

摘要

FAIR 数据管理原则是各界人士协商形成的科学数据管理与开放共享规范，旨在促进数据的有效共享利用。通过设计可查找、可访问、可互操作和可重用的数据和算法，以及产生这些数据的工具和工作流程，可以生成更多有价值的研究。FAIR 原则自提出后在研究界得到广泛传播。2018 年我国出台了《科学数据管理办法》，对我国科学数据全生命周期的管理做出了详细规定，初步明确了科学数据开放共享的管理机制和工作内容，但对于如何确保科学数据有效利用的关注还不够。因此本文将考察《科学数据管理办法》对 FAIR 原则内涵的支持程度，探讨是否可以将 FAIR 原则的相关理念纳入政策中以推动该原则在我国的实施。由于 FAIR 原则不是标准，是指导方针，并未涉及具体的实现技术、途径，以及具体的实施标准与规范，因此有必要先确定支持 FAIR 原则的政策要素，进而考察《科学数据管理办法》对这些要素的支持情况，最后再提出我国科学数据开放共享工作完善建议。

本文在第一章阐述了文章的研究背景和研究意义，梳理国内外的研究现状，分析当前国内外研究的特点并指出不足，简要概述了本文的研究内容、研究方法和研究创新点。第二章在详细介绍 FAIR 原则产生背景及内涵、《科学数据管理办法》内容体系的基础上，总结出二者的共性与差异，为后文的研究奠定基础。第三章运用网络调查法和质性文本分析法从两方面提取 FAIR 政策要素，一是从欧洲国家的 FAIR 数据政策文本中提炼，二是从相关报告中筛选出针对政策制定者的有关建议并进一步总结，再将两份 FAIR 政策要素进行比较，初步得出 FAIR 政策要素框架。通过电子邮件的形式向研究领域内的专家发放咨询表征集专家意见，最终确定 26 个 FAIR 政策要素。第四章探讨了《科学数据管理办法》对 FAIR 政策要素的支持情况，并对结果进行详细分析。第五章提出基于 FAIR 原则的《科学数据管理办法》优化具体建议，为完善和修订我国科学数据管理政策、优化我国科学数据开放与共享工作提供一些参考。第六章对本文的研究内容和结论进行整理并总结，指出研究中的不足，对 FAIR 未来进行展望。

关键词：FAIR 原则；《科学数据管理办法》；数据开放共享；科学数据政策

Abstract

FAIR data management principles are scientific data management and open sharing norms negotiated by people from all walks of life, aiming to promote the effective sharing and utilization of data. More valuable research can be generated by designing findable, accessible, interoperable, and reusable data and algorithms, as well as the tools and workflows that produce this data. The FAIR principle has been widely disseminated in the research community since its inception. In 2018, my country promulgated the " *Scientific Data Management Rule* ", which made detailed regulations on the management of the full life cycle of scientific data in my country, and initially clarified the management mechanism and work content of the open sharing of scientific data. Attention is not enough. Therefore, this paper will examine the degree of support for the connotation of the FAIR principle in the " *Scientific Data Management Rule* ", and explore whether the relevant concepts of the FAIR principle can be incorporated into policies to promote the implementation of the principle in my country. Since the FAIR principle is not a standard, but a guideline, it does not involve specific implementation technologies, approaches, and specific implementation standards and norms. Therefore, it is necessary to determine the policy elements that support the FAIR principle, and then examine the " *Scientific Data Management Rule* " on these. Finally, we put forward suggestions for improving the open sharing of scientific data in my country.

In the first chapter, this paper expounds the research background and research significance of the paper, sorts out the research status at home and abroad, analyzes the characteristics of the current research at home and abroad and points out the shortcomings, and briefly outlines the research content, research methods and research innovations of this paper. The second chapter summarizes the commonalities and differences between the two on the basis of introducing the background and connotation of the FAIR principle and the contents system of the " *Scientific Data Management Rule* ", which lays the foundation for the following research. Chapter 3 extracts FAIR policy elements from two aspects by using network survey method and qualitative text analysis method. One is to extract FAIR data policy texts from European countries. After further summarizing, the two FAIR policy elements are compared, and the FAIR policy element framework is initially obtained. The consultation characterization set expert opinion was distributed to experts in the research field by e-mail, and the 26 FAIR policy elements were finally determined. The fourth chapter discusses the support of the " *Scientific Data Management Rule* " to the FAIR policy elements, and analyzes the results in detail. The fifth chapter puts forward specific suggestions for the optim

ization of the " *Scientific Data Management Rule* " based on the FAIR principle, and provides some references for improving and revising my country's scientific data management policies and optimizing my country's scientific data opening and sharing. The sixth chapter organizes and summarizes the research content and conclusions of this paper, points out the deficiencies in the research, and looks forward to the future of FAIR.

Key words: FAIR principle; *Scientific Data Management Rule*; Open Data Sharing; Science Data Policy

目录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.1.1 研究背景	1
1.2.1 研究意义	2
1.2 国内外研究现状	2
1.2.1 FAIR 原则研究现状	2
1.2.2 《科学数据管理办法》研究现状	5
1.2.3 国内外研究现状述评	6
1.3 研究内容与方法	7
1.3.1 研究内容	7
1.3.2 研究方法	7
1.4 研究创新点	8
第 2 章 FAIR 原则与《科学数据管理办法》内容及比较	9
2.1 FAIR 原则的产生背景与内容	9
2.1.1 FAIR 原则产生背景	9
2.1.2 FAIR 原则具体内容	9
2.2 《科学数据管理办法》内容体系	13
2.2.1 科学数据的采集、汇交与保存	13
2.2.2 科学数据的共享与利用	13
2.2.3 科学数据的保密与安全	13
2.3 FAIR 原则与《科学数据管理办法》的比较	13
2.3.1 制定背景与目标	13
2.3.2 内容与性质	14
2.3.3 实施重点	14
2.3.4 比较结果	14
第 3 章 FAIR 政策要素提取与确定	15
3.1 基于政策内容分析的 FAIR 政策要素提取	15
3.1.1 文本收集	15
3.1.2 开放式编码（一级编码）	16
3.1.3 主轴式编码（二级编码）	19
3.1.4 选择性编码（三级编码）	22

3.1.5 理论饱和度检验	23
3.2 基于政策建议报告的 FAIR 政策要素提取	23
3.2.1 政策建议报告选取	23
3.2.2 FAIR 政策特征要求	23
3.2.3 FAIR 政策要素筛选	24
3.3 FAIR 政策要素的对比取舍及其初始框架	25
3.3.1 FAIR 政策要素对比取舍	25
3.3.2 FAIR 政策要素初始框架	26
3.4 FAIR 政策要素最终确定	27
3.4.1 专家咨询表设计	27
3.4.2 专家的基本情况	28
3.4.3 调查结果量化分析	28
3.4.4 FAIR 政策要素修改	29
3.5 FAIR 政策要素描述	31
3.5.1 外部特征维度	33
3.5.2 内容要素维度	33
第 4 章 《科学数据管理办法》对 FAIR 原则的支持程度	35
4.1 支持级别的确定	35
4.1.1 支持级别判断方法	35
4.1.2 与 FAIR 要素匹配情况	35
4.2 《科学数据管理办法》对 FAIR 原则的支持情况分析	37
4.2.1 《科学数据管理办法》完全支持的 FAIR 要素	37
4.2.2 《科学数据管理办法》部分支持的 FAIR 要素	38
4.2.3 《科学数据管理办法》不支持的 FAIR 要素	41
第 5 章 《科学数据管理办法》优化建议	45
5.1 适当将 FAIR 原则理念要求纳入法律规范中	45
5.1.1 尽量融入 FAIR 原则中的内涵与要求	45
5.1.2 使用结构化数据标记模式描述政策	46
5.1.3 重视科学数据重用中元数据的价值	46
5.2 借助工具强化政策的指导性与可操作性	47
5.2.1 制定数据管理计划以推动数据共享	47
5.2.2 制定政策指南以提升政策可操作性	48
5.3 提供强有力的保障确保政策的贯彻落实	48
5.3.1 加强科学数据管理财政支持与激励	48

5.3.2 完善科学数据管理的评价考核制度	49
5.3.3 建立科学有效的监督机制保障实施	49
第 6 章 总结与展望	50
6.1 研究总结	50
6.2 研究不足	50
6.3 研究展望	51
参考文献	52
致谢	59
附录	61
个人简历、在学期间发表的学术论文及研究成果	74

第 1 章 绪论

1.1 研究背景与意义

1.1.1 研究背景

近年来，随着科学数据数量的增加，人们越来越重视科学数据的价值，对科学数据开放共享的看法也发生了很大转变。欧洲核子研究中心（CERN）的一项特别研究发现，开放科学数据的价值主要在于其质量，由两个因素决定——组织内强大的数据管理实践以及开放数据在未来使用和重用的潜力^[1]。一方面，开放科学数据通过增加科学知识和数据的吸收以及再利用来推进民主化科学，能提高已发表科学成果的质量和透明度，使科学成果能够得到验证和可重复性，并促进科学生产和传播向数字模式的持续转变。另一方面，全球变暖，粮食安全和贫困，对能源和资源的无止境的需求，污染加剧，城市化进程日益加剧，以及寻求增加知识，长寿和提高生活质量，这些都越来越依赖于科学和技术的应用，开放科学数据有助于打破不同国家、不同领域之间应用科学技术的壁垒。当前科学大数据潜力在不断增长，只有产生数据的研究人员共享它们的科学数据，让数据可以被其他人挖掘和重用，才能从泛滥的数据中获得收益。为了推动科学数据的大规模集成和高效重用，2014 年众多学术界、出版商和政府等利益相关者在洛仑兹举办了研讨会，会上推出“FAIR”一词，所形成的 FAIR 原则草案最初通过同行倡议的网站（Force11）向公众征求意见，根据反馈，最终于 2016 年发布。FAIR 是一组指导原则，用于使数据可查找、可访问、可互操作和可重用。FAIR 数据管理原则（简称 FAIR 原则）通过指出使用清晰的机器可操作元数据来丰富数据集的相关性和重要性，为数据的 FAIR 性制定了指导方针，其目的是促进数据的有效共享利用，确保研究人员不会花费不必要的时间、资源来收集和整理已经存在的数据。欧盟一项调查报告显示，不共享 FAIR 数据产生的欧洲年度财务成本至少为 102 亿欧元，而 FAIR 对潜在经济年增长的影 响估计为每年 160 亿欧元^[2]。

近十多年来，科学数据开放共享一直受到国家的重视。我国科技部于 2002 年就开始实施科学数据共享工程，经过近 20 年的探索，在基础科学、农业、林业、海洋、气象、地震、地球系统科学、人口与健康等领域建成形成了一批资源优势明显的科学数据中心。在 2001 年至 2021 年间，在国家层面正式出台的 科学数据开放共享政策就有 114 项^[3]，如《气象资料共享管理办法》（2001）、《关于推进交通运输行业数据资源开放共享的实施意见》（2016）、《科学数据管理办法》（2018，以下简称《办法》）、《工业和信息化部关于工业大数据发展的指导意见》（2020）等。显然我国在

科学数据开放共享工作上已取得一些成效，但整体而言，我国科学数据数据的开放共享水平还落后于发达国家，数据开放共享的制度保障体系有待健全，在指导政策和法规制定的原则上尚未达成共识是主要原因之一。当前，FAIR 原则已成为国外众多政府、科研机构、资助机构、出版商等制定数据战略、政策和法规的重要指导原则之一，在保障开放科学数据的可持续发展上发挥着积极作用^[4]，适当引入 FAIR 原则的先进理念可以减少数据共享中的阻碍，有助于推动我国科学数据开放共享工作的发展。

1.2.1 研究意义

(1) 拓宽了科学数据开放共享的研究视角。从 FAIR 原则角度探讨《办法》优化，丰富了我国科学数据共享与利用领域的研究内容，同时研究国际科学数据管理 FAIR 原则也有助于推进我国科学数据理论研究领域中研究范围的拓展。

(2) 提出《办法》优化建议，为制定实施细则或修订政策提供参考借鉴。本文在考察我国《办法》对 FAIR 原则支持程度的基础上，提出的诸多优化建议，对我国科学数据政策的完善以及科学数据开放共享工作的改进具有理论指导意义。

(3) 探讨从政策层面纳入 FAIR 原则相关理念，推进我国科学数据管理与共享实践。FAIR 原则是在科学数据开放共享的背景下提出的，目的是实现科学数据的充分共享利用，吸收 FAIR 原则中的合理部分，可以更好地推进我国的科学数据共享利用。

1.2 国内外研究现状

近年来，国内外学者们从不同角度对科学数据管理 FAIR 原则和《办法》进行了研究，产出了丰硕的学术成果。调研国内外学者的相关研究文献，有助于笔者全面把握当前研究趋势。

1.2.1 FAIR 原则研究现状

(一) 国外 FAIR 原则研究现状

自 2016 年 FAIR 原则在 *Scientific data* 上发布以来，得到了众多利益相关者的引用、认可与采纳。欧洲理事会将开放科学和研究数据的可重用性作为优先事项。科学欧洲 (Science Europe, SE) 积极倡导 FAIR 科学数据的共享和重用，协调政策和程序，并提供明确的指导方针，以支持研究人员的数据管理。美国国立卫生研究院 (National Institutes of Health, NIH) 制定《数据科学战略计划》(NIH Strategic Plan for Data Science)，要求在管理 NIH 所资助项目产生的数据时遵循 FAIR 原则^[5]。欧盟委员会 (European Commission, EC) 很早就将 FAIR 原则纳入欧洲科学研究战略

政策，法国、芬兰、奥地利、瑞典和英国等众多欧洲国家将其开放科学政策向 FAIR 原则作出内容调整^[6]。与 FAIR 实践进展同步的是，国外学者在理论研究上围绕 FAIR 原则展开了诸多有益探讨，其研究内容可从以下几个方面进行概括：

(1) FAIR 原则内容阐述及其实施框架研究。2016 年，Wilkinson MD 等发表了《科学数据管理 FAIR 指导原则》，阐述了 FAIR 基本原理，并介绍了一些已有实施案例，是首篇 FAIR 原则学术文献^[7]。Boeckhout 等（2018）指出 FAIR 原则为更系统地促进数据共享指明了前进的道路，有望成为生命科学研究的基础，但还要解决一些道德、方法和组织方面的挑战^[8]。De Smedt K 等（2020）认为 FAIR 数字对象（FDO）实现了更高级别的互操作性，建议欧洲开放科学云（European Open Science Cloud, EOSC）考虑使用 FDO 作为实现数据管理全球互操作性的基本机制^[9]。Koers H 等（2020）为协调实现 FAIR 数据生态系统向数据和基础设施服务提供商提出多项建议，如机构应通过建立数据管理计划（Data Management Plan, DMP）来支持 FAIR 意识和实施，为研究人员提供简单直观的培训^[10]。Annika Jacobsen 等（2020）对 FAIR 原则实施的注意事项做了说明^[11]。Pujol Priego L 等（2022）通过深入了解来自高能物理和分子生物学领域研究人员共享科学数据的方式，探讨科学数据共享过程中障碍和促进因素^[12]。

(2) FAIR 数据政策制定及其实施效果研究。随着 FAIR 原则的实践，国外高校、数据期刊、资助机构的数据管理政策向 FAIR 原则作出了调整，其中不乏一些优秀的案例。Coll I S 等（2019）指出 FAIR 原则政策涉及的内容有：机构如何通过将研究数据管理、共享和重用的支持和实践指南制度化来应对符合 FAIR 数据要求相关的众多问题^[13]。Jetten M 等（2019）对科研信息管理系统（Current Research Information Systems, CRIS）的价值进行论证并设计了调整方案，目的是在荷兰拉德布德大学实施 FAIR RDM 政策^[14]。Cannon M 等（2022）概述了从 2019 年 11 月开始，在 Taylor & Francis 出版的六种地球和环境科学期刊上实施 FAIR 数据共享政策的积极与消极影响，研究发现，引入数据共享政策对期刊投稿率、期刊的接受率以及同行评审的周转时间的影响比较小^[15]。

(3) 支撑数据管理 FAIR 化的技术平台基础。研究数据联盟可观察属性术语可互操作描述工作组（RDA I-ADOPTION WG）构建出 I-ADOPTION 互操作性框架，重点解释了 FAIR 语义组件，旨在为开发更加用户友好的数据注释工具提供语义基础^[16]。Wilcox D 等（2018）指出 Fedora 作为一个灵活、可扩展的开源存储库平台，在支持 FAIR 原则方面发挥着关键作用^[17]。Anjaria（2020）系统地分析了两个现有的数据管理平台 FAIRDOME 和 Dataverse，并基于 Petri 网络的形式化模型和 Petri 网络代数开发出数据管理模型——Datalection，帮助计算机系统设计师实现和分析 FAIR 原则^[18]。Wieser Florian 等（2021）描述了怎样建立本地网络基础设施以便于

数据管理 FAIR 化^[19]。Welsh Ciaran 等（2021）引入 libOmexMeta 用于生物模拟模型的语义注释来支持 FAIR 原则^[20]。Neumann Janna（2022）阐明 FAIR 数据基础设施的概念，认为当前研究数据管理、FAIR 数据管理、出版以及数据基础设施正在朝着更高的意识和科学界接受度方向迈进^[21]。

（4）FAIR 服务评估与 FAIR 数据激励机制研究。Wilkinson Mark D 等（2019）提出了一个可扩展、可自动化的框架来评估数字资源，其中包括可衡量 FAIR 成熟度指标、开源工具和参与指南，帮助数据管理员逐步实现数据资源 FAIR 化^[22]。Bloemers M 等（2020）针对研究资助组织（RFO）构建出 FAIR 资助模型，该模型与研究团体中的标准、元数据相结合，能自动评估研究数据管理的进展和产出^[23]。Devaraju A 等（2021）通过制定 17 个核心指标来支持对 FAIR 数据对象的系统评估，FAIR 指标可用于评估数据存储库是否在致力于实现 FAIR 数据对象，同时，他们还提供了一个开源工具（F-UJI）将指标应用于实践中。该工具及其基础指标是 Horizon 2020 FAIRsFAIR 项目的重要成果之一^[24]。

（5）FAIR 原则在各个科学领域的应用研究。国外作者还非常关注 FAIR 原则在不同领域的实际应用问题。John Wise 等（2019）认为实施 FAIR 原则能提高生物制药行业以及其他生命科学，如生物医学、环境、农业和食品生产的研发效率^[25]。Delgado Jaime 等（2020）探讨了将 FAIR 原则应用于基因组信息时的安全与隐私问题^[26]。Georgiana 等（2021）建议使用 FAIR 原则，结合区块链技术改进 B2B 流程^[27]。Jacobsson T Jesper 等（2022）在钙钛矿数据库项目中，为钙钛矿太阳能电池设备数据和交互式数据探索的可视化工具创建了一个开放访问数据库，根据 FAIR 数据原则制定程序和协议，以简化新钙钛矿数据的传播和收集^[28]。

（二）国内 FAIR 原则研究现状

2016 年，在杭州召开 G20 峰会，会议通过了《2016 年二十国集团创新行动计划》，该计划支持集团成员在 FAIR 原则指导下，探讨公共财政资助的开放科学研究成果和数据管理相关问题^[29]。国际科学理事会数据委员会（CODATA）于 2019 年 9 月召开开放科学数据政策与实践国际研讨会，《科学数据北京宣言》（*The Beijing Declaration on Research Data*）^[30]是该会议主要成果之一，北京宣言将全球数据同盟与开放数据的 FAIR 原则作为十条原则之一，指出应按照 FAIR 原则精神推动科学数据的开放和共享利用。笔者在 2022 年 3 月 20 日，以“FAIR 原则”OR“FAIR 数据”OR“FAIR 化”在中国知网上进行全文检索，通过手动筛选和剔除了会议、报告、与 FAIR 数据原则提出时间相差甚远的文献后，得到 227 条文献。我国的 FAIR 原则研究成果主要集中在 2017-2022 年，这与 FAIR 原则提出及其在国际上得到认可与应用的时间特征基本一致。经笔者深入分析，将国内学者对 FAIR 原则的相关研究从以下几个方面进行归纳：

(1) FAIR原则产生背景及主要内容介绍。宋佳等(2021)介绍了FAIR原则的产生背景、研究与实践进展,认为融合、智能和生态化的科学数据共享系统将成为主流趋势^[31];邢文明等(2021)对FAIR原则的内容及其相关要求进行了详细阐述,并考察了国外JRC-MRDI基础设施平台对FAIR原则的遵循情况^[32]。

(2) 实施FAIR原则的国外经验及其启示。翟军(2020)与邱春艳(2021)从内容、政策、实施框架、技术标准及实践路径等方面调查了欧盟开放科学数据FAIR原则落实情况^{[33][6]};刘凤红等(2021)总结和归纳国际出版集团如Springer Nature、Taylor&Francis、Wiley以及Elsevier推动数据开放共享和复用的政策、实践以及推行FAIR原则的具体举措^[34]。

(3) 不同学科领域FAIR原则的应用研究。段青玉等(2019)基于FAIR原则,调研了人文社科领域的荷兰国家级数据平台——Data Archiving and Networked Service (DANS)、法国超大型研究基础设施TGIR Human-Num、北京大学开放研究数据平台这三大代表性数据出版平台的FAIR原则实践,认为平台型数据出版至少需要在五个资源层面共同推进数据的可发现、可访问、可互操作与可重用^[35]。叶未央等(2021)研究发现,中医药古籍数字化出版存在盗版现象严重、维持运营成本大、古籍出版质量参差不齐等问题,并基于FAIR原则对古籍出版商关于中医药古籍数据库的建设提出建议^[36]。朱妍昕等(2022)基于FAIR原则,采用NeOn方法,构建了涵盖58个类的循证医学文献本体,并进行实例验证^[37]。秦顺等(2022)从DataCite收集到的128条数据为样本,梳理DMP与FAIR原则的融合实践^[38]。邢文明等(2022)对国内外知名高校图书馆推进FAIR原则实施的情况进行调查、归纳和分析,总结图书馆参与FAIR原则实施的有效途径^[39]。孙清玉等(2022)以河海大学为例,从系统架构、质量控制、服务体系等方面探讨了高校机构知识库在科学数据管理方面的应用以及科学数据FAIR原则实践^[40]。

1.2.2 《科学数据管理办法》研究现状

自《办法》出台以来,我国地方政府逐渐开始重视科学数据管理政策的价值和意义,陕西省、黑龙江省、重庆市、四川省等十余个省市积极响应《办法》的要求,陆续印发了本地区的科学数据管理实施细则。在科研院所中,中科院和农科院分别发布《中国科学院科学数据管理与开放共享办法(试行)》(2019年2月)、《中国农业科学院农业科学数据管理与开放共享办法》(2019年7月)。整体而言,《办法》在以政府部门为主导的层面已有所落实。近年来,国内学者从多角度对《办法》内容、实施情况及其指导意义展开研究。

(1) 《办法》内容体系解读及完善建议。张洋等(2019)^[41]从生命周期视角、邢文明等(2019)^[42]从生命周期与利益相关者二维视角、秦顺等(2019)^[43]从管理

体制及安全措施角度、温亮明等（2020）^[44]从政策文本量化视角、白锐等（2019）^[45]从政策议程角度、都平等（2022）^[46]从概念域视角，这些学者从不同角度对《办法》的结构、内容体系、语言、政策议程、数据集属性边界展开深入又全面的解读与分析。在探讨内容的同时，学者们还提出了及时且精准的优化建议，如盛小平等（2020）希望在科学数据开放共享政策中增加科学数据治理措施，为科学数据质量治理提供有效的行动指南^[47]；在《办法》中明确对科学数据质量的要求，为今后完善或修订科学数据管理办法提供参考^[48]。

（2）《办法》实施现状与落实策略研究。李洋等（2021）调研发现，在众多利益相关机构中，省级政府相关部门是落实《办法》的主力军，不同利益相关机构的落实程度存在不平衡和不充分的现象，整体而言《办法》落实的不太理想^[49]。温亮明等（2021）分析了区块链与《办法》的耦合性，指出在数据汇交与保存、数据保密与安全等各方面出现的《办法》落实难题，可以利用区块链核心技术解决^[50]。

（3）对科学数据开放与共享的指导意义。如结合《办法》相关规定，储文静等（2019）借鉴国外经验，设计出适用于我国高校层面的科学数据管理规范管理流程，包括科学数据采集、汇交、处理、公开和利用等内容^[51]。刘敏（2020）建议高校落实《办法》要求，确定符合自身建设目标的科学数据管理制度^[52]。

1.2.3 国内外研究现状述评

纵观国内外研究成果可发现，当前对 FAIR 原则的内容、提出背景及其重要意义的研究成果较为丰富，对 FAIR 原则在不同领域的实际应用进行了探索，这些研究对促进 FAIR 原则实践提供了较好的理论指导。国外早已将 FAIR 原则纳入科学研究战略与政策中，且目前国外研究已经侧重于 FAIR 基础设施、FAIR 工具、FAIR 服务评估和 FAIR 语义互操作性等实现数据 FAIR 化的研究上，努力构建 FAIR 数据生态系统。我国学者对 FAIR 原则的关注度也在持续上升，主要集中在 FAIR 原则的解读、国外 FAIR 经验介绍等理论研究上，对 FAIR 原则实践的研究较少，缺乏直接探讨 FAIR 原则在我国实施的路径研究，如制定 FAIR 实施指南、将 FAIR 原则理念纳入政策中，进而促进我国科学数据 FAIR 化。其原因可能是我国还未将 FAIR 原则纳入科学数据开放共享政策或 DMP 中，缺乏相关领域的 FAIR 实践，对 FAIR 数据价值的重视程度不够。本文基于国外 FAIR 数据政策基础，结合《办法》内容，提出促进我国科学数据 FAIR 化的实质性建议，丰富当前 FAIR 原则政策研究。

1.3 研究内容与方法

1.3.1 研究内容

本文的研究思路是基于 FAIR 政策与相关报告提炼出支持 FAIR 的政策要素，考察这些要素与《办法》的契合程度，提出基于 FAIR 原则的优化建议，以期推进我国科学数据开放共享工作可持续发展。论文总共包括六个部分，主要研究内容如下：

第1章 绪论。阐述本文研究背景与意义，梳理国内外研究现状，分析当前国内外研究特点与不足，概述本文研究内容、研究方法和创新之处。

第2章 FAIR 原则与《办法》内容及比较。对 FAIR 原则的背景与内容、《办法》中有关科学数据采集、汇交与保存、共享与利用、保密与安全等内容进行阐述与梳理后，对 FAIR 原则与《办法》进行多角度比较，以便展开后文的研究脉络。

第3章 FAIR 政策要素提取与确定。运用网络调查法和文献调查法，调研了国外 FAIR 政策与相关报告，选取合适的文本提炼出 FAIR 政策要素，咨询专家意见后，确定最终要素。

第4章 《办法》对 FAIR 原则的支持程度。确定《办法》对 FAIR 政策要素的支持等级，对支持情况进行逐一分析。

第5章 《办法》优化建议。提出优化《办法》的具体建议，为我国科学数据政策的完善以及优化科学数据开放共享工作提供参考依据。

第6章 总结与展望。对本文的研究内容和结论进行整理并总结，指出研究中的不足，对 FAIR 未来进行展望。

1.3.2 研究方法

(1) 文献研究法。为全面客观地了解国内外 FAIR 原则研究现状与《办法》研究现状，利用万方、维普、中国知网、ScienceDirect、SpringerLink、CRS 核心论文库等数字资源，解析国内外的相关研究理论和现状，阐述本研究的合理性和必要性。

(2) 网络调查法。本文的研究涉及到了国外开放科学数据 FAIR 原则的相关理论，FAIR 原则相关政策、我国《办法》落实现状等内容采用了网络调查法。

(3) 专家调查法。本文在对 FAIR 政策要素的确定处采用了专家调查法，向本研究领域的专家咨询意见，以确保每个要素的设置都是科学合理的。

(4) 比较研究法。将 FAIR 原则与《办法》进行对比，将 FAIR 政策要素与《办法》内容进行对比，从多个方面提出《办法》改进建议。

(5) 内容研究法。本文运用 NVivo 质性分析软件对国外 FAIR 政策内容进行编码和处理，从中提取 FAIR 政策要素。

1.4 研究创新点

其一，提炼 FAIR 政策要素。目前，国际上没有统一的 FAIR 政策模板，但在有些国家的 FAIR 政策与相关报告中明显或隐含了支持 FAIR 原则的政策要素，笔者从欧洲国家的 FAIR 政策文本和相关报告中提取支持 FAIR 原则的政策要素，得到 FAIR 政策要素初始框架，运用专家调查法对 FAIR 政策要素初始框架进行完善，确定 FAIR 政策要素。

其二，对《办法》的优化提出可行建议。在考察《办法》对 FAIR 原则支持度后，从适当引入 FAIR 原则理念、强化政策可操作性、为政策实施提供保障措施三方面提出基于 FAIR 原则的我国《办法》优化建议，以期为落实主体更好地贯彻执行政策提供一些实施方案，促进我国开放共享事业的繁荣发展。笔者建议可以适当在政策中融入 FAIR 原则中的内涵与要求，重视科学数据重用中元数据的价值，通过制定 DMP 和政策指南来强化政策的指导性与可操作性，同时加大财政支持、完善相关考核制度与监督机制来保障政策的贯彻落实。

第 2 章 FAIR 原则与《科学数据管理办法》内容及比较

2.1 FAIR 原则的产生背景与内容

2.1.1 FAIR 原则产生背景

数据密集型科学的一个重大挑战是通过帮助人类和机器发现、访问、集成和分析与任务相关的科学数据及其相关算法和工作流来促进知识发现。考虑到这一目标，2014年在荷兰莱顿举办了一个名为“共同设计数据博览会”（Jointly Designing a Data FAIRport）的交流会，本次会议汇聚了来自不同国家和不同研究领域的专家，共同探讨开放科学范围内研究数据的使用、处理和再利用。会议结束时提出了一套基本原则的制定草案，这些原则指出科学数据对所有研究对象都应该是可查找的、可访问的、可互操作的和可重用的（FAIR），无论是对机器还是对人。随后，由FORCE11社区的几名成员成立了一个专门的FAIR工作组对原则内容进行了微调和改进，将FAIR原则正式概念化，并提出在科学数据管理领域引入FAIR原则。2016年3月15日，FAIR原则正式发表在自然出版集团期刊《科学数据》（《Scientific Data》）上，这是FAIR原则首次出现在学术文献中。FAIR原则很快在全球范围内得到认可和采用，并作为数据管理和研究政策的基石。

2.1.2 FAIR 原则具体内容

FAIR是由4个原则的英文单词首字母组合而成的。FAIR原则包括可查找（Findable）、可访问（Accessible）、可互操作（Interoperable）和可重用（Reusable）4个基本原则和15条具体指导原则（如表2-1所示），每个细则相互关联、独立且可分离。

表 2-1 FAIR 原则主要内容

基本原则	细化条目
可发现 (Findable)	F1.数据（元数据）被设定全球唯一、持久的标识符
	F2.使用丰富的元数据描述数据
	F3.元数据清晰、准确地包含所描述数据的识别符
	F4.数据（元数据）在可检索系统中注册或索引
可获取 (Accessible)	A1.数据（元数据）可通过其标识符使用标准化通信协议进行检索
	A1.1协议是开放的、免费的、可普遍实施的
	A1.2协议允许在必要时进行身份验证和授权
	A2.即使数据已不再使用，元数据也是可访问的

续表 2-1 FAIR 原则主要内容

基本原则	细化条目
可互操作 (Interoperable)	I1.数据（元数据）使用一种正式的、可获取的、可共享且广泛应用的语言进行知识呈现
	I2.数据（元数据）使用遵循 FAIR 原则的受控词表
	I3.数据（元数据）包含对其他数据（元数据）的合法引用
可重用 (Reusable)	R1.数据（元数据）由一系列准确且相关的属性进行充分描述
	R1.1 数据（元数据）通过一个清晰且可访问的数据使用许可证进行发布
	R1.2 数据（元数据）应关联数据来源
	R1.3 数据（元数据）符合领域相关标准

(1) F (Findable) —可发现

这一原则要求为数据和元数据赋予永久标识符，并提供丰富元数据描述数据。这是确保数据实现共享与重用的前提。

F1: 数据（元数据）被设定全球唯一、持久的标识符

F1 指出，必须为数字资源（即数据和元数据）分配一个全球唯一且持久的标识符，以便计算机找到和解析。这是 FAIR 原则中最基本最重要的部分，全球唯一且持久的标识符是所有其他 FAIR 原则中的基本要素。F1 为数据标识符规定了两个条件：①它必须是全局唯一的（即其他人不能在不引用该数据的情况下重用/重新分配相同的标识符）；②它必须是持久的。注册表服务保证了标识符链接在未来的可解析性，即使该资源不再存在或移动。

F2.使用丰富的元数据描述数据

在创建 FAIR 数字资源时，元数据应包括有关数据背景、质量、条件或特征等描述性信息。丰富的元数据信息有助于计算机在无人干预的情况下自动完成目前需要研究人员大量关注的常规且繁琐性任务。这一原则背后的基本原理是，即便缺乏数据标识符，人们也能借助元数据信息轻松溯源数据。因此，遵守 F2 有助于人们找到数据，并增加重复使用和引用。

F3.元数据清晰、准确地包含所描述数据的识别符

元数据和它们描述的数据集通常是单独的文件。元数据文件与数据集之间的关联应通过在元数据中提及数据集的全局唯一和持久标识符来显式化。

F4. 数据（元数据）在可检索系统中注册或索引

仅实现 F1 和 F2 还不能确保数据能在互联网上被人们发现。高质量的数据资源可能只是因为没有被人们找到而影响其价值发挥。如果数据流、服务或存储库等数字资源的可用性未知，那么没有人（也没有机器）可以发现它。当前有许多方法可以发现数字资源，如搜索引擎，能很好的满足普通搜索者的基础信息需求，但对于

科研人员所需的大部分研究数据，则需要更精准地索引。原则 F1-F3 将为当前一些存储库和未来服务提供高精度索引所需的核心元素。

(2) A (Accessible) ——可访问

这一原则要求明确数字资源的访问方式，如进行身份验证和授权。

A1.数据（元数据）可通过其标识符使用标准化通信协议进行检索

识别数字资源的主要目的是同时提供使用某种明确定义的机制以某种格式检索该数字资源记录的能力：因此可检索性是 FAIR 可访问性的一个方面。

A1.1 协议是开放的、免费的、可普遍实施的

协议需是免费的（无成本的）和开放的（有源代码的），才能在全球范围内实施以促进数据检索，且任何拥有计算机和互联网连接的人都至少可以访问元数据，进而促进数据能被最大限度地重用^[32]。

A1.2 协议允许在必要时进行身份验证和授权

这一原则清楚地表明，FAIR 中的“A”并不意味着“开放”或“免费”。某些数字资源，例如基于道德、法律或合同约束而具有访问限制的数据，需要采取额外措施才能访问。这通常涉及确保访问请求者确实是请求者（身份验证），请求者的配置文件和凭据与资源的访问条件（授权）匹配，并且预期用途与允许使用情况相匹配（如仅用于非商业目的）。

A2.即使数据已不再使用，元数据也是可访问的

由于维护数据资源在线状态是需要成本的，长此以往，数据集会因为维护困难而降级或消失，这将导致人们会浪费很多时间在查找已失效的数据上。相较之下，存储元数据不仅成本低且维护简单。因此，原则 A2 规定，即使数据不再使用，元数据也应保留。A2 与 F4 中描述的注册和索引问题相关。

(3) I (Interoperable) ——可互操作

即数据需要与应用程序或工作流进行互操作以进行分析、存储和处理。

I1.数据（元数据）使用一种正式的、可获取的、可共享且广泛应用的语言进行知识呈现

FAIR 最终目标是优化数据的重用。为此，需要很好地描述元数据和数据，以便可以在不同的设置中复制/组合。

I2.数据（元数据）使用遵循 FAIR 原则的受控词表

用于描述数据集的受控词汇需要使用全局唯一和持久的标识符进行记录 and 解析，使用数据集的任何人都需要能够轻松找到和访问此文档^[32]。使用共享的、正式结构化的数据集是 FAIR 的重要组成部分。因此，用于数据（元数据）的词汇表本身也需要是可查找的、可访问的、可互操作的和可重用的，以使用户（包括计算机）能够完全理解元数据中使用的术语的含义。

13.数据（元数据）包含对其他数据（元数据）的合法引用

一般来说，数据或元数据不存在于孤岛中，必须采取必要的措施来确保代表资源的知识与其他资源的知识相关联，以创建一个有意义的相互关联的数据和服务网络。“合法引用”旨在（元）数据资源之间建立尽可能多的有用链接来充实数据的上下文内容，并与构建数据模型所花费的时间/精力相平衡。

（4）R（Reusable）——可重用

这一原则要求对数据引用来源进行标记与记录，保障数据能被机器可读，使数据能被重复利用。

R1 数据（元数据）由一系列准确且相关的属性进行详细描述

为了更简单与迅速的发现与重用数据，需要给数据赋予多个标签。R1 与 F2 非常相似，然而，原则 F2 背后的基本原理是实现有效的基于属性的搜索和查询（可查找性），而 R1 的重点是使机器和人类能够评估发现的资源是否适合重用，给定特定的任务。为此，数据发布者需要确保提供的元数据能充分描述数据来源信息且易于被机器和人找到，诸如实验数据、使用的仪器设备等信息。R1 指出数据发布者不应尝试预测数据使用者的身份和需求。

R1.1 数据（元数据）通过一个清晰且可访问的数据使用许可证进行发布

该项是关于法律互操作性的。数字资源及其元数据必须始终无一例外地包含一个许可证，描述资源可以在哪些条件下使用，即使这是“无条件的”。默认情况下，如果没有这种明确性，资源就不能合法使用，而无法找到许可证实际上与没有许可证相同。没有许可证并不表示“开放”，而是会产生法律不确定性，这将阻碍数据重用。

R1.2 数据（元数据）应关联数据来源

数据来源包括以下方面：资源是如何生成的，为什么生成资源，由谁生成，在什么条件下，使用什么原始数据或源资源，使用什么资金/资源，谁拥有数据以及生成后的应用处理过程。来源信息可帮助人员和机器评估资源是否符合其预期重用的标准，以及为了适当地重用资源可能需要哪些数据操作过程。

R1.3 数据（元数据）符合领域相关标准

对于类型不一、组织标准各异、文件存储格式不稳定以及不采用通用词汇的数据集重用起来比较困难。如果已有相关领域标准或存在数据存档、共享的良好例子，则可以优先选择遵循这些标准或参考这些优秀实践案例，以减少重复工作。例如，国外很多研究社区使用 MIAME、MIAPE 等作为社区内应遵守的最低限度信息标准。因此，在实施 FAIR 原则时也应该要求数据符合相关标准。

2.2 《科学数据管理办法》内容体系

在借鉴国内外先进经验的基础上，结合我国科学数据管理的实际情况，2018年3月17日，我国国务院办公厅正式印发《办法》。《办法》要求加强科学数据全生命周期管理，对科学数据的采集、汇交与保存、共享与利用、保密与安全等内容进行了规范；按照“开放为常态、不开放为例外”这一共享原则，突出科学数据开放共享是《办法》的一大亮点；平衡好对外交流过程中数据共享与国家秘密之间范围，保障我国数据安全；《办法》还明确了科技部、主管部门、法人单位和科学数据中心的主要职责。

2.2.1 科学数据的采集、汇交与保存

《办法》第十条至十八条对科学数据的采集、汇交与保存进行了规范。在采集和生成数据时应按照相关标准进行，由政府预算资金资助形成的科学数据应汇交到数据中心，对社会资金资助形成的不同性质的科学数据进行了不同规定。要求科技部整合国家科学数据中心，主管部门建立科学数据汇交管理制度，法人单位建立激励机制、科学数据质量控制体系和科学数据保存制度。

2.2.2 科学数据的共享与利用

《办法》第十九条至二十四条规定了科学数据共享与利用的相关事宜，遵从“开放为常态，不开放为例外”的原则。要求主管部门、法人单位将科学数据资源目录面向社会开放共享，鼓励为科学数据开展增值服务。要求对科学数据使用者注明数据使用来源。明确了法人单位无偿和有偿提供科学数据的情形。

2.2.3 科学数据的保密与安全

《办法》在第二十五条至第二十九条对科学数据保密与安全的大概范围进行了规定。对涉及国家秘密的科学数据不得对外开放，同时也要防止泄露和篡改。要求主管部门和法人单位建立相应的安全保密审查制度，数据中心应建立应急管理和容灾备份机制，完善科学数据安全防护体系。

2.3 FAIR 原则与《科学数据管理办法》的比较

2.3.1 制定背景与目标

(1) 制定背景。在科学数据开放共享的大背景下，各界人士在总结多年开放数据的经验后开始意识到要使数据易于被机器或人们发现、查找和互操作才能保障科学数据的有效共享与利用，因此协商形成具有适普性的科学数据管理 FAIR 规范。

我国也非常重视科学数据的共享与利用，从科技部资助大批科学数据共享项目到颁布一系列法规政策和领域标准再到建立国家数据共享平台和科学数据中心，在推动科学数据共享利用上的投入力度不断加大，2018 年我国首个国家层面的科学数据管理与共享法律规范《办法》的颁布更是有力地推动了我国科学数据的开放共享。

(2) 制定目标。管理和整合来源不同、标准不一、分布广泛的数据以推动科学数据的高效重用，这是科学界提出 FAIR 原则的主要目的。而《办法》的制定目的是“进一步加强和规范科学数据管理，保障科学数据安全，提高开放共享水平，更好支撑国家科技创新、经济社会发展和国家安全”。

2.3.2 内容与性质

(1) 内容侧重点。在内容上，FAIR 原则侧重于科学数据的重用，用四个主原则及若干子原则构成科学数据实现 FAIR 的标准。而《办法》内容不仅从数据生命周期的角度明确了科学数据管理与共享各环节任务，还从管理的角度明确了各责任相关者的职责^[42]，其内容侧重于科学数据的管理、安全与共享。

(2) 性质。一个是指导方针，一个是政策。通俗的说，FAIR 原则是各界人士通过广泛协商形成且被广泛认可的管理科学数据的一种手段、一套规范科学数据管理活动的指导准则。通过实现数据可操作、可获取、可互操作和可重用，来促进数据的共享与复用，推动科学数据开放共享愿景的实现。但其内容没有规定具体负责实施的主体，没有涉及 FAIR 原则实施的技术标准，也没有涉及实现科学数据 FAIR 化的基本步骤。而《办法》则是规范我国科学数据生命周期各阶段管理，保障科学数据安全，提高开放共享水平的国家层面政策，明确了不同责任主体在科学数据管理不同阶段的不同分工与责任。

2.3.3 实施重点

在具体的实施上，FAIR 原则注重“FAIR”这一目标的实现，即不论实施者采取何种方法和技术，旨在确保科学数据逐渐结构化，达到可查找、可访问、可互操作和可重复使用的目的，并尽可能向所有人开放。而《办法》则更加注重科学数据管理的过程，对于科学数据生命周期的各个阶段进行了详细的规定，如《办法》“共享与利用”一章中，涉及较多的是科学数据共享与利用过程中各个利益相关者的行为规范，对于如何实现科学数据的共享与利用的内容很少。

2.3.4 比较结果

由上述分析可知：(1) FAIR 原则内容侧重于科学数据的利用，但没有提及具体实施，《办法》注重科学数据的管理过程，但对科学数据的有效利用关注不够；(2) FAIR 原则与《办法》拥有同一愿景，即规范科学数据管理，促进科学数据共享利用。

第 3 章 FAIR 政策要素提取与确定

FAIR 不是标准，是指导方针。标准是规范性的，而指导方针是宽松的。它允许采用不同的方法来提供数据可查找、可访问、可互操作服务，以实现最终目标。可见，它并未涉及具体的实现技术、途径，以及具体的实施标准与规范，对 FAIR 原则的作任何解释或实施都可以，只要它们能产生机器可操作的结果，这就意味着各个利益相关者可以根据实际情况制定自己的 FAIR 实施方案，并且随着技术的发展，这些实施方案可以随着时间的推移而调整。因此，制定统一的 FAIR 数据政策或相应指南是利益相关者实施 FAIR 原则促进数据重用至关重要的一步。

FAIR 政策要素则是 FAIR 数据政策的浓缩与凝练，详细的政策要素可帮助政策制定者评估其数据政策是否支持 FAIR，并就如何与 FAIR 原则保持一致这一问题提供实质性的政策优化建议，为未来政策调整或制定 FAIR 数据政策/指南提供参考。因此，为了后续考察《办法》对 FAIR 原则的支持程度，必须先确定支持 FAIR 的政策要素，这正是本章将重点探讨的内容。目前，国际上没有统一的 FAIR 政策模板，但在有些国家的 FAIR 政策与相关报告中明显或隐含了支持 FAIR 的政策要素，笔者将通过这些文本提取支持 FAIR 原则的政策要素。拟从两方面获取 FAIR 政策要素，一是从欧洲国家的 FAIR 政策文本中抽取，二是从政策建议报告中筛选。将两者进行比较与取舍，初步得出 FAIR 政策要素框架，再运用专家调查法对 FAIR 政策要素框架进行完善，经统计分析后，确定最终的 FAIR 政策要素。

3.1 基于政策内容分析的 FAIR 政策要素提取

3.1.1 文本收集

早在 2014 年启动的“地平线 2020”（Horizon 2020）科技计划中，欧盟委员会就要求 Horizon 2020 资助的项目要按照 FAIR 原则进行数据管理，受其资助的欧洲众多科研机构积极参与到推动 FAIR 原则的政策制定与落实中。近年来，在资助机构与科研机构的影响下，一些在开放科学运动中较为活跃的国家纷纷将 FAIR 原则纳入其开放科学政策中，借助政策的强制性，使数据符合 FAIR 原则的要求，并尽可能向所有人开放。为了考察《办法》对 FAIR 原则的支持程度，本研究将选取这些国家级政策作为本节的研究对象。

本节选取政策文本的标准是：国家层面颁布的涉及如何实施 FAIR 原则的开放科学政策。主要搜集渠道及理由：一是 SPARC Europe 发布的《欧洲开放科学政策分析（第 7 版）》^[53]报告，该报告对 2021 年 4 月之前的欧洲开放科学国家政策进行了扫描，统计了涉及 FAIR 原则的国家政策；二是 EOSC 网站，它提供了欧洲国家

开放科学政策最新信息^[54]。笔者最终筛选出了 10 个符合标准的开放科学政策作为本节 FAIR 政策要素提取来源，如表 3-1 所示。

表 3-1 国外开放科学政策

编号	发布主体	政策文件	发布时间
1	荷兰教育、文化和科学部	国家开放科学计划（荷兰） ^[55]	2017 年 2 月
2	英国经济和社会研究理事会	ESRC 研究数据政策 ^[56]	2018 年 5 月
3	法国高等教育、研究和创新部	国家开放科学计划 ^[57]	2018 年 7 月
4	爱尔兰国家开放研究论坛	向开放研究环境过渡的国家框架 ^[58]	2019 年 7 月
5	芬兰开放科学协调组织	2020-2025 年开放科学研究（芬兰）宣言 ^[59]	2019 年 12 月
6	丹麦高等教育和科学部	丹麦 2020 年研究基础设施路线图 ^[60]	2020 年 2 月
7	意大利研究部	2021-2027 年国家研究计划 ^[61]	2021 年 2 月
8	拉脱维亚教育和科学部	拉脱维亚开放科学战略 2021-2027（草案） ^[62]	2021 年 8 月
9	瑞典研究委员会	FAIR 研究数据标准实施指南 ^[63]	2021 年 10 月
10	奥地利联邦事务部	奥地利开放科学政策 ^[64]	2022 年 2 月

本节研究选择借助 NVivo 质性分析软件对这 10 个政策文本进行节点编码，形成初始化概念和范畴，通过理清各范畴之间的关系得出高层次的主范畴，即支持 FAIR 原则的政策要素。

3.1.2 开放式编码（一级编码）

开放式编码也称一级编码，是将原始资料打散，赋予概念化标签，然后再以新的方式重新组合起来进而界定概念、发现范畴的操作过程^[65]。笔者按照“政策编号—政策编码次序”对原始材料进行区分，经过对比综合、整理归纳、筛除矛盾概念后，得到若干个初始化概念，再对其精炼概括后形成 27 个范畴，表 3-2 为得到的部分初始概念和若干范畴。

表 3-2 开放式编码范畴化（部分）

范畴	初始化概念	原始材料
支持建立数字存储库	建立安全的数据存储库	8-8 科学机构必须根据 FAIR 原则和 OAI-PMH 协议创建安全的研究数据存储库。

续表 3-2 开放式编码范畴化（部分）

范畴	初始化概念	原始材料
允许不共享数据	共享科学数据有限制	4-14 科学数据的共享可能会因正当理由而受到限制，例如商业利用、机密性、安全性、个人数据保护、项目目标的实现以及与研究结果的进一步利用不相容或其他规定的合法理由。
元数据描述要尽可能丰富	发布足够的元数据以促进数据重用	2-7 为了确保研究数据可以被其他人有效地重复使用，必须公开记录和发布足够的元数据，以确保研究数据既可发现，又可独立理解，无需求助于创建者。
要求元数据符合通用标准	遵循元数据标准	2-17 将数据存放到负责的存储库时，应根据该实体的标准并遵循数据文档倡议 (DDI)、SDMX 或 INSPIRE 等元数据标准，提供或创建标准化的结构化元数据记录，并解释目的，数据的来源、时间参考、地理位置、创建者、访问条件和使用条款。
要求数据引用并提供详细出处	提供数据来源信息	2-16 研究数据应附有高质量的文档和元数据，以便为二级用户提供基本信息，以独立理解数据、进行发现并允许科学重用。文件应至少描述数据的来源、实地工作和数据收集方法、处理和/或研究人员对数据的管理。应清楚地标记和描述变量或转录本等单个数据项。
要求数据管理符合 FAIR 原则	确保产生的数据符合 FAIR 原则	3-1 我们的目标是确保法国政府资助的研究产生的数据逐渐结构化，以符合 FAIR 数据原则（可查找、可访问、可互操作和可重复使用），并确保它们得到保存，并尽可能向所有人开放。
提交数据使用声明	要求受资助者提交数据共享声明	2-28 在 Je-S 提案的相关部分中包含关于数据共享的声明
说明不共享数据的理由	要求受资助者说明不共享数据的理由	2-29 如果数据共享被认为是不可能的，请提出强有力的论据来证明这一点
要求将数据存放在经认证的数字存储库中	默认将数据存放在安全存储库中	8-3 根据 FAIR 原则，公共资助研究项目的数据及其元数据必须默认存放在安全存储库中
为数据资源赋予 PID	以 PID 的形式为数据提供持久性信息	2-26 研究人员应向出版商（或负责的数字存储库）提交研究数据，这些出版商以持久标识符的形式为数据提供持久信息。出版商、负责的数字存储库和数据服务提供商必须为其保管的数据提供足够的持久性信息。

续表 3-2 开放式编码范畴化（部分）

范畴	初始化概念	原始材料
通信协议	使用通信协议检索数据	9-5 可以通过标准化通信协议检索数据和/或元数据。
提供支持、指导、建议、培训	培养研究人员技能和文化	8-9 为了培养研究人员在高质量研究数据管理方面的技能和文化，鼓励研究机构引入数据管家，培养研究人员的一般和行业特定数据管理技能，促进电子基础设施和工具的使用，帮助规划数据管理和完成 DMP，发布和存档研究数据集。
数字基础设施建设	数字基础设施	4-13 通过国家和国际数字基础设施，研究将在安全和可信的环境中变得并保持可查找、可访问、可互操作和可重复使用。
数据与研究成果可同时查找与访问	数据与研究成果可同时查找与访问	2-12 一般而言，支持已发表研究成果的数据应与已发表研究成果本身同时可找到和可访问。
数据保护	数据保护	2-4 访问级别是用于对数据进行分类以确定允许访问的条件的类别。数据根据其详细程度、敏感性和机密性进行分类。这种分类确保了适当的数据处理和访问保护措施到位。
设置禁锢期	实施禁锢期	2-23 如果需要延迟发布已保存的数据，以允许资助者公布其研究结果，则可以对数据实施禁锢期。禁锢期一般不超过 12 个月，从赠款结束算起，但视情况而定。
设立数据首席官 (CDO)	为研究数据设立数据首席官	3-4 高等教育、研究和创新部将为研究数据设立首席数据官 (CDO) 职位。CDO 将与法国政府的首席数据官合作，并监督相关机构的研究数据官网络。
明确基本原则	遵守“默认开放”原则	8-1 研究数据必须遵守“默认开放”原则，即研究数据必须是公开的或有正当理由的。不公开数据的原因可能是数据集包含敏感数据，出于法律或道德原因无法公开，或者数据集非常大且传播成本高昂。
伦理与道德	高道德标准与伦理审查	2-21 为了促进创新和高质量的研究，ESRC 要求其支持的研究将以高道德标准进行。ESRC 研究伦理框架对此提出了期望，并就如何以伦理的方式开展研究提供了有用的指南。ESRC 认识到，在一些新的研究领域，尤其是在公共领域使用或收集数据，或重新调整行政、交易和社交媒体数据的用途时，无法提前获得个人的知情同意；然后建议研究人员寻求数据共享批准，作为其伦理审查的一部分。

续表 3-2 开放式编码范畴化（部分）

范畴	初始化概念	原始材料
开放源代码软件和工具	鼓励开放源代码的 软件和工具	1-21 积极鼓励开放源代码的软件和工具，以便在未来记录和制作所有科学工作流程。
建立技能能力中心	建立高水平技能能 力中心	8-10 在某些领域（高性能计算、语言和量子技术），计划建立与研究数据管理密切相关的高水平卓越能力中心。在其他部门，也鼓励科研机构和政策制定者建立高水平的技能能力中心，任何科研机构的研究人员都可以将其用作资源。
监督与处罚	监督	6-2 丹麦高等教育与科学局在为国家研究基础设施提供新资助的谈判中，应特别要求通过新研究基础设施产生的研究数据符合 FAIR 原则。丹麦高等教育和科学署将监督这种合规性。
加强对 DMP 的要求	强制要求制定 DMP	8-2 强制要求研究人员为公共资助的研究计划中的所有项目制定 DMP。
鼓励合作与跨学科研究	支持跨科学项目	10-12 奥地利支持提高科学数据跨学科可用性的项目（避免数据孤岛）。
鼓励、认可、奖励、激励开放科学实践	奖励和激励	10-1 科学工作必须考虑到开放科学原则，并且必须通过适当的认可和奖励来促进。建立激励制度，鼓励科学家和公民公开开展日常工作。开放的科学实践应该映射到科学事业中，并让研究人员更加渴望。
共享元数据	共享元数据	4-15 共享元数据对于受限数据尤为重要，如果可能，应使其可访问。
资金与成本	投资成本	2-25 使用公共资金来支持公共资助产生的研究数据的管理和共享，包括在其资助计划中共享数据的数据管理和数据准备成本，以便数据可以重复使用。

3.1.3 主轴式编码（二级编码）

主轴式编码也称二级编码，在一级编码得到的若干范畴基础上，再进一步寻找能涵盖范畴之间相关关系的主范畴，即通过梳理各个范畴之间的类属关系，建立逻辑联系层，形成主范畴并对其进行深度分析。本文将部分范畴按照数据生命周期管理顺序进行排序，这些范畴与主范畴之间的类属划分参考了温芳芳（2019）^[66]、姜鑫（2020）^[67]的相关研究。最终总结出 8 个主范畴，具体主范畴、范畴内容、文献参考来源如表 3-3 所示。借助 NVivo 软件对原始材料进行编码，通过层次图表可以更加清晰直观地识别 FAIR 数据政策中的突出主题，编码项数越多，面积越大，在

FAIR 数据政策中被提及的频率越高，具体如图 3-2 所示。

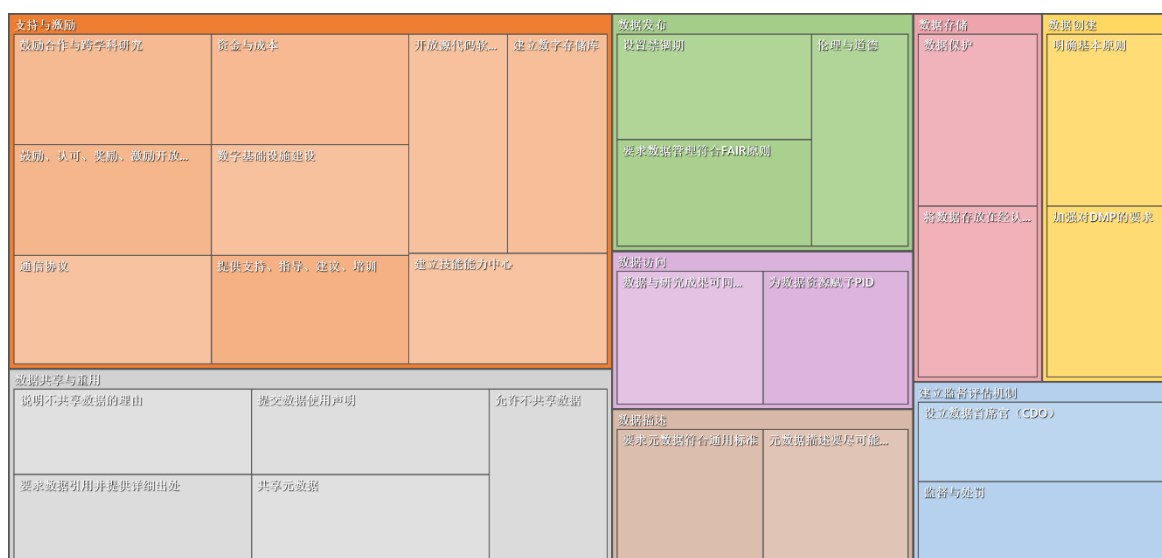


图 3-2 按编码项数比较节点

表 3-3 主轴式编码及其内涵

主范畴	范畴	范畴内涵	文献来源
数据创建	明确基本原则	要求在遵守开放科学数据基本原则的前提下，创建与收集数据。	温芳芳 (2019)、姜鑫 (2020)
	加强对 DMP 的要求	规定资助者和机构在资助条件和其他研究支持中应要求受资助者提交数据管理与共享计划。	
数据描述	要求元数据符合通用标准	规定数据需按照相关元数据标准生成，以确保数据在通用知识共享空间 (CC0、CC-by) 或同等许可下可用。	温芳芳 (2019)
	元数据描述要尽可能丰富	要以机器可读格式且用丰富的元数据描述数据。	
数据存储	将数据存放在经认证的数字存储库中	鼓励或要求研究人员将数据归档到经过认证的数据存储库中	姜鑫 (2020)
	数据保护	要求建立数据安全防护体系，防止数据泄露、病毒攻击等，如设置访问级别、身份验证。	
数据发布	要求数据管理符合 FAIR 原则	要求按照 FAIR 原则发布数据，确保公共资助产生的数据逐渐结构化，并在可能的情况下可供所有人访问。	温芳芳 (2019)、姜鑫 (2020)
	设置禁烟期	为允许数据延迟发布设置合理的时间期限。	
	伦理与道德	要求在发布和传播数据之前寻求获得知情同意，或者适当地匿名数据	

续表 3-3 主轴式编码及其内涵

主范畴	范畴	范畴内涵	文献来源
数据访问	数据与研究成 果可同时查找 与访问	规定描述数字对象或对象集合的元数据 应包含可搜索的全局唯一持久标识符。	姜鑫（2020）
	为数据资源赋 予 PID	规定描述数字对象或对象集合的元数据 应包含可搜索的全局唯一持久标识符。	
数据共享 与重用	允许不共享数 据	允许当涉及的数据是出于国家安全、国 防或公共卫生安全的原因被视为敏感的 数据时，可以不共享数据。	姜鑫（2020）
	要求数据引用 并提供详细出 处	规定必须通过正式引用来承认来源使用 的数据，以表示对产生、保存和共享关 键研究数据集的研究人员的智力贡献认 可。	
	共享元数据	要求元数据应共享，可被访问。	
	说明不共享数 据的理由	要求数据生产者在不共享数据时，应给 出详细说明。	
	提交数据使用 声明	要求提交 DMP 时应包括关于数据共享 的声明。	
监督与评 估	监督与处罚	要求有关部门监督政策的执行，对于未 按要求管理数据的情形实施处罚。	-
	设立数据首席 官（CDO）	要求设立 CDO 职位，监督相关机构的 科学数据管理。	-
支持与激 励	提供支持、指 导、建议、培 训	为科学数据管理和共享提供指导/建议/ 培训。	-
	鼓励、认可、 奖励、激励开 放科学实践	鼓励数据重用，为支持开放科学和研究 提供奖励/激励和服务。	-
	资金与成本	对共享数据的数据管理和数据准备成本 进行支持	-
	鼓励合作与跨 学科研究	鼓励利益相关者之间进行交流，鼓励参 与国际开放科学倡议，支持和鼓励跨学 科研究，以促进数据重用。	-
	建立数字存储 库	要求建立安全的科学数据存储库。	-
	开放源代码软 件和工具	鼓励开放源代码软件和工具，使所有数 据存储库都能被搜索、数据被链接等	-
数字基础设施 建设	要求创建高质量和可靠的数字基础设 施，为数据归档和重用提供支持服务。	-	

续表 3-3 主轴式编码及其内涵

主范畴	范畴	范畴内涵	文献来源
支持与激励	通信协议	表示可以通过标准化通信协议检索数据/元数据。	-
	建立技能能力中心	计划或鼓励在某些领域建立高水平的技能能力中心，为研究人员提供技能服务。	

3.1.4 选择性编码（三级编码）

选择式编码也称三级编码，是对上述开放式编码和主轴编码得到的主范畴、范畴的进一步分析，通过关系结构来表示主范畴与 FAIR 政策要素存在的关系。从表 3-4 可以看出，8 个主范畴均对支持 FAIR 原则有一定的影响和作用，因此可以在这些范畴中提炼出 FAIR 政策要素。其中数据创建、数据描述、数据存储、数据发布、数据访问、数据共享与重用是 FAIR 政策要素构建的核心要求，是科学数据管理与共享的主要内容。监督与评估、支持与激励是 FAIR 数据政策顺利实施的重要基础与动力，为科学数据管理与共享提供支持与服务是实施 FAIR 数据政策的基础保障。

表 3-4 主范畴典型关系结构

典型关系	关系结构	关系结构的内涵
数据创建-FAIR 政策	整体与部分关系	在 FAIR 数据政策中应明确基本原则，要求在数据创建初始阶段制定 DMP。
数据描述-FAIR 政策	整体与部分关系	在 FAIR 数据政策中应要求元数据符合通用标准，描述要尽可能丰富。
数据存储-FAIR 政策	整体与部分关系	在 FAIR 数据政策中应要求将数据存放在经认证的数字存储库中，以确保数据安全。
数据发布-FAIR 政策	整体与部分关系	在 FAIR 数据政策中需要数据管理符合 FAIR 原则、设置禁锢期，注意伦理与道德考虑。
数据访问-FAIR 政策	整体与部分关系	在 FAIR 数据政策中应要求为数据资源赋予 PID、数据与研究成果可同时查找，以确保数据可长期访问。
数据共享与重用-FAIR 政策	整体与部分关系	在 FAIR 数据政策中应要求共享元数据、允许不共享数据、说明不共享数据的理由、要求数据引用并提供详细出处、提供数据使用许可与共享声明
监督与评估-FAIR 政策	调节关系	明确监督与处罚、评估标准、设立数据首席官（CDO）等对 FAIR 数据政策的实施产生积极影响。

续表 3-4 主范畴典型关系结构

典型关系	关系结构	关系结构的内涵
支持与激励-FAIR 政策	调节关系	为科研人员数据管理提供支持、指导、建议、培训，鼓励、认可、奖励、激励开放科学实践等对 FAIR 数据政策的实施产生积极影响。

3.1.5 理论饱和度检验

为确保研究的范畴具有较好的信度和效度，需要进行理论饱和度检验。一般认为，当新收集的数据不再出现新的概念或范畴时即达到理论饱和^[68]。在完成开放式编码、主轴式编码和选择性编码后，随机抽取一个政策文本重新进行相同流程的编码，并没有发现新的范畴和关系结构，由此可以推断，已构建的理论研究通过了饱和度和检验。

3.2 基于政策建议报告的 FAIR 政策要素提取

3.2.1 政策建议报告选取

2016 年 8 月欧盟委员会成立“FAIR 数据专家组”（European Commission Expert Group on FAIR Data），以支持开放科学的研究和创新政策制定。2018 年发布“将 FAIR 变成现实”^[69]（Turning FAIR into Reality，简称 TFiR）报告，在定义 FAIR 概念、建设 FAIR 文化、创建 FAIR 数据技术生态系统、FAIR 能力建设、FAIR 指标衡量六个方面提出 34 项实施建议。TFiR 是欧盟委员会“FAIR 数据专家组”的最终报告和行动计划，为不同利益相关者（如政策制定者、研究资助者、科研机构、出版商等）提供了切实可行的建议，是诸多利益主体实施 FAIR 原则的战略方向和具体行动的主要来源。如 FAIR4Health 参考了 TFiR 中针对机构和研究团体的部分建议，制定了《在健康研究中实施 FAIR 开放数据政策的指南》^[70]。为了最大限度地减少重复工作，促进 TFiR 报告的实施，FAIRsFAIR 成立了 Synchronisation force 团队，发布《对 EOSC 的 FAIR 建议》^[71]白皮书，就如何实施 TFiR 报告中的建议提出了进一步落实方案。因此，笔者选取 TFiR 报告中针对政策制定者的相关建议作为本节 FAIR 数据政策要素提取来源。

3.2.2 FAIR 政策特征要求

这里借鉴 FAIRsFAIR 于 2019 年 11 月发布的“FAIR 政策现状扫描”^[72]（Fair Policy Landscape Analysis）报告中采用的政策分析方法。利用该分析方法基于 TFiR 报告中提出的关键政策建议，提炼出支持和促进 FAIR 原则的数据政策应具备的特

征和要求（部分示例见表 3-5）。

表 3-5 TFiR 相关建议与行动计划对科学数据政策的要求（示例）

TFiR 相关建议	TFiR 相关行动建议	对政策的要求	FAIR 政策要素
建议 1: 定义 FAIR 以便实施	行动建议 1.3: 应澄清 FAIR 和 Open（开放）之间的关系，因为这两个概念经常被混淆。FAIR 并不意味着开放，然而，在全球推动开放科学的背景下，使 FAIR 数据成为现实应得到需要适当开放和保护的政策的支持，可以表述为“尽可能开放，尽可能封闭”。	为了限制对数据的访问/获取，需要对数据共享给予合法例外；元数据应公开。	政策应对数据共享作出要求； 政策应允许数据共享的例外情况； 如果允许例外，应要求说明理由； 政策应要求元数据共享。
建议 3: 为 FAIR 生态系统开发组件	行动建议 3.2: 默认情况下，整个 FAIR 生态系统及其各个组成部分都应该为人类和机器工作。政策和 DMP 应该是机器可读和可操作的。	政策本身应该是人类和机器都可读的。	为政策赋予永久标识符（PID）； 政策应是机器可读的。
建议 16: 广泛应用 FAIR 原则	行动建议 16.1: 政策必须声明 FAIR 原则应适用于研究数据、元数据、代码、DMP 和其他相关数字对象，以及政策本身。	政策应明确涵盖包括数据在内的广泛的产出，这些产出期望都能得到管理与共享。	明确界定政策范围； 明确提及 FAIR 原则。

3.2.3 FAIR 政策要素筛选

笔者进一步根据上述要求提炼出 FAIR 政策的相关要素（见表 3-8）。由表 3-8 可知，基于 TFiR 报告，共提炼出 13 个 FAIR 政策要素。笔者将其划分为外部特征和内容要素两个维度，其中外部特征要素 3 个，包括：政策制定/更新日期；为政策赋予永久标识符（PID）；政策是机器可读的。内容要素 10 个，包括：明确提及 FAIR 原则；对数据进行明确定义；要求制定 DMP；要求数据共享；要求元数据共享；允许不共享数据；如果允许不共享数据，应要求说明理由；支持与 RDM 相关的、使数据 FAIR 化（显式或隐式）的合理成本；要求数据引用；提供指导、培训或支持以促进政策的实施。这些要素有助于将 FAIR 原则的要求融入政策，借助政策的强制约束力推动 FAIR 原则在实践中落实。

表 3-6 FAIR 政策要素及内涵

考察维度	FAIR 政策要素	要素内涵
外部特征	政策制定/更新日期	政策应明确包含制定/实施/生效/更新的日期，以及将要审查和更新的日期。
	为政策赋予永久标识符（PID）	为政策赋予唯一且永久标识符（如 DOI），以便查找、引用和追踪。
	政策是机器可读的	政策内容不仅应使用机器可读的格式（如 HTML 或 PDF），还应使用结构化数据标记模式 ^① ，以便于搜索引擎或程序能识别/理解其内容。
内容要素	明确提及 FAIR 原则	-
	对数据进行明确定义	对“数据”进行清晰定义，明确哪些科研产出适用于本政策（如数据、软件、算法、代码等）。
	要求制定 DMP	政策应要求制定 DMP。
	要求数据共享	政策对数据共享做出要求，数据共享应明确共享的许可、方式和范围。
	要求元数据共享	政策要求对元数据进行共享。
	允许不共享数据	对于涉及国家安全、个人信息、商业秘密，以及其他不适宜共享的数据，应允许其不予共享。
	如果允许不共享数据，应要求说明理由	如果认为数据不适宜共享，应提供正当合理的理由。
	支持与 RDM 相关的、使数据 FAIR 化（显式或隐式）的合理成本	为使科研数据 FAIR 化，需要付出额外的努力和时间精力，这些应得到资金支持。
要求数据引用	政策应明确要求数据使用者应进行引用。	
提供指导、培训或支持以促进政策的实施	为政策的实施提供相应支持。	

3.3 FAIR 政策要素的对比取舍及其初始框架

3.3.1 FAIR 政策要素对比取舍

基于质性文本分析国外数据政策内容获得了 27 个范畴，即 27 个 FAIR 政策要素，又基于 TFIR 报告，提炼出 13 个 FAIR 政策要素。笔者将两者进行对比分析，合并同类或相似的要素，并规范与优化了个别要素名称，相似要素及优化后的要素名称见表 3-7。经考虑，删掉的要素有：“数据保护”、“设立数据首席官（CDO）”、“通信协议”、“建立技能能力中心”。

^① 结构化数据标记模式（structured data markup schema）：指使用专门的格式进行标记、嵌入在 HTML 网页中、有着良好的结构且相互关联的数据。

表 3-7 FAIR 政策要素同类合并

基于政策内容的要素	基于 TFiR 报告的要素	合并后的要素概念
明确基本原则	明确提及 FAIR 原则	明确提及 FAIR 原则
加强对 DMP 的要求	要求制定 DMP	要求制定 DMP
为数据资源赋予 PID	为政策赋予永久标识符 (PID)	为数据及政策本身赋予 PID
允许不共享数据	允许不共享数据	允许不共享数据
共享元数据	要求元数据共享	要求共享元数据
说明不共享数据的理由	如果允许不共享数据, 应要求说明理由	如果允许不共享数据, 应要求说明理由
提供支持、指导、建议、培训	提供指导、培训或支持以促进政策的实施	提供指导、培训或支持
资金与成本	支持与 RDM 相关的、使数据 FAIR 化 (显式或隐式) 的合理成本	支持与 RDM 相关的、使数据 FAIR 化 (显式或隐式) 的合理成本

3.3.2 FAIR 政策要素初始框架

通过将两份政策要素进行对比与取舍后, 初步确定了 27 个 FAIR 政策要素, 划分为外部特征与内容要素两个维度, 其中外部特征要素有 3 个, 内容要素有 24 个, 详见表 3-8。

表 3-8 FAIR 政策要素初始框架

考察维度	FAIR 政策要素
外部特征	政策制定/更新日期
	政策是机器可读的
	为数据及政策本身赋予 PID
内容要素	对数据进行明确定义
	明确提及 FAIR 原则
	要求制定 DMP
	鼓励遵循通用元数据或领域标准
	元数据描述要尽可能丰富
	建议将数据存放在经认证的数字存储库中
	要求数据管理符合 FAIR 原则
	设置禁锢期
	伦理与道德

续表 3-8 FAIR 政策要素初始框架

考察维度	FAIR 政策要素
内容要素	数据与研究成果可同时查找与访问
	允许不共享数据
	要求数据引用并提供详细出处
	建议提交数据使用声明
	要求数据共享
	要求共享元数据
	如果允许不共享数据，应要求说明理由
	鼓励建立数字存储库
	鼓励合作与跨学科研究
	开放源代码软件和工具
	鼓励数字基础设施建设
	提供指导、培训或支持
	支持与 RDM 相关的、使数据 FAIR 化（显式或隐式）的合理成本
	鼓励、认可、奖励、激励开放科学实践
要求建立监督评估机制	

3.4 FAIR 政策要素最终确定

本研究选用专家调查法对初步形成的 FAIR 政策要素框架进行最终验证和确定，即基于专家调查法对选取的每个要素指标进行打分以表明专家对要素的认可程度，利用专家的判断减少个人判断的误差与不准确，进而确保最终确定的要素是科学合理的。

3.4.1 专家咨询表设计

调查问卷分为两部分。第一部分是要素重要性评价。该部分由专家为 2 个考察维度和初步提取的 27 个 FAIR 政策要素赋分，并给出认为该要素不重要的理由，也可以提出修改建议。问卷参考李克特五点量表，将重要程度分为 5 个等级：1 代表非常不重要、2 代表不太重要、3 代表一般重要、4 代表比较重要、5 代表非常重要，数字越大代表越重要。第二部分是专家自评对调查内容的熟悉程度和评判依据。该部分的设立了四个判断依据，分别是实践经验、理论分析、参考文献和直观选择，由专家评价这些依据对其做出本次判定的影响程度。两轮问卷详见附录 A 和附录 B。

3.4.2 专家的基本情况

由于本文的研究主题是科学数据管理与共享，因此笔者选择了对本领域熟悉程度较高的专家作为咨询对象，他们的研究领域及方向均涉及到了科学数据管理、科学数据开放共享或科学数据管理与共享政策。为保证所选专家的权威性和代表性，笔者选择的专家均已在本学科领域内的核心期刊发表过数篇重要文献。最终选取 32 名专家学者作为调查对象，共进行了两轮专家咨询。第一轮通过邮件的方式发放调查问卷 32 份，共收回 4 份调查问卷，专家积极系数为 12.5%。第二轮发放问卷 4 份，回收 4 份，专家积极系数为 100%。其中中国科学院大学 2 人，湖北大学 1 人，武汉科技大学 1 人；在学历方面，4 名专家均为博士学历；在职称方面，副教授 2 人，馆员 1 人。

3.4.3 调查结果量化分析

(1) 专家意见权威程度

专家的权威程度 (Cr) 由专家的熟悉程度系数 (Cs) 和判断依据系数 (Ca) 决定。其中专家判断依据 (Ca) 分为实践经验、理论分析、参考文献、直观选择，影响程度为大、中、小，具体赋值见表 3-9。专家对调查内容的熟悉程度 (Cs) 分为不熟悉、不太熟悉、一般熟悉、比较熟悉、非常熟悉，分别赋值为 0、0.2、0.5、0.8 和 1 分。一般来说，当 $Cr > 0.7$ 时，表明专家意见权威程度较好。由表 3-10 可知，两轮专家权威系数分别为 0.85 和 0.86，都高于 0.7，表明参与此次调查的专家具有良好的代表性和权威性，调查结果可靠性较高。

表 3-9 专家权威程度赋值表

判断依据	对判断专家的影响程度		
	大	中	小
理论分析	0.3	0.2	0.1
实践经验	0.5	0.4	0.3
参考文献	0.1	0.1	0.1
直观选择	0.1	0.1	0.1

表 3-10 专家权威程度统计表

系数	第一轮	第二轮
判断依据系数 Ca	0.85	0.875
熟悉程度系数 Cs	0.85	0.85
专家权威系数 Cr	0.85	0.8625

(2) 专家意见协调程度

本研究采用 Kendall 协同系数 (W) 来反应专家意见的协调程度。一般认为, 当 W 的显著性检验结果 $P < 0.05$ 时, 说明专家意见协调度较好, 评分结果可信, 反之则一致性越差。

利用 SPSS25.0 进行分析, 两轮专家咨询意见协调程度如表 3-11 所示, 协调系数分别为 0.302 和 0.260, 没有过半, 说明专家的意见不一致, 存在分歧, 而两轮调查的 P 值均为 0.00, 小于 0.05, 表明研究具有统计学意义, 有很好的可信度。

表 3-11 两轮专家意见协调程度表

咨询轮次	Kendall's W	卡方 (χ^2)	渐近显著性 (P)
第一轮	0.302	26.271	0.000
第二轮	0.260	24.178	0.000

3.4.4 FAIR 政策要素修改

将专家打分结果进行整理分析, 均值数值越高表明专家认可度越高, 反之越低。变异系数数值越低表示专家意见协调度越大, 即要素指标的选取更具有合理性。本研究以重要性赋值算术平均数大于 3.5 且变异系数小于 0.25 为筛选标准^[73], 充分考虑专家意见和统计结果, 进行 FAIR 要素的筛选。

第一轮专家咨询: ①增加: 新增“要素类别”, 即将 21 个“内容要素”细分为数据管理、数据共享、监督与激励。②修改: 将“鼓励数字基础设施建设”与“鼓励建立数字存储库”两个要素合并为“鼓励建立统一的数字存储库并定期汇交”。③删除: 无。经统计, “政策是机器可读的”和“支持与 RDM 相关的、使数据 FAIR 化 (显式或隐式) 的合理成本”这 2 个要素的均值和变异系数都不符合筛选标准 (见表 3-12), 对于是否予以删除, 将在下一轮继续征集专家意见。

第二轮专家咨询: ①增加: 无。②修改: 将 FAIR 要素“明确提及 FAIR 原则”改为“提及 FAIR 或其相关子原则”。将“建议将数据存放在经认证的数字存储库中”和“伦理与道德”调整到“监督与激励”要素类别中, “设置禁锢期”、“数据与研究成果可同时查找与访问”和“开放源代码软件和工具”调整到“数据共享”要素类别中。补充 PID 等英文缩写的中文全称。③删除: 无。对于第一轮中是否建议删除的两个要素, 有两位专家给出了意见, 一位专家建议均保留, 一位专家认为“政策是机器可读的”可有可无, 建议删除“支持与 RDM 相关的、使数据 FAIR 化 (显式或隐式) 的合理成本”这一要素。由于只有一位专家给出删除意见, 另外在第二轮调查数据中这两个要素的评分均值和变异系数均已符合筛选标准, 因此笔者选择保留这两个要素。

从表 3-13 中可以看出, 所有评分内容的均值都是大于 3.5 的, 说明专家对于选

取的各要素指标均有较高的认可度；而在专家意见协调度方面，每个指标的变异系数数值都是小于 0.25 的，说明专家对于选取的要素指标意见分歧较小。因此，评分结果可取，可以直接结束调研，确定最终的 FAIR 要素。

表 3-12 第一轮专家调查数据分析表

评分内容	重要性均值	标准差	变异系数	评分内容	重要性均值	标准差	变异系数
外部特征	3.75	0.83	0.221	允许不共享数据	4.5	0.50	0.111
内容要素	5	0.00	0.000	要求数据引用并提供详细出处	4.75	0.43	0.091
政策制定/更新日期	4.5	0.50	0.111	建议提交数据使用声明	4.25	0.83	0.195
政策是机器可读的	2.75	1.09	0.396	要求数据共享	4.5	0.87	0.192
为数据及政策赋予 PID	3.75	1.30	0.346	要求共享元数据	4.75	0.43	0.091
对数据进行明确定义	5	0.00	0.000	如果允许不共享数据，应要求说明理由	4.5	0.50	0.111
明确提及 FAIR 原则	3.75	0.83	0.221	鼓励建立数字存储库	3.75	1.30	0.346
要求制定 DMP	4	1.00	0.250	鼓励合作与跨学科研究	4	1.00	0.250
鼓励遵循通用元数据或领域标准	4.5	0.50	0.111	开放源代码软件和工具	4	1.22	0.306
元数据描述要尽可能丰富	4.25	0.83	0.195	鼓励数字基础设施建设	3.75	0.83	0.221
建议将数据存放在经认证的数字存储库中	4.5	0.87	0.192	提供指导、培训或支持	3.5	0.87	0.247
要求数据管理符合 FAIR 原则	4.5	0.50	0.111	支持与 RDM 相关的、使数据 FAIR 化（显式或隐式）的合理成本	3.25	0.83	0.255
设置禁锢期	3.75	0.83	0.221	鼓励、认可、奖励、激励开放科学实践	4.75	0.43	0.091
伦理与道德	4.25	0.83	0.195	要求建立监督评估机制	4.5	0.50	0.111
数据与研究成果可同时查找与访问	4.75	0.43	0.091				

表 3-13 第二轮专家调查数据分析表

评分内容	重要性均值	标准差	变异系数	评分内容	重要性均值	标准差	变异系数
外部特征	4.4	0.42	0.096	伦理与道德	4.50	0.50	0.111
内容要素	4.95	0.09	0.017	数据与研究成果可同时查找与访问	4.23	0.39	0.092
数据管理	4.98	0.04	0.009	允许不共享数据	4.50	0.50	0.111
数据共享	4.90	0.17	0.035	要求数据引用并提供详细出处	4.73	0.42	0.089
监督与激励	4.33	0.41	0.094	建议提交数据使用声明	4.35	0.41	0.094
政策制定/更新日期	4.45	0.46	0.102	要求数据共享	4.70	0.41	0.088
政策是机器可读的	4.00	0.71	0.177	要求共享元数据	4.73	0.42	0.089
为数据及政策本身赋予 PID	4.08	0.13	0.032	如果允许不共享数据，应要求说明理由	4.45	0.46	0.102
对数据进行明确定义	4.75	0.43	0.091	鼓励建立统一的数字存储库并定期汇交	4.38	0.41	0.095
明确提及 FAIR 原则	4.25	0.43	0.102	鼓励合作与跨学科研究	4.23	0.81	0.191
要求制定 DMP	4.43	0.44	0.099	开放源代码软件和工具	4.20	0.79	0.187
鼓励遵循通用元数据或领域标准	4.68	0.41	0.087	提供指导、培训或支持	3.55	0.55	0.156
元数据描述要尽可能丰富	4.70	0.41	0.088	支持与 RDM 相关的、使数据 FAIR 化（显式或隐式）的合理成本	3.75	0.83	0.221
建议将数据存放在经认证的数字存储库中	4.95	0.09	0.017	鼓励、认可、奖励、激励开放科学实践	4.08	0.72	0.176
要求数据管理符合 FAIR 原则	4.38	0.82	0.187	要求建立监督评估机制	4.43	0.44	0.099
设置禁铜期	3.75	0.43	0.115				

3.5 FAIR 政策要素描述

通过两轮专家调查，对 FAIR 政策要素初始框架进行了优化和完善，最终形成了如表 3-14 所示的 FAIR 政策要素最终框架。共 26 个 FAIR 政策要素，其中外部特

征 3 个，内容要素 23 个，且内容要素又细分为数据管理、数据共享和监督与激励三种类别。以下将对每个要素的内涵进行描述。

表 3-14 FAIR 政策要素最终框架

考察维度	要素类别	政策要素
外部特征	-	政策制定/更新日期
		政策是机器可读的 为数据及政策本身赋予永久标识符（PID）
内容要素	数据管理	对数据进行明确定义
		提及 FAIR 或其相关子原则
		要求制定数据管理计划（DMP）
		鼓励遵循通用元数据或领域标准
		元数据描述要尽可能丰富
	数据共享	要求数据管理符合 FAIR 原则
		数据与研究成果可同时查找与访问
		设置禁锢期
		允许不共享数据
		要求数据引用并提供详细出处
监督与激励	建议提交数据使用声明	
	要求数据共享	
	要求共享元数据	
	如果允许不共享数据，应要求说明理由	
	开放源代码软件和工具	
监督与激励	伦理与道德	
	建议将数据存放在经认证的数字存储库中	
	鼓励建立统一的数字存储库并定期汇交	
	鼓励合作与跨学科研究	
	提供指导、培训或支持	
监督与激励	支持与科研数据管理（RDM）相关的、使数据 FAIR 化（显式或隐式）的合理成本	
	鼓励、认可、奖励、激励开放科学实践	
	要求建立监督评估机制	

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/238107126045006026>