应文件细化了《办法》实施的策略和实施细则，为《办法》的实施提供了政策、法律、实施规则等方面的保障。

从上述文件对数据范围、获取、存储、挖掘、管理、共享、管理、规划、发展、指导、开放、复用、交汇、数据资产、保护、安全、权利、使用范围等都有所涉及。

但对有关数据的版权、数据、完整性验证、有效性确定，对专门数据元数据设计及元数据框架体系、登记制度、数据管理机构设定及权限、数据资产的交易、数据使用范围限定等方面缺乏特别专指的文件、规范、法律条文及实施范例。

5 科学数据分类管理要素与管理边界

5.1 科学数据分类管理的现状

数据的分类是数据管理的基础，有效的分类，能很好的确定数据范围边界、管理边界及使用权限，主要分类有：从内容关联角度：“科学数据相关的内容领域对科学数据进行分类”^[27、28]，“科学数据的分类方式以内容相关的角度进行划分，就是领域数据与使用者需求内容的匹配情况”^[29]；从数据完整前面角度：“从数据内容类型与数据格式类型完整的角度对科学数据进行分类，指在整个科学数据生命周期中数据内容保存的准确与完整”^[30]；从数据来源可靠性的角度：“对数据准确时效的处理过程，渠道可靠的数据来源”^[29、31、32]；从数据权威性的视角：数据获取组织、个人、机构的权威性，因内外部因素影响而具有典型性、不可获缺、代表性等，“按组织机构类型对科学数据进行分类，可以不同类型机构的权威程度不同的思路出发进行研究”^[33]，“科学数据的开放共享是数据进行分类的一种方式”^[13、34]。

科学数据的分类依据如数据边界、数据价值等性质进行分类，目前报道不多。本文从《办法》中对科学数据的含义或描述侧重点来梳理其分类要素，以为科学数据的边界管理与面向科学数据的使用及价值再造等提供理论支撑。

5.2 科学数据的管理要素与管理边界

科学数据的分类、概念范畴、内容限定等都影响其数据的边界及管理因素，从上述科学数据边界范畴及数据域的解读及分析来看。

科学数据的边界确定：①科学数据要素：指影响科学数据所在的数据领域与数据使用关键因素，不同的数据要素影响数据的管理内容与管理边界，根据科学数据的概念及概念范畴，数据要素主要有数据主体、属性、质性、关联、再生性、生命周期、有效性、价值、复用、共享等；②数据相关性：原生、演绎、再生等数据及数据关联（关联数据）、数据管理等；③数据价值及价值再造：数据资源、数据资产及资产化后成为资本等。④数据生命周期：数据效应、半衰期、有效性、无用性等；⑤数据版权：数据的 IP（创新版权）、登记、交汇、出版、传播等；⑥数据利用：数据的协议、共享、复用、权限等。确定数据边界的因素很多，本研究主要就这些主要因素进行分析和研究与复杂关系下的管理边界问题。

表 6 数据科学的要素及实施边界

	边界特征	边界要素	边界限定	说明
1	数据来源	发布主体特性、数据关联、内容相关性、	领域分类、实践分类、过程分类	有利于领域的 数据归纳和实际使用
2	数据价值	数据本身价值（数据资源、数据资产、数据资本等）、价值再创造特性	过程分类	有利于数据增值
3	数据范畴	数据集名称范畴、数据集划分依据、数据生命周期与时效控制特征、学科域	学科分类、管理分类	有利于区分有价值的数据集无用的数据
4	数据属性	学科属性、类型质性、使用权限、数据有效性、数据版权	学术分类、理论分类、过程分类	有利于归类使用
5	数据利用	复用、共享、利用协议	实践分类、过程分类	有利于保护数据的权限

6 科学数据边界管理的现行规则、工具及发展

6.1 科学数据管理规则及主要管理内容

目前，科学数据相对认同的规范主要有：见表 7：

表 7 数据科学实施管理的认同规则

来源	规则名称	规则内容	规则的边界
Wilkin-son ^[24]	FAIR 数据原则	可找寻性 (Findable)、可获取访问性 (Accessible)、可交互操作性 (Interoperability)、重复利用性 (Reusable),	提供了研究领域数据共享的准则，按数据规则分享
多方	科学数据管理生命周期模型	以生命周期模型为基础的科学数据管理架构及模式	
美国国会图书馆	RDA 国际编目规则	RDA 是在数字环境下资源描述与检	数字环境而设计的数

等编制	(Resource Description and Access, 资源描述与检索) 的国际标准	索的新标准。	字资源编目(资源描述与检索), 强调关联关系的描述, 可关联研究成果和研究数据。
JISC 与 2016 年出版	RDM (科研数据管理 Research Data Management, RDM) 政策	基于《科研人员参与管理科研数据指南》主要面向大学的研究人员以及科研数据管理支持人员, 如向科研人员提供科研数据存储, 管理, 出版和存档服务或建议支持的人。	数据管理策略: 对于基于数据的服务、安全、出版、质量评价、科学管理政策研究等。
多方	数据资产框架(Data Asset Framework, DAF)	DAMA 数据治理框架、DGI 数据治理框架、IBM 数据治理框架、EDM 模型、层框架模型、决策域模型等	对数据资产管理行使权力和控制的活动集合 ^[35]
国家新闻出版广电总局	《版权信息核心元数据标准》	2015 年制定发布了 (CY/T 134—2015)	实施数据版权的信息标准

6.2 不同类型数据管理要旨及主要策略

从科学数据“数据域”角度考虑,《办法》中的“数据域”是自然科学、工程技术科学等领域,但在实施过程中要进行细化,从自然科学和人文社会科学两大类数据进行管理^[36-37]。制定并完善科学数据管理计划并依据各个学科领域的不同,计划要依据数据管理的环节有所不同,但宗旨不变,为了科学数据更好地共享利用^[38、39]。《办法》中也强调了“要以建立的各领域的数据中心的方式,统一加强管理科学数据的全流程使用”^[4],从数据管理初期的采集汇交、管理中期的挖掘分析、管理中期的保存以及管理后期的共享再利用四个步骤进行解读^[40]。

6.2.1 科学数据初期管理的采集汇交

科学数据的采集汇交数据管理的首要基础^[41]。《办法》从国家层面强行规范各个法人主体单位/机构实施科学数据汇交要求,目的是保障数据的完整性、使用性^[42]。以国家农业科学数据中心为例,在农业领域的数据管理汇交严格依照《农业科学数据汇交管理办法》进行管理^[43]。2021 年在对科技部科技计划项目科学数据汇交工作中,遵守先由数据汇交的项目方提交数据汇交计划,国家农业科学数据中心审核通过后,方可提交科学数据,由领域专家对数据进行审核后反馈给国家农业科学数据中心,通过的出具科学数据汇交凭证。建立起了先汇交科学数据再验收项目的管理机制。来弥补了在各领域基础研究中数据难以收齐的短板。

6.2.2 科学数据中期管理的挖掘分析

科学数据在中期管理都要历经挖掘分析这一阶段。收集汇交属于第一步骤,拥有了数据,不可能置之不理,而是要更好地进行挖掘分析,从而使得数据可用。根据不同学科、不同领域、不同类型,利用相关的适合的数据管理、数据存储、数据挖掘与分析的软件工具,采用相应和先进的数据处理技术,将数据采集、登记、汇交的数据进行处理,来提高数据的使用价值,开展对应的市场增值服务,如对企业进行数据决策支持等,实现科学数据的内在价值。

6.2.3 科学数据中期管理的保存

科学数据都需要合理、有效、长期的存储。这是对科学数据重用以及管理具有可持续发展意义的至关重要的环节^[44]。《办法》对法人单位/机构/部门要求要建立数据保存制度，以确保未来的科研活动不会失去原有的数据资源。同时做好数据存储、管理、使用和安全等设施，保障科学数据的完整性、可持续性与安全性。针对不同学科领域，使用合理科学的备份方式。

6.2.4 科学数据后期管理的共享利用

科学数据都需要被合理合规的共享利用^[45]。数据的共享，对于将研究成果快速转化为知识以及数据产品起到了重要的作用。目前尚未出台合理的有效的数据共享机制，各个行业、各个领域的大量科学数据依旧分散在各个政府部门、各个科研团队手中。《办法》中规定了以开放为常态的原则，统筹编制科学数据的资源目录，面向社会开放共享努力提由主管部门组织高共享效率^[4]。数据主管的部门也应积极鼓励数据的出版工作，完善数据孤岛现象有利于科研活动高效开展，公益性的数据可无偿共享，对使用科学数据的盈利性活动，双方签订有偿合同，以保证双方利益共赢。

6.3 基于概念域高校和科研院所等落实《办法》的保障措施

高校和科研院所是科学数据产生的主要机构，通过以上的阐述和分析，为了保障《办法》的落实，主要要做到以下几点：

① 充分理解办法的内涵，在重视自然科学、工程技术科学等领域的科学数据及科学数据集，也要重视社会科学数据、人文社会科学数据及数据集，并且要重视有价值数据的科学数据档案归档的收集和登记管理。

② 重视数据平台的建设，通过对办法中科学数据的概念域分类，了解高校及科研院所的各类数据类型，利益相关者之间的相互依赖关系为科学研究创造了一个富有成效的环境^[46]，对科学数据资源进行有序的存储、安全、组织和管理

③ 高校及科研院所具有一定的领域优势和专业特色，“建立跨机构的领域科学数据汇交制度，在国家统一政务网络和数据共享交换平台的基础上开展本领域（本专业）的科学数据汇交工作”^[4]，在目前缺乏科学数据登记制度的情况下，把科学数据的数据汇交和登记管理协同实施，建立数据分级分类制度与机制^[47]，科学数据登记制度建立打好基础。

④ 科学数据具有资产的属性，在十九届四中全会从国家治理体系和治理能力现代化的高度把数据也视为生产要素，并要求加强数据有序共享，这就需要推进“数据资源-数据汇交-数据登记管理-数据资产化-数据复用^[48]”的管理进程。

7. 结语

通过对《办法》科学数据概念认知和相关定义的梳理，明确了科学数据的对象、管理服务、数据边界的限定，对于落实《办法》，加快相关机构针对性《科学数据管理办法》相

关政策出台及实施,有利于分类别、分区域、分领域的明确科学数据采集管理机制、数据登记、主体责任、规范业务、实施流程、权限归属、数据平台汇交、数据复用等,为我国科学数据管理与开放共享提供保障机制,汇交各级各类数据进行融合管理,保障国家数据战略的实施和发展。

参考文献

- [1]陈柄臣. 侵害公民个人数据犯罪研究[J]. 社会科学家, 2021(8):113-118.
- [2]刘建平. 科学数据用户相关性判断模型研究[D]. 北京: 中国农业科学院, 2020.
- [3]杨从科. 中国农业科学数据资源建设研究[D]. 北京: 中国农业科学院, 2007.
- [4] 《科学数据管理办法》
[EB/OL]. [2022-01-01]. <https://baike.so.com/doc/27159694-28552208.html>.
- [5]徐天雪. 基于 AHP 的科学数据开放共享政策评估指标体系研究[D]. 哈尔滨: 黑龙江大学, 2019.
- [6]秦小燕, 初景利. 科学数据素养内涵结构研究[J]. 图书情报工作, 2019, 63(18):30-39.
- [7]张莉. 中国农业科学数据共享发展研究[D]. 北京: 中国农业科学院, 2006.
- [8]何思源. 科研档案和科学数据协同治理的动因分析[J]. 档案管理, 2020(6):64-66.
- [9]司莉, 邢文明. 国外科学数据管理与共享政策调查及对我国的启示[J]. 情报资料工作, 2013(1):61-66.
- [10]严春来. 基于知识图谱的国内科学数据研究热点和趋势分析[J]. 电脑编程技巧与维护, 2020(12):137-138; 156.
- [11]章琰, 杨一图, 吴健, 等. 我国科学数据共享运行机制模式创新探讨——以产业技术联盟为例[J/OL]. 科学学研究, 2021(11):1970-1979.
- [12]唐素琴, 赵宇. 《数据安全法》突出科学数据的必要性研究[J]. 中国科技资源导刊, 2021, 53(2):19-25; 110.
- [13]闫鑫, 黄国彬. 科学数据分类研究述评[J]. 图书馆论坛, 2020, 40(5):45-54.
- [14]杨林, 钱庆, 吴思竹. 科学数据管理生命周期模型比较[J]. 中华医学图书情报杂志, 2016, 25(11):1-6.
- [15]潘教峰, 张晓林. 第四范式:数据密集型科学发现[M]. 北京:科学出版社, 2012.
- [16]Borgman Christine L.. Big Data, Little Data, No Data:Scholarship in the Networked World[M]. The MIT Press:2015-01-02.
- [17]Li Cai, Yangyong Zhu. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era[J]. Data Science Journal, 2015, 14: 补页码.
- [18]Rice R, Haywood J. Research data management initiatives at University of Edinburgh[J]. International Journal of Digital Curation, 2011, 6(2): 232-244.
- [19]洪程. 我国高校科研数据管理政策研究[D]. 湘潭: 湘潭大学, 2019.
- [20]Borlund Pia. The concept of relevance in IR[J]. Journal of the American Society for Information Science and Technology, 2003, 54(10): 补页码.
- [21]柴会明, 张立彬, 赵雅洁. 国内图书馆科学数据研究述评[J]. 图书情报工作, 2019, 63(7):116-126.
- [22]迟玉琢. 科学数据能力研究: 内涵、框架和影响因素[J]. 农业图书情报学报, 2020, 32(1):23-29.
- [23]都平平, 李雨珂, 陈越. 高校科研数据资产化存储及数据复用权益许可研究[J/OL]. 图书情报工

作:1-9[2022-01-11]. <http://kns.cnki.net/kcms/detail/11.1541.G2.20211222.1112.002.html>.

[24] Latham, Bethany. Research Data Management: Defining Roles, Prioritizing Services, and Enumerating Challenges[J]. *The Journal of Academic Librarianship*, 2017, 43(3):263-265.

[25] Wilson J A J, Martinez-Uribe L, Fraser M A, et al. An institutional approach to developing research data management infrastructure[J]. *International Journal of Digital Curation*, 2011, 6(2): 274-287.

[26] 顾立平. 数据治理-图书馆事业的发展机遇[J]. *中国图书馆学报*, 2016, 42(5):40-56.

[27] 周力虹, 段欣余, 宋雅倩. 我国高校图书馆科研数据管理服务调查与分析[J]. *图书情报工作*, 2017, 61(20):77-86.

[28] Christine L. Borgman, Peter T. Darch, Ashley E, et al. Knowledge infrastructures in science: data, diversity, and digital libraries[J]. *International Journal on Digital Libraries*, 2015, 16(3-4): 补页码.

[29] 闫鑫, 黄国彬. 科学数据分类研究述评[J]. *图书馆论坛*, 2020, 40(5):45-54.

[30] 公晓. 图书馆科研数据联盟管理模式建设研究[D]. 大连: 辽宁师范大学, 2017.

[31] 周倩. 面向科学数据出版的信息资源开发利用研究——以国防科技领域为例[J]. *情报理论与实践*, 2019, 42(2):140-144.

[32] 朱兰, 左维, 李春来. 绕月探测工程卫星数据的存储与管理[J]. *天文研究与技术*, 2008, 5(4):365-372.

[33] 李志芳, 邓仲华. 国内开放科学数据的分布及其特点分析[J]. *情报科学*, 2015, 33(3):45-49.

[34] R Sinnott, M Bayer, A Stell, et al. Grid infrastructures for secure access to and use of bioinformatics data: experiences from the BRIDGES project[M]. *First International Conference on Availability, Reliability and Security (ARES'06)*, 2006, 8-957.

[35] 刘桂锋, 阮冰颖, 刘琼. 加强数据安全防护 提升数据治理能力——《中华人民共和国数据安全法(草案)》解读[J]. *农业图书情报学报*, 2021, 33(4):4-13.

[36] 宋永辉, 姜鑫, 马廷灿. 科学数据政策研究的热点及演化趋势分析[J/OL]. *科学观察*:1-19[2021-11-15]. <https://doi.org/10.15978/j.cnki.1673-5668.202106001>.

[37] 孟银涛, 赵蕾霞, 于倩倩. 面向科学数据保存的科学数据评价指标体系研究[J/OL]. *农业图书情报学报*:1-12[2021-11-15]. <https://doi.org/10.13998/j.cnki.issn1002-1248.21-0271>.

[38] 王勋荣, 蒲筱哥. 高校科学数据共享影响因素与策略研究[J]. *图书馆*, 2021(10):41-46;62.

[39] 方恒, 宋立荣. 国内外期刊机构科学数据出版政策研究[J/OL]. *情报杂志*:1-10[2021-11-15]. <http://kns.cnki.net/kcms/detail/61.1167.G3.20210929.1715.043.html>.

[40] 陈飞飞, 宋丹阳. 图情期刊科学数据规范调查与分析[J]. *数字图书馆论坛*, 2021(9):56-61.

[41] 翟军, 梁佳佳, 吕梦雪, 等. 欧盟开放科学数据的 FAIR 原则及启示[J]. *图书与情报*, 2020(6):103-111.

[42] 蔡盈芳. 推进科学数据与科研档案的协同管理[J]. *中国档案*, 2021(9):60-61.

- [43]柴苗岭,黄琳,任运月.重要开放农业科学数据资源建设现状综述[J].农业图书情报学报,2020,32(10):25-34.
- [44]张群,印熙媛.国内外数据素养教育研究现状及启示[J].数字图书馆论坛,2020(12):65-72.
- [45]周成虎,孙九林,苏奋振,等.地理信息科学发展与技术应用[J].地理学报,2020,75(12):2593-2609.
- [46]王成一.加拿大科学数据管理及启示[J].图书馆杂志,2020,39(6):77-82;91.
- [47]周振国.治理框架视域下的数据治理研究[J].农业图书情报学报,2020,32(7):57-62.
- [48]刘静羽,贾毓洁,张婧睿,等.大数据环境中的数字资源权益管理:关键问题与应对策略——2021年大数据环境中数字资源权益管理研讨会综述[J/OL].图书馆杂志:1-12[2021-11-30].

都平平 女,中国矿业大学图书馆,研究馆员。研究方向:知识产权、数据管理。作者贡献:文章框架设计、论文撰写。E-mail:ppdu168@126.com 江苏徐州 221116

李雨珂 女,上海出版印刷高等专科学校文化管理系,讲师。研究方向:数字经济。作者贡献:英文文献翻译与国外现状部分查阅。上海 200093

张雪媛 女,中国矿业大学图书馆,馆员。研究方向:资源管理。作者贡献:参与技术部分撰写。江苏徐州 221116