

中国科学院文献情报中心科技信息政策中心
系列编译资料

科研数据管理服务的需求探索方法

**How to Discover Requirements for Research Data
Management Services**

原编著者：D. Scott Brandt, Susan Wells Parham, Sherry Lake

2014年8月



BY NC ND 本作品采用[知识共享署名-非商业性使用-禁止演绎 3.0 中国大陆许可协议](https://creativecommons.org/licenses/by-nc-nd/3.0/)进行许可。

使用须知

中国科学院文献情报中心为促进学术交流、促进文献情报服务创新发展，特组织编译《科研数据管理服务的需求探索方法》，供个人学习和研究使用。文献情报机构可以在保证《科研数据管理服务的需求探索方法》完整性和所有编译者信息完整准确性的条件下在网站上整期上载和传播《科研数据管理服务的需求探索方法》的 PDF 版本，并明确说明来源。

任何机构或个人在引用《科研数据管理服务的需求探索方法》内的具体编译内容时，请按照学术规范注明来源，包括原始文献的著者、题名和来源网址等，也包括编译者姓名和编译内容来源。如果任何机构或个人要直接整条采用具体编译内容（包括仅对文字进行非实质调整后的采用）、或者要对较长编译内容直接采用其较大篇幅内容（例如超过五百字以上），应事先征得编译者的同意。任何机构和个人，未经中国科学院文献情报中心许可，不能直接把《科研数据管理服务的需求探索方法》的内容大规模直接编撰为新的作品或作品的一部分。

编译者：张瑶

审校者：丁瑾一校，顾立平二校

审核者：顾立平



本作品采用[知识共享署名-非商业性使用-禁止演绎 3.0 中国大陆许可协议](https://creativecommons.org/licenses/by-nc-nd/3.0/)进行许可。

科研数据管理服务的需求探索方法

1.前言	4
2.数据管理角色和职责	6
3.放在科研环境中的科研数据管理	8
3.1 将需要什么服务?	8
3.2 研究思路、工具与设施	9
3.3 利益相关者：生产者、使用者和决策者	10
3.4 机构规则和研究规范：为什么要改变?	12
4.发展阶段	13
4.1 设想	13
4.2 启动	14
4.3 探索	15
4.4 设计	16
4.5 执行	17
4.6 评价	18
5.启动与发现需求	19
5.1 深度个案研究	19
5.2 调研数据实践和标准服务的能力	19
5.3 利用数据保存用户文档使数据生命周期文档化	25
5.4 利益相关者用户文档、角色和方案	27
5.5 发展研讨会——从编程马拉松到混搭	30
6.下一步计划和未来的挑战	31
6.1 管理要求	31
6.2 规模和复杂性的挑战	32



科研数据管理服务的需求探索方法

1.前言

本指南主要适用于高等教育机构员工或者跨机构合作项目的工作人员,其工作职责是开发服务或工具来支持科研数据管理(Research Data Management, 简称 RDM)和数字保存。您的 RDM 开发工作可能嵌入并且受到科研群体的关注,或者是进行更高水平的合作,如图书馆或者计算服务。您需要一个系统的方法去计划、引出、分析、证明和优先考虑各种用户的需求。这类“需求发现”在本指南中,与 RDM 服务开发过程相联系。而数字内容管理中心(Digital Curation Centre, 简称 DCC)的指南《如何发现科研数据管理服务的需求》详细地描述了这些服务。更多的资料可从 DCC 和 DataONE 网站下载,也可从本指南最后列出的参考文献中获取。

本指南的第 2 节总结了数据管理工作的角色和职责,并介绍了运用系统的方法去满足数据管理需求是有价值的。有效的科研数据管理涉及许多人员、支持技术和组织,包括协调人力资源和财政资源。在某种程度上,支持服务需要集成他们的信息系统。研究项目需要系统帮助管理项目生命周期内的数据。一旦数据创造者的需求得到满足,其他人员和系统可能参与数据存档和保存,来满足后续用户和重用的需求。

非学术利益相关者也有相关需求,例如任何外部合作伙伴或者研究参与者,他们也会使用或促进数据产生。其他机构和商业服务提供商提供的服务也扮演着重要的角色。例如满足个人研究者对知识库或档案的长期保存需求,为他们的学科提供最佳的服务,并确保以上这些服务也符合资助主体和机构的期望。

在第 3 节中我们考虑了 RDM 需求的背景。我们从通常所需要的高级服务模式出发,然后我们考虑为什么研究背景与其他地区的业务流程变革不同,以及确定需求有什么样的挑战。为了促进多样性和复杂性的研究工作,我们提出了可能符合要求的三个关键领域:

- 思路、工具和设施: 你需要确认什么样的信息被科研人员认为是“数据”,考虑其他记录或信息的隐含关系,并且它们采取什么样的呈现形式。考虑一下科研人员需要多大程度的帮助以完成工作,以及他们是否使用集中的设施或标准化的协议。科研人员在其所属领域进行更广泛研究的任何驱动因素,都可能会增加他们对 RDM 的需要,而科研过程中这些技术具有不确定性使 RDM 更具有挑战性。



•科研利益相关者：生产者、使用者和决策者；在本机构或跨机构发展任何 RDM 服务都要对利益相关者进行分析，减少因部分团体或服务提供者没有正确评估所带来的风险，他们通常包括跨越组织边界的工作角色，比如学科馆员或科研计算器员工。

•机构原则和研究规范：如果利益相关者购买了本指南，需要制定指导方针，以阐明变更所带来的好处和风险。这尤其适用于涉及保存和共享数据的决策。最大的好处是在项目期间提供技术支持，促进科研人员共享的数据得到承认，有助于决定保存数据的类型，并支持存储和使用数据目录。

RDM 开发周期可包括六个循环的阶段：设想、启动、探索、设计、实施和评估。图 1 说明了前五个核心活动的背景。第六阶段“评估”单独显示，因为它往往涉及不同的因素，但是其“评估”的结果和这些不同因素也适用整个开发过程。

在本指南的第 4 部分，概述 RDM 发展的关键阶段和这些关键阶段中涉及的活动。在第 5 部分中，描述了支持启动阶段改变和发现需求的工具。本指南介绍了一个广泛的策略并进一步描述服务的实现过程。采用常用的技术收集信息需求，如调查、访谈和小组讨论。我们需要考虑这些技术怎样适应服务的发展，以及研究背景的各个方面。



图 1 科研数据管理服务的开发过程



在第 5 部分中，我们也确认了常见的 RDM 需求方法的主要因素。这些包括：

- 进行深度案例研究
- 调查数据实践和标准服务能力，使用数据资产框架，科研数据基础设施和目标合作评估工具，或数据管理必要条件（DMVitals）等。

档案化数据生命周期与数据监管

- 利益相关者用户文档、型人和情境
- 开展研讨会活动，如编程马拉松（hackdays）和混搭（mashup）

最后，在第 6 节我们考虑数据管理需求的下一步计划，以及 RDM 需求未来的挑战范围。

DCC 是由英国联合信息系统委员会资助的，用于在英国高等教育部门构建良好的数据管理实践能力。DCC 提供协调培训，并支持机构运行，使知识和实践实现最佳转移。本指南涉及英国大学参加的 DCC 项目（其与相关项目一起开发 RDM 服务）。

DataONE 是一个由美国国家科学基金会 (National Science Foundation，简称 NSF) 资助的基础设施网络。它旨在促进生物和环境数据的发现，访问分布式知识库，为科学家提供一组集成的工具，支持数据生命周期的所有阶段，包括数据管理和内容管理。

2. 数据管理角色和职责

资助机构和大学等机构政策制定者越来越重视科研数据管理和数字内容管理，对机构和科研人员赋予的新职责。确定他们的需求范围可能比较复杂，并且利益相关者可能不熟悉数据管理和数字内容管理术语。因为他们的需求与预期会改变，需求管理过程应解决以下挑战：

- **数据政策：** 出版商和融资机构要求科研人员使数据可公开访问和重用。科研数据政策继续有助于开发有效的服务和基础设施。

- **科研的不同利益相关者：** 包括数据生产者、研究资助者、参与者和用户，其中任何一个人都可能影响数据管理需求，研究项目中后期，依据他们的需求产生的数据是可重用的。

- **数据密集型的研究范式：** 新的数字数据来源和建模技术可供集成新的和已归档的数据集，寻求集成模式可以生成新的假说和理论。这种“归纳推理”方法的使用，对科学来说是一种新颖的方法，在新兴生物科学领域较为常见。有效的数据管理和数字内容管理是数据密集型领域的关键，需求分析需要考虑他们需要的专业数据管理工具和平台。

上述所有基本上都需要优先考虑依据预期收益和预期成本得出的需求。发展 RDM 基础设施的很大一部分成本，是解决上述问题的“启动”成本：发展政策、捕获需求、实施调查方案——规模够大，经验够多时，成本将缩小。因此这些问题都不应阻碍进步，他们应该有条



不紊的识别需求、管理需求变化。

资助机构在一个广泛的层面上定义了数据管理的责任。例如英国研究理事会（Research Councils UK，简称 RCUK）对数据政策制定共同原则。个人研究委员会采取自己的数据政策，明确职责确保科研人员个人和机构之间共享成果。

大多数大学和其他研究机构也已经开发了一个出版物的知识库。为了方便访问科研数据，他们中的许多现在需要提供一个“核心”数据成果的知识库，帮助科研人员确保给别人提供更多的有价值的数字数据，通常认为这是有价值的资产与服务。积极开发的另一个领域是信息系统研究——管理科研产出的信息，并与其他信息管理功能集成。

美国国家科学委员会的一份报告《长久生命力的期数字数据集成：促进 21 世纪研究和教育》：给个人和机构一个广泛的的角色和责任，促进数据访问。资助机构的报告确定了角色，例如创建一种文化，尽可能多的像关注期刊一样关注数字数据；鼓励创建易访问的数字共享，支持规范和标准的发展，并提供资源和监管。

报告还确定了数据作者和用户的职责(见第 3 节的科研利益相关者)以及主要参与人员：

- 数据管理：负责数据库/知识库的操作和维护的组织和数据科学家
- 数据科学家：信息和计算机科学家、数据库和软件工程师和程序员、学科专家、管理者和专家注解者、图书馆员、档案和其他数字集成成功管理的关键因素。

数据管理员通过一个机构库或主题档案馆向科研人员提供支持。他们通常要具备软件开发能力或访问数据的能力，以分析需求和选择支持技术。数据密集型领域可能要求对当前的实践进行详细分析，涉及到有关团体的数据科学家。这种分析可能还包括电子研究或基础设施的专家。

机构内开发基础设施项目，通常会涉及图书馆的研究支持部门、计算服务和研究管理。所涉及的信息专业人员可能来自不同的背景，而且非技术问题同技术问题一样重要。任何组织变革的过程可能产生新的政策发展和培训需求，科研数据管理也不例外。这可能还需要信息治理或档案管理单位，或知识产权管理的专家(其作用和电子研究或科学计算是一样重要)的支持。学术联络或学科馆员角色可能会涉及一些服务角色。

除了个人研究机构，还存在许多基于主题的知识库，并且有一些资金机构对特定的数据中心提供资助。一些研究小组可能与学科电子研究基础设施或网络基础设施保持一致，如 DataONE。有其他团体参与协作，标准和工作流机构包括国家科学院、学术团体和专业机构。



3.放在科研环境中的科研数据管理

3.1 将需要什么服务？

任何开发项目都需要根据其用户和利益相关者及他们的组织目标，确定他们需要的服务。存在一些关于支持服务范围的共识，而它们通常来自科研机构，这有助于制定选项。

下图 2 给出了一个高层次视图。

服务将遵循研究周期的典型阶段，从规划要使用的数据到存储、出版它。所有这些都需利用到指令或者软件工具所形式的在线支持，并且，根据需求和负担能力的，需要，由某些人进行指导和培训。

在大学里开展科研数据管理服务，在某些方面就像进行业务流程重组（BPR）。例如高级管理人员必须支持它，客户需求必须得到满足，内部利益相关者的买入是必不可少的。RDM 和 BPR 一样，它的输出通常包括工具，用于转换数字数据和数据产生过程中的工作流程。而且，文化变迁对成功的重要程度与提供增强功能对成功的重要程度一样。

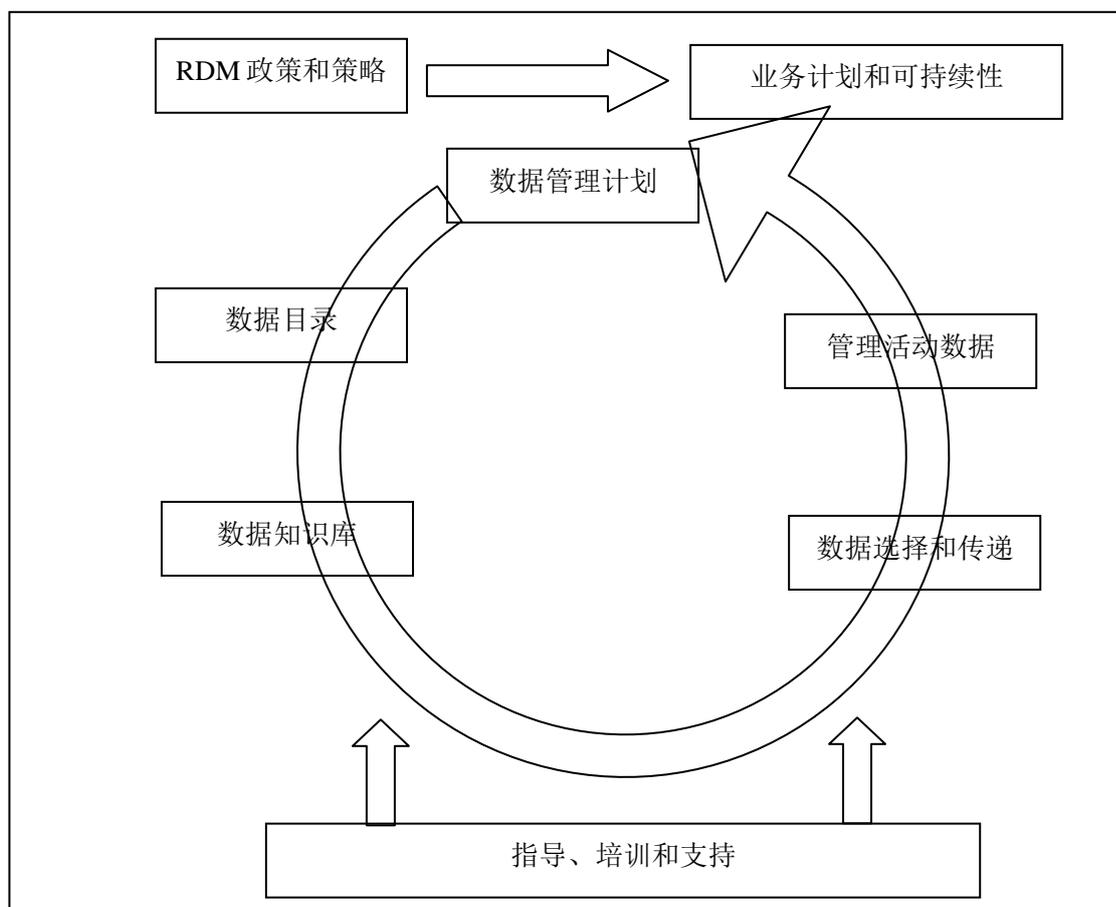


图 2 科研数据管理支持服务概述



我们还需要考虑 RDM 情景和业务流程变化之间的重要差异，这些是目前服务发展所面临的挑战。科研环境背景和业务流程不同，因为它们通常更复杂——即所涉及的活动是很难的、不确定的，高度依赖于相互依存性和标准化水平的。科研领域具有模糊性和不断变化的界限，通过非正式网络组织起来，并依据允许学术独立的规范，追求高风险想法，抵制制度控制。

迎合研究环境会限制引入标准化“企业”解决办法的范围，因为这些可能无法满足研究的需要，以探索处理数据的新方法。

设计跨学科 RDM 基础设施是一种相对较新的挑战，但对电子基础设施的研究显示，就像其他形式的共享资源制度设计一样，需要重视当前规范和日常实践原则、功能、服务组件相关的“互动模式”，以及资助者和其他利益相关者所设想的结果。规范可以以“理所当然的”方式渗入日常实践研究中。当新的流程威胁到它所服务的研究的进程时，只有假设清晰，才能避免代价高昂的变动，所以，在设计早期识别它们，是有意义的。

理想情况下，信息专业人员应该与研究团体共同设计，即与科研人员密切合作，以了解他们如何看待需求，并使他们积极参与，形成满足这些需求的解决方案。如果可行的话，RDM 应建立在他们已经使用的平台上，以最小化努力实现数据共享——例如，通过从现有的工作流程自动提取背景信息和元数据。

3.2 研究思路、工具与设施

科研数据的一个定义是 “……信息的主要构建块包括最低级别抽象任何领域的知识，它是可识别的数字、字符、图像的集合，或以某种方式语境化的其它象征（将事实、数字或想法表示为传染病信息）。这意味着，什么算作研究领域的“数据”取决于对它是什么及其代表意义的共识。这反过来可能取决于该领域中的思想、工具和设施（用于测试给定数据集是什么及其代表意义）。

资助主体将科研数据归为“公共利益”。为了履行这一职责，科研数据需要与足够的背景信息一起打包，以便（至少）原研究者的同行可以理解，并采用标准定义通用的细节，并按学科来描述内容。信息将存储在管理良好的档案室、知识库或社区数据库，使数据为其他人所用。

所需信息转化为特定比特流的“科研数据”，它可以出现在研究计划、数据管理计划中，也可以出现在捕获它们的仪器和软件中。这些与他们对任何断言（数据是特别的研究调查结果）的了解和检验程度有关。因此，通过数字数据与思想的关系（体现在背景描述和表示信息中）而成为“科研数据”。



在越来越多的领域，“数据”被公认是作为科研项目进展的输出物，并且逐步与同行和利益相关者共享。在其他领域，数据可能是分析和诠释的原材料，但只共享研究文章或书籍中描述的部分。这些文化差异影响了公共数据的共享程度。研究组之间合作与竞争的水平会影响这一点，特别是当技术的发展带来新的机遇，可以跨研究领域汇集数据和转换分析方法时。

科研领域在许多方面有所不同，但有两个共同点，即科研人员是如何相互依存的，他们的任务涉及多少不确定的技术。更多确定性技术手段意味着标准更可行，数据或代码更加可以复制。随着经济规模增长，以及资源围绕大型设施集中化，相互依赖性越来越强。在具有这些特征的领域（如天文学、地球科学和基因组学），数据管理和共享理念更好地确立起来，因此这不是偶然。在数据密集型学科，存在广泛多样的数据类型和标准（包括元数据）。集中式服务可能需要研究者找到折衷的方法——即在提供标准数据和通用目录所需的元数据之间找到平衡，并将他们的研究归入不合适的标准类别。

科研设施、仪器和分析平台对数据产生和管理有重大影响。它们的位置、容量与功能会限制 RDM 的作用，通过开发新的服务于整个机构的 RDM 基础设施，能够更多的实现什么。在地化的“数据密集型”研究集群也可以提供积极实践的案例。确定需求和评估结果时，这些都需要考虑。考虑到机构的备份解决方案，例如，用户对数据检索的需求可能因不同的仪器（这些仪器产生不同的文件数量、大小、复杂度和总体容积）而发生变化。更普遍的是，通用方法需要在极小的范围内使用，并适应研究小组不同的背景和不断变化的需求。

收录数据以及其生产者所表达的想法的工具将决定数据如何获取和汇集。RDM 工具或服务简化“数字研究对象”的使用方式，使用数字化工作流程链接到数据。例如，这些可能包括电子实验室笔记本。在评估操作和利益时，需要考虑程度，考虑数字化研究对象在多大程度上满足“可重复使用、可改变用途、可复验的、可复写的、可复制的、可引用的、明显的和值得尊重的”这些理念。在介绍工具或服务（如数字笔记本电脑以及以协作方式使用的工具）时，观察人们怎样与这些工具进行交互，这是很重要的。人们确保数字工具真正提高用户能力，在这种情况下，数据更有意义。

3.3 利益相关者：生产者、使用者和决策者

在大多数研究领域都会有一系列利益相关者，包括数据生产者、学术科研人员、数据管理人员以及受聘于研究机构的数据科学家。例如，这些利益相关者可能包括公司、政策制定者、非政府组织。RDM 发展需要考虑这些利益相关者的需求，因为他们与计划服务有交集。例如，DataONE 生物和环境研究基础设施的开发商确定主要利益相关者为“科学家”，次



本作品采用[知识共享署名-非商业性使用-禁止演绎 3.0 中国大陆许可协议](https://creativecommons.org/licenses/by-nc-nd/3.0/)进行许可。

要利益相关者为科研过程中经常与科学家互动的群体。如图 3 所示，图书馆被确定为关键的次要利益相关者，因为它们存在于科学家工作的每个机构环境中。

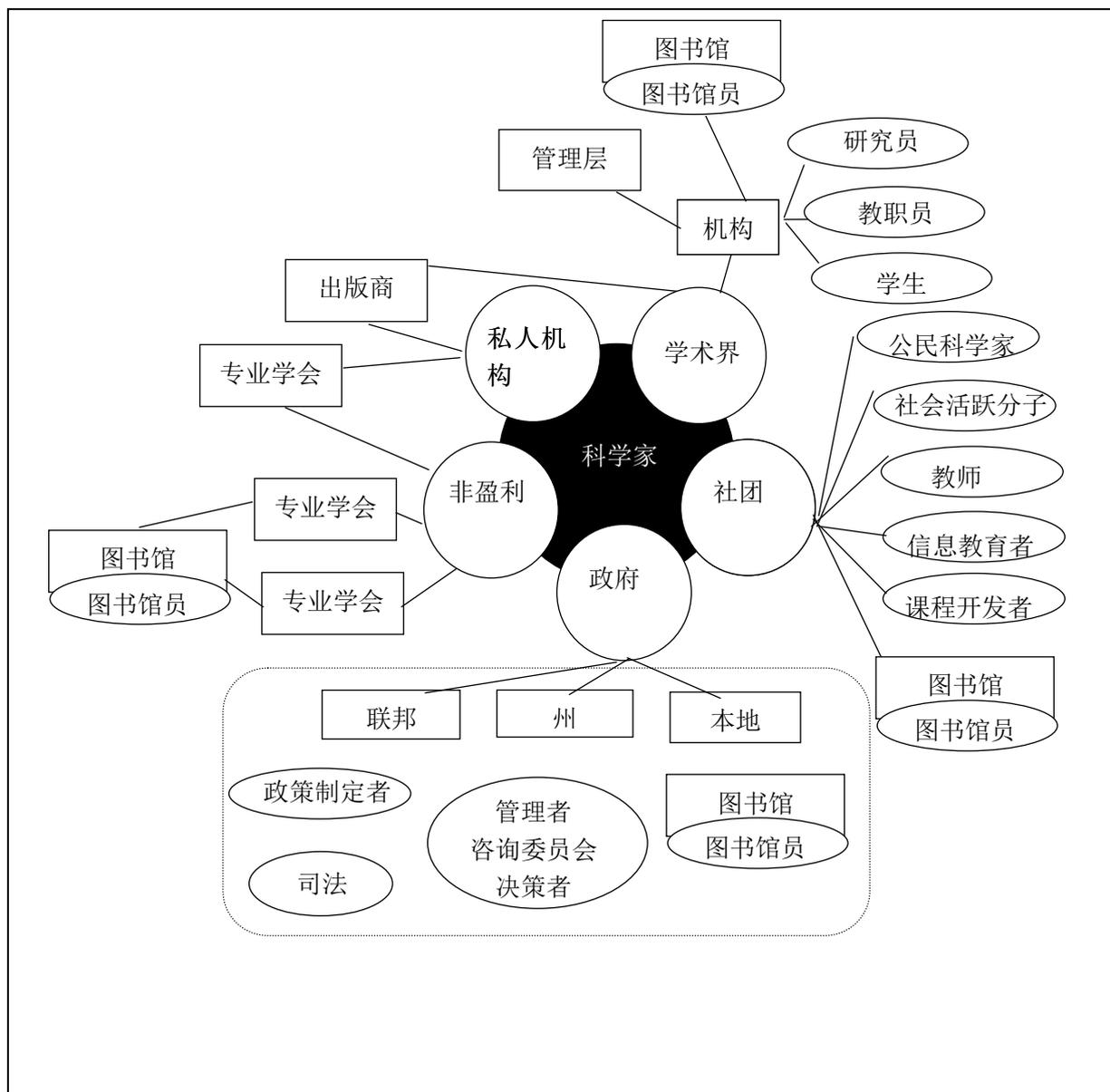


图 3.RDM 利益相关者：来自 DataONE 的案例。椭圆形代表个人角色，方框代表组织机构（经授权转载， Michener 等人）

开发项目时应明确，在科研机构内外部，RDM 服务如何与其他服务供应商进行互操作。这将包括网络基础设施，如美国的 DataONE，以及欧洲迎合特定学科社区的研究基础设施。

在机构内，高级管理人员需要引导项目发展，以确保计划输出成果是可行的、可取的和



本作品采用[知识共享署名-非商业性使用-禁止演绎 3.0 中国大陆许可协议](https://creativecommons.org/licenses/by-nc-nd/3.0/)进行许可。

可持续的。他们还需要支持任何作为发展一部分引入的政策或指导方针，通过（如）有关委员会构成和员工沟通方式进行咨询，并参与机构业务规划过程。

正如前面提到的，科研支持和行政服务通常涉及一个机构的图书馆、电脑、档案管理和科研管理职能。在许多情况下，他们的科研数据管理合作将涉及探索和协商新的角色。通常需要参与 RDM 团队，以实施由督导组规定的操作。确定技能发展的需求，培训和政策宣传很可能被列入团队职权范围内，并且这些将是发现和设计阶段主要关注点之一。计算科研人员和信息专业人员属于图书馆角色，例如知识库管理人员和学科馆员，在确定信息系统需求时，他们很可能被认为是主要执行者。

图书馆员（尤其是承担学术联络或学科馆员角色的馆员）可能已经有了“跨越边界”的技能，能适应 RDM 服务的发展。在数字图书馆发展进程中，对跨学科、IT 等专业领域的工作需求也一样强烈。对于元数据和科研数据编目的需求，将扩展这些技能。数字图书馆研究显示，利益相关者的不同实践网络将影响他们如何构建框架来记录有关数据，以及什么用来共享。

利益相关者将跨越机构界线。例如，他们包括学科数据知识库、将其连接起来的网络基础设施或研究基础设施（如 DataONE），提供跨机构和学科的数据管理服务。这种基础设施可以从服务层的角度来理解。基础层提供网络计算和存储服务，包括云服务提供商和公共国家研究与教育网络（NRENs）。在此之上，中介服务提供标准和系统，以支持数据知识库之间的互操作性。开发人员可能需要与各种服务提供商接触，例如作者识别机构（如 ORCID）、引文数据集机构（如 DataCite）或研究信息标准化机构（如 OpenAIRE、CASRAI、EuroCRIS）。

3.4 机构规则和研究规范：为什么要改变？

近年，数据共享和治理的“自上而下”政策从公共资助方流传到监管机构。这反映了在科学方面应对技术变化的需求，以及关于科研透明度和完整性的更广泛公众利益。数据治理的原则（如由英国研究理事会或美国国家科学基金会定义的）有助于识别机构对功能的需求，以促进科研数据共享。它可以用于区分机构和科研团体需要在操作层面制定的“宪法规则”和“集体选择”政策。例如，有规定明确定义：谁可以将科研数据提交到知识库以及谁可以访问它。

RDM 服务也取决于社会规则——研究小组和他们的利益相关者关于哪些数据是值得保存和共享的问题达成的共识。机构政策不能确定每个决策，但是可以通过设置范围，允许在较低级别设定细节，如研究小组或部门，或特定服务领域。例如，机构哪个部门提供哪种服



务这种决策问题，应该考虑到科研人员如何看待当前产品（如图书馆、计算机、研究支持和培训机构），以及是以集中形式还是院系形式提供服务。

数据共享规范是一个存在争议的话题，它会强烈地影响数据存储服务。“自上而下”政策和科研人员的信念及惯例之间可能存在巨大差异。许多研究调查员将访问条款的决定权看作其道德权利，而且调查显示，只有在直接有利于科研人员研究的情况下，他们才会选择共享数据，如协作或联合署名。

政策必须包括资助者和机构的期望，但这不足以支持数据共享实现。RDM 同开放获取出版一起宣传，可能带来协同作用，因为有证据表明，在开放获取期刊发表文章的科研人员更容易共享数据。发展机构知识库出版物的经验显示，服务提供商需要用站在教师的角度与他们沟通，并强调资源库给他们带来的价值。对于科研人员，他们从他们的组织得到的支持是说服他们共享数据的强有力因素，尤其是以下形式的支持：

- 他们项目期间，在数据管理方面获得的技术支持
- 数据创造者的正式确认或归属
- 选择要共享的数据量
- 数据存储支持
- 搜索和编目数据技巧

科研人员要考虑如何分享，分享什么，从长远来看，他们的研究取得更大的知名度的可能性可能是有说服力的，并且某些学科的研究表明，数据共享和高引用率之间存在相关性。研究合作规模也会影响这个，因此 RDM 服务提供工具和支持是有道理的，这有助于科研人员日益扩大合作范围，建立更加公开分享的信任机制，且有助于确定这样做的成本和效益。

4. 发展阶段

4.1 设想

超越初期的现场服务试点来发展 RDM 基础设施，需要组织高层拥有共同愿景。高级管理人员需要建立一个督导组或咨询委员会来协助他们，或者现有的 RDM 模范机构或团体来协助提升 RDM 措施。理想的情况是，负责科研的高级管理组成员担任小组主席（如负责科研的副主任或副校长）。

督导组的总体目标包括：为 RDM 制定政策原则，帮助协调匹配组织总体任务，并且定义指南中的角色和职责。考虑到该机构的研究策略、外部政策驱动、服务重点和技术机会，督导组也会为高级别审批制定一项战略或路线图。督导组的总目标是完成一份业务状况



报告，它可以提交给任何组织，以保证建立必要服务的经费。一旦这种情况发生，督导组可能重组，以完善既定服务计划和指导项目。

关键要素

- 建立管理承诺和愿景
- 探索科研政策和战略机遇
- 识别技术，驱动变革
- 划定初步调查范围

4.2 启动

取得初步调查承诺后，督导组的工作将集中在提高利益相关者的认识，以及促进“买入”的进一步发展。RDM 督导组常常召集服务提供商，与他们交流他们的教学支持活动，但很少交流科研支持或任何直接相关的数据。

小组成员相互理解并找出突出问题。这包括识别满足资助机构和监管机构的政策要求的能力差距，以及寻求好的机遇（例如，新的研究，或数据管理效率等）；若与科研人员、机构、外部客户及利益相关者就有关的好处和风险进行对话，将有助于确定预期结果和成功标准。

各级科研人员的参与是至关重要的，并且高级科研人员应列入督导组。为了了解数据工作流正式化或开放数据的意义，有必要咨询科研用户或知识交流专家的意见。

确定数据生产者和用户的真实需求范围，可采取分组座谈会或研讨会、调查和结构访谈方法。与督导组所能提供的资源相比，这很可能需要更广泛的资源。运营团队也需要得自图书馆、信息技术以及科研支持服务的支持。

适用本阶段的工具包括在第 5 节中介绍的数据资产框架（DAF）。KRDS 效益分析工具（26）确定了科研人员、服务提供商和外部利益相关者的潜在回报，也可用于本阶段。通用项目调度和预算编制技术也很可以在这个阶段使用。

关键要素

- 识别、咨询，以及告知利益相关者
- 确定科研人员和用户优先级别
- 确定预期成果和成功标准
- 组织运营团队



- 进行项目规划

4.3 探索

从较高层面确定优先事项工作，起始阶段确定改变当前做法的需求，了解所需经费。重要的是要明白一些关于数据生产、共享和使用的学科形势和现行规范。这项研究也将考虑支持服务的形势——哪些人参与其中，以及研究者和其他利益相关者遇到的问题。探索阶段应该以使用案例或用户情景的形式，规定所要实现的 RDM 服务。

科研自然会涉及到数据采集的高度专业化知识和非标准化技术。不同的科研群体，甚至在同类学科中，都会有自己的方法。鉴于机构中科研小组的数量和多样性，将需要挑选 RDM 督导组或项目经理，之后再界定服务需求的范围。理想的情况是，该组的成员来自不同资金来源、数据类型和研究团队规模（即从个人科研人员到大联盟）的团队，从而使科研项目避免第 3 节中强调的因素，可以解释变化。让不同职业阶段的科研人员参与，这也是非常有用的，例如，博士学位学生的需求和关注点与资深教授就非常不一样。

确定人们需求的变化是非常重要的，需求如何表达，以及如何使可能存在的障碍符合 RDM 战略和法规要求。因此，探索阶段可能包括对相关政策认识的评估，并以图表展现典型数据资产的生命周期和相关研究对象（软件、协议、日志等）。在实践中，参与程度将取决于科研人员的兴趣或关注度。

因此，对需求发现方法的选择应包括一系列方法，研究者可以快速、便捷的参与到其中，并对整体需求和政策义务产生基本认识，对那些花费更多时间，但产生更深层信息的当前资产和做法有基本了解。通常情况下，项目经理或业务责任组将在开展这项工作前，进行一系列短期探讨，包括：选择研究组和任何相关服务（备份、存储或图书馆支持）的当前提供商。有用的工具包括：数据保存用户文档、数据管理的必要条件、利益相关者简介和数据资产框架（DAF）等。

访谈和研讨会会产生大量的定性描述，需要进行甄别，以确定最需要实现的功能。科研数据基础设施和目标联合评估工具（CARDIO）能够补充这一分析。它给 RDM 项目经理提供了 RDM 服务理想功能的通用模型。该模型以比例量表的形式应用，使利益相关者能够对当前提供的服务进行数字评分，并交流不同的看法，用图表展现来自访谈或研讨会（由科研人员和其他服务的用户参与访谈或研讨会）的数据。通过对大量定性的需求采集资料进行总结，得出初步框架，本模型有助设计服务。



这种分析应纳入多种标准设计方法，用于记录用户需求，如用例和用户故事（27）。验收标准将用于判断如何实现用例，它是本阶段同样重要的成果。在这时，RDM 利益框架也是有用的，例如前面提到的 KRDS 效益分析工具（27）。

关键要素：

- 证明现有数据的做法和支撑
- 分析现有数据的做法和支撑
- 确定所需的组织、技术和资源能力
- 识别用户需求和数据相关验收标准

4.4 设计

设计 RDM 服务是一个反复的过程，在探索阶段的第一第二步骤（Alpha 和 Beta）是改进与建立原型。正如其他任何服务设计项目一样，新服务的基本概念将在探索阶段里确定。在设计阶段，应逐步确定每个服务将要为其用户提供什么，需要什么样的功能来支撑，需要具有怎样的价值，以及如何宣传其价值。目的、功能、质量和性能这些元素可以用来描述所提供的服务（28）。

灵活的设计方法可以产生效益。起草使用案例或用户情景，首先是创建角色和职责，以提供所需的功能和支持级别，例如，在线或某种程度的面对面建议或咨询。这些应与服务的高级别说明保持一致，之后再行服务的精炼，标识实现所需用例的模块化服务及其相互作用。“精选微服务”模式（29）是这种模块化方法的一个例子，这有助于减少成本和个别服务可用性变化的影响。

对于任何在线 RDM 系统，在 alpha 阶段都需要进行详细的业务规划，才能进入 beta 阶段。无论采取哪种设计方法，alpha 阶段都应征求用户的反馈意见，根据用例、线框图或纸本原型，形成基本的工作原型。如果这意味着可以利用现有资源创建可行的办法，beta 阶段将落实所学知识，并产生一个在线服务的完整工作原型（参见《政府服务设计手册》）（30）。

设计阶段还应考虑与其他服务集成的需求。这些最可能包括：成本核算系统、目前的科研信息系统（CRIS）和科研成果信息库。设计过程还需要考虑到现有或计划中的较低级别基础设施，如网络附加存储或外部云存储服务。

RDM 基础设施成熟的学科（如天文学，或基因组学），可能有完善的平台和工作流程，来使用和存储外部基础档案和虚拟研究环境（VREs）中的数据。也可能有本土专家资料库，



可以与任何 RDM 服务提供的中央数据知识库结合。按设计方案来改造现有平台，不可能在很短的时间内完成，但可以采取循序渐进的方式，与研究团体一起寻求机会，将元数据管理工作流程与这些集成，以在一个机构目录发布元数据。可用于工作流程建模的工具包括：科研活动信息发展（RAID）图（31）以及 Web Curator and MyExperiment（参见 DCC 工具服务目录（32））等。

关键要素

- 定义和分析新的服务理念
- 新服务的原型和详细设计
- 设计人力资源结构
- 分析和设计数据管理工具及基础设施

4.5 执行

设计 Beta 阶段应确定各支持功能（如 IT、图书馆或研发处）分别提供哪些“业主”所需的服务，以及这些服务中个人角色转变的需求。它还应确定是否需要新的关系和工作流程，以连接这些新的角色。实现这些更改还需要进行宣传、培训和进修。随着新的工作流程落实到位，所有已出版的指导意见可能需要更新，以传达这些改变。由于扰乱了服务提供商现有的规范，以及长期以来项目负责人（PIs）承担唯一责任这一做法，实施过程中可能会有详细谈判。

数据管理工具和基于软件服务的关键问题可能是与其他系统集成的问题，这要求与其他系统遵循的一致标准。对很多大学而言，RDM 支持接口标准（如科研信息系统输出报告）不仅新颖而且还是资助者传播政策里的要求（例如，提供开放获取）。这就尤其影响系统之间的元数据交换条件。工具应当具有足够的灵活性，能满足基于标准的用户文档，尤其是那些基于 CERIF 研究资料（见 EuroCRIS（33））的文件，作为广泛地用于科研数据集的相关信息。

RDM 服务机制需要符合网络无障碍标准，或 ISO 27000 信息安全管理标准（34）。当然，所有的合规性需求在探索阶段就应明确。然而，这些会随着 RDM 领域标准进一步发展而改变。例如，在一些机构为其 RDM 服务进行“受信任的知识库”进行认证时，采用数据封装标准和 ISO16363 标准（见 APARSEN，2012），它的合规性需求可能会发生改变。



数据管理工具和基础设施实施之前应该确定服务级别，以反映出市场对可用性和可靠性的期望。验收措施、测试计划以及其它性能标准应该在 beta 阶段确定。Live 阶段中应包括服务运营流程和业务模型在内。

关键要素

- 重组支持服务
- 实施数据管理工具和服务
- 确保符合相关标准

4.6 评价

RDM 服务是否“上线”，将取决于是否向高级管理人员提供全面的服务。资源配置决策可能取决于是否向用户和其他利益相关者展示出实实在在的好处。如果这些在探索阶段早一点确定，探索阶段，可以根据收集的可用数据反映的实际问题，在后期阶段进行完善。它应该提供基础反馈和分析机制，持续改进提供的服务。

项目组以路线图或行动计划来控制进度。除此之外，标记能够协助项目团队和其他人继续完善所提供的功能，并引导从 alpha 阶段进入到 beta 阶段的决策，使该项目获得更多的资金。CARDIO 工具（见第 6 节）和其他通用模型可用于（例 35）其能力范围内指导开发团队。

项目应广泛征询预期效益，并将切合实际的指标作为其成绩的证明。在项目开始时就实施这一工作，并且贯穿整个项目过程，而不仅作为项目收尾活动，这有助于评价服务是否满足用户的实际需求。例如，在 JISC 科研数据管理计划中，有一批“证据采集者”收集个别项目的益处证明（这些提炼于叙述性、简短的个案研究）以及定量指标（例如，下载）。

关键要素

- 评估服务改善后的效益和成本
- 确定持续改善计划的指标



5.启动与发现需求

本节推荐的方法已用于 RDM 领域,并且有众多的更为通用的工具和方法用于需求发现,这些也可以借鉴(见例 37)。

5.1 深度个案研究

尽管研究小组进行数据实践案例研究非常耗时,并且花费高昂,但这对于理解实际需求有很高的价值。进行案例研究就可能“身临其境”,它涉及研究者背后的信息或数据科学专家,以及他们获得数据过程中的故事。假设数据案例研究花费的时间由数据管理问题的复杂性决定,那么,可能需要进行数周或数月的参与以及观察研究。例如,最近的一些研究指出,数据的再利用领域存在很大的挑战,特别是数据新用途或跨学科方面(11,38)。案例研究也涉及科研人员参与的研讨会或分组座谈会,主要是关于数据政策和支持需求问题。

下面介绍的轻量级方法也会采用访谈、分组座谈会或研讨会的方式。显然,需要平衡投入和回报。在开发环境下,轻量级方法往往更适用,但在信息或计算科学领域,可以与专门科研人员协作,此时深入研究更适用。这些可以促进新见解和 RDM 模式的产生,或说明事实真相。

深度研究可以便捷的判断出工作流程哪部分需要进行重大改变,并且预测重大影响。例如,将数据存储进资料库的程序自动化,可能需要对工作流程进行深入分析,以确保正确的理解需求。如果 RDM 发展方案内没有这种研究,可以从其他来源找。例如:国际数字管理杂志(www.ijdc.net)、数据保护用户文档目录(<http://docs.lib.purdue.edu/dcp/>)或 DCC 和 DataONE 网站。

5.2 调研数据实践和标准服务的能力

许多公共机构的 RDM 项目受益于科研人员和其他利益相关者进行的数据管理在线调查。这些通常借鉴类似的问题,进行较大规模的调查,调查人们的一般做法及对数据共享和再利用的态度。例如,在 2009-2010 年,DataONE 项目调查了数千名科学家,总结出“标准评估”,之后,又利用调查结果完成了详细的利益相关者角色定位。

数字资产框架



本作品采用[知识共享署名-非商业性使用-禁止演绎 3.0 中国大陆许可协议](https://creativecommons.org/licenses/by-nc-nd/3.0/)进行许可。

通过在线和面对面调查和访谈，数据资产框架（DAF）给出了一个快速和轻量级的方法来发现数据管理实践。DAF（39）的主要阶段是：

- 第一阶段：规划，确定调查的目的和范围（例如，全机构或特定教员团体或支持功能），并进行初步的研究。一些机构的 DAF 研究有助于 RDM 政策草案咨询，通过信息治理和研究委员会，这些信息有助于大学的正式决策。

- 第二阶段：识别数据资产并将它们分类，以确定在何处集中力量进行更深入的分析。这个阶段可能仅仅确定数据类型和数量，或者可以在数据实践的详细审查中，进行更彻底的科研小组资产评估

- 第三阶段：识别并遵循定性科研人员研究工作流程的典型生命周期，并识别数据创建和保存实践中的机会及威胁/风险。

- 第四阶段：汇集所有收集的信息，并提供数据管理改进建议。

第二和第三阶段采用网上调查和半结构化访谈相结合。网上调查通常有 10-20 题，涵盖了研究活跃的工作人员对政策要求的认识、数据管理规划的责任、利益的预期、对培训和指导的需求、备份和存储方面的当前做法、工作数据的获取、具有长期价值数据的共享，以及他们提供支持服务的优先事项。

半结构化访谈通常涉及多个试点组，跨学科、资金来源和研究团队规模。依据团队大小，访谈一般涉及 3-6 名研究者，从博士研究生到教授水平的小组领袖。访谈主题与调查问题类似，但以更轻松的谈话方式来了解研究者当前领域的研究和背景。这包括了第 3 节所述的部分。例如，数据采集时使用的仪器和工具，用于处理数据的工具、标准和基础设施，对数据再利用的看法以及政策走向。

为这些研究提供资源需要仔细规划。结果将是 RDM 需求范围的高层次分析，并向 RDM 督导小组提供关于服务发展优先事项的建议。对于定性研究案例，DAF 研究可以得出类似结果。例如，虽然目的是发展而非研究，但是伦理审查可能涉及访谈问题、知情同意书，以及要求匿名方式进行等。

备注或记录的任何副本都将需要加以分析，虽然这和研究的严谨要求级别不同（通常需要大约 6 小时来分析 1 小时的访谈内容），但至少需要和访谈相同的时间来完成。可以通过增加每次访谈的人数，减少每次访谈的时间（每次大约 15-30 分钟），或减少开放式问题并增加封闭式问题来实现（后两种方法会降低结果的丰富性）。

在 Georgia Tech 使用 DAF



本作品采用[知识共享署名-非商业性使用-禁止演绎 3.0 中国大陆许可协议](https://creativecommons.org/licenses/by-nc-nd/3.0/)进行许可。

科研数据项目馆员 Susan Wells Parham

虽然 Georgia Tech 因其工程项目而闻名，它在一系列技术和数据丰富的领域（内科学、社会科学和人文学科），拥有强大的科研能力。我们进行科研数据评估的目标是，在这些不同的领域，推广广义上理解的科研数据环境。

不对单一的学校或研究小组进行全面审计，我们使用 DAF 创建一个在线调查，收集基本数据保存和管理的信息，确保各学校的研究机构至少有一个科研人员参加调查，并且他们来自不同研究中心。我们招募具有广泛实践工作的科研人员，并接受各种层次的技术和财政支持（共 63 个响应）。

Georgia Tech 的研究数据环境概况使我们可以完善现有的服务，并对未来的数据保存做出计划。调查结果显示，在无资助的科研项目中存在服务差距——许多受访者不参与赞助研究，这些研究表现出对数据共享和保存服务以及数据管理最佳做法相关信息的特定需求。我们主要针对教师、研究生和科研管理人员进行宣传，发展向外联系（包括演示文稿、文章、网页指南和印刷材料）。我们还为编写数据管理计划的科研人员制定了培训和咨询方案。

我们基于 DAF 的调查还显示，许多研究者创建了易于访问的数据集，它们的大小相对较小，并可以无限期对公众开放。这些标准适用于我们机构知识库的收集策略——SMARTech，所以我们为科研数据集开发政策，并开始与研究者合作，以通过我们的知识库收集和保存数据。

我们要求受访者从预定义列表中选择数据保存服务（不限定选择的数量），表明他们对数据保存服务的兴趣点。73%的受访者表示对数据存储和保存感兴趣；67%的受访者表示对数据共享工具感兴趣；52%的受访者表示对数据管理最佳做法感兴趣。大约有 40%的受访者表示对以下几点感兴趣：编写正式数据管理计划的相关信息；协助会议资助机构整理数据管理要求以及长期保存数据的选择问题。这些调研结果体现了我们与 Georgia Tech 其他单位的协作，以及我们为提供长期数据存取和保存而制定的科研数据管理合作策略。

科研数据基础设施及其对象的联合评价

科研数据基础设施和目标的联合评估（CARDIO）是可以用来评估当前和所需支持能力之间差距的标准方法。CARDIO 和 DAF 在寻求当前支持科研数据管理的信息。DAF 收集半结构化信息，CARDIO 采用比例量表对不同时间或不同组间的维度进行比较。

CARDIO 标准最初涉及协调员或项目经理，通过对给定范围内的支持元素分级来作出初步评估，然后招募参与者，并邀请他们给自己评级。然后进行比较评价，达成共识，例如，



本作品采用[知识共享署名-非商业性使用-禁止演绎 3.0 中国大陆许可协议](https://creativecommons.org/licenses/by-nc-nd/3.0/)进行许可。

通过讨论，责任已充分明确的领域应给予高评级。可以在访谈、研讨班或进行调查时，进行参与者评级和讨论。另外，CARDIO 在线工具（40）可以用于收集网上应答，任选一个聊天工具即可实现。

CARDIO uses a rating scale to assess organisational, resourcing and [technology elements](#) of RDM service provision. These elements can be assessed to different degrees of granularity according to the level of engagement required:

CARDIO 使用比例量表来评估组织 RDM 服务提供的资源和[技术元素](#)。根据所需的参与程度，这些元素可以按不同程度来评定：

1、一个简单的“CARDIO 测验”有助于启动对提供服务的审查。测验包括 13 个问题，每个问题都是在三个代表其机构当前状况的选项之间做出选择。然后将测评结果用于优势和劣势评估，这之后可以进行更深入的评估。

2、“路线图矩阵”用来评估提供 RDM 服务能力的提升，这是英国资助机构期望受资助机构开发的。其包括一个在研讨会中使用 9×5 的表格。

该表格的九行可按以下三个标题分组：

- 政策、发展战略和可持续性
- 数据管理支持和员工发展
- 科研数据的存储、保存和共享

对于表格中的每一行，用户可以在五个选项中进行选择，该五个选项表示该机构嵌入所需服务的阶段。这五个步骤大致与本指南第 3 节所述的步骤一致：

- 1) 构想&启动：识别对变革的需求，并采取行动
- 2) 探索：调查并确定需求范围
- 3) 设计&试点：解决方案进行小规模试点测试
- 4) 推出：解决方案受到资助，并更广泛的进行试点
- 5) 嵌入：服务到位并持续改进

3、服务可能会符合资助机构的期望，但不一定是最优的方式。所以，第三种选择是评估提供服务的成熟度，而不是积极实践积极实践所定义的内容元素。整个 CARDIO 模型使用了 30 个元素描述，涉及组织、RDM 服务提供的技术和资源方面。这些元素和语句描述了积极实践，可从整体或院系或更小的单位，评估该机构，例如，根据提供服务决定是否下放或集中来划定范围。



CARDIO 工具可以收集在线评估。面对面研讨会也是有价值的，尤其，来自不同研究和服务背景的个人进行评估时，通常不进行交互。这两种方式也可以混合使用，例如，面对面研讨会用来介绍该工具以及参与者相互认识，随后，项目经理跟进，使用在线工具，以便进行更详细的评价和讨论。

在 Warwick 大学使用 DAF 和 CARDIO

DAF 和 CARDIO 可以相互补充。例如，DAF 可以实现“情报收集”角色，然后使用 CARDIO 进行评估。收集科研人员的现行做法以及其对所提供服务的意见，然后可以对已达标的服务进行重新评估。另外，通过确定问题的优先次序，CARDIO 研讨会会有助于确定 DAF 调查范围。

2012-2013 年间有几所大学举办了 DAF 和 CARDIO 研讨会。其中之一是 Warwick 大学。DCC 员工和专业信息工作人员在学术联络、资源库管理和研究方面协作，分析研讨会收集的需求信息。这些涉及学者和提供商，审查影响 RDM 支持、挑战和能力差距的治理机制因素。该研讨会结合专题介绍和小组讨论。根据前面提到的三类成立了三个小组，即：

- 政策、战略发展和可持续性
- 数据管理支持和员工发展
- 科研数据的存储、保存和共享

每一个类别有三种能力级别。首先作为个人，然后讨论他们得分的理由，以及他们所认为的优先事项，最后，针对每一种能力的级别达成共识。DCC 整理会议笔记，总结了一份关于级别评定及其理由的报告。

后期，该研讨会的意见可以用来制定一套访谈问题，其改编自先前的 DAF 研究。然后，DCC 访谈两个试点组的科研人员，他们受到了参加研讨会的 Warwick 信息专业人员的影响。

访谈涵盖以下主题：

- 数据管理和共享规划
- 数据收集
- 数据处理和分析
- 数据可获取性、有价值数据安全性
- 支持和培训
- 数据管理方面的挑战
- 对服务和优先顺序的预期



将访谈记录整理成文字，并将这些总结反馈给参与者，以检查是否准确的表达出他们的意见。在随后的研讨会上，告知服务提供商有关该访谈所凸显的意见，以促进制定执行计划，将大学的科研数据管理战略付诸实践。

数据管理必要条件（DMVitals）

Virginia 大学开发的 DMVitals 方法与 DAF 和 CARDIO 有相似之处。像 DAF 一样，它是针对个别科研人员的访谈，其目的是记录他们数据管理方法的详细信息，以了解如何进行改进。DMVitals 采用高度结构化访谈协议，该访谈由封闭式问题组成，而不是本指南中其他方法所使用的开放式问题。

DMVitals 访谈所提的问题都是基于数据管理实践中的八个“组成部分”，它们是：

- 1) 文件格式和数据类型
- 2) 组织化文件
- 3) 安全/存储/备份
- 4) 融资指南
- 5) 著作权以及隐私/保密性
- 6) 数据文档和元数据
- 7) 档案与共享
- 8) 引用数据

每题的可用选项都涉及积极实践积极实践的声明，这起源于 Virginia 大学的指导方针，以及澳大利亚国家数据服务（ANDS）的长期存储选项评价模型。（41）

与 CARDIO 类似，“可持续发展水平”评级考虑到每个组件可能的响应范围。五级比例量表是由澳大利亚国家数据服务（ANDS）提供的科研数据管理能力成熟度模型派生出来的，从“初始”，经过“发展”、“定义”、“管理”，最后“优化”。（42）

虽然 CARDIO 评级被看作利益相关者之间对话（关于如何改善支持）的基础，而 DMVitals 工具是使用得分评估个别科研人员 DM 实践的可持续性，并按质量最差到最好排序。这借鉴了对可持续水平的预定义“行动声明”，而该科研人员的测评结果表明了可持续水平。提供用于定义和改善科研人员 DM 实践的框架，通过 DM 可持续性比率的平均值来定义数据管理成熟度水平。因此，该 DMVitals 工具针对数据管理访谈，提供即时和可操作的反馈，并且可以用于数据管理规划的自动化咨询指导。



该 DMVitals 工具的优点在于创建数据管理报告，它为每个科研人员指定定制的任务。这些任务可以很容易地划分成不同的阶段，基于他或她的个人资料、访谈记录和之后收集的信息，为每个科研人员创建一个数据管理实施计划。将此工具与评价和规划方法相结合，有助于加快建议报告的过程，并提供有价值的可操作性反馈，这些可以供科研人员立即使用，以提高他或她的数据的可持续性。

需求在已知范围内时，自动响应方法有助于需求收集，但不能充分考虑个人的研究目标和情况。这不一定导致对需求的“一刀切”做法。如果作为更为开放式的访谈标准测试工具，DM Vitals 应该帮助科研人员明确在“坏”的数据管理实践中，做“好”的原因，并引导他们重视该工具的建议，以适应其特定的数据生命周期。

5.3 利用数据保存用户文档使数据生命周期文档化

数据保存用户文档（DCP）是另一个工具，用于收集有关科研人员数据集的信息：他们在做什么，他们想怎么做。使用工具时常常假定：科研人员没有时间去思考用知识知识库存储数据需要什么。用户文档可以帮助科研人员阐明数据的各个方面，如什么实验成果应该共享（例如，原始的或分析后的数据）。它还要求科研人员描述在使用、引用、权利等方面的需求或要求。因此，通过用户文档可以洞察科学数据的工作流程，以及推广到其他人所遇到的障碍。

将该数据保存用户文档作为访谈工具是一个研究项目的成果，它探索了“谁愿意用什么和谁在什么时间分享”这个问题。（43）该项目最初访谈了 21 个科研人员，讨论他们的数据，并通过一个迭代过程，进行了一系列科研人员认为重要的调查。其中包括：什么样的数据将是重要的分享、分享的数据的格式或形式、所有权涉及的参数、使用条件等。

数据保存用户文档（DCP）概述了数据集或集合的“故事”，描述了一个研究项目的起源和生命周期。开发该方法是用于处理“……在形成保存要求和服务时，需要涉及哪些学科和子学科”（44）这个问题带来的挑战。

从数据保存用户文档可以看出应用于特定集合的决定，例如，在选择保留的数据集时，以及提供元数据时。该方法的关注点是对跨具体研究社区时数据共享做法差异的分析，以及这些差异的影响因素，例如，所使用的数据类型和该数据所处的生命周期阶段。



DCP 方法在某些方面与 DAF 类似，包括所询问问题的范围。像 DAF 一样，用户文档是供图书馆员和他人使用的决策信息。然而，DAF 注重对数据集的普遍认识和跨部门或机构实践，而 DCP 更注重子学科间的关联数据。

研究馆员“数据访谈”使用 DCP 的过程是有用的培训和发展，可以培养参与者的信心，使其参与到有关数据管理的讨论中（例如，参见 MANTRA ）（45）。设定的数据保存用户文档访谈是两个小时的谈话。像 DAF 一样，该方法采用了半结构化访谈形式，访谈内容有：受访者在研究过程中遇到的问题、收集数据的类型、可访问性和所有权问题、在科研究过程的数据转换方式、共享各个阶段成果的相关实践等。

工具包包括用户文档模板、一张工作表（科研人员、访谈员会在访谈过程中填写完成）、访谈手册和用户指南（46）。已完成的用户文档会提交到出版物引用资源——数据保存用户文档目录。该 DPC 目录提供一套服务，以支持用户文档的出版物，包括：为每个已出版的 DCP 指定一个 DOI、通过列入索引和发现工具以提高用户文档可见性、通过 CLOCKSS 和 Portico 承诺保存 DCPs。

用户指南旨在指引馆员有关访谈访谈的整个过程，从开始接触，到录音访谈，以及将谈话内容合成用户文档。用户指南还针对可能出现的问题，给出了建议，例如，在讨论相关数据集时，保持科研人员谈论的方面正确；如果科研人员没有足够的时间，该把访谈的重点放在哪一部分。

通常一位科研人员基于以下原因而被标识：他们对数据保管的需求，以前的关系，或者是因为他们的研究数据在其他方面很重要。确定重点关注的具体项目，这作为准备工作的一部分，是至关重要的。事前应该通过网站进行搭配组合的审阅，确定近期出版物或预印本研究，让访谈访问者熟悉科研人员的工作。访谈访问者应该能够根据“英特尔”这个标志，确定访谈的科研人员以及讨论的数据集。（注：这次访谈并不收集有关科研人员的个人信息，如果访谈访问者希望以人类主题研究为主题，以已批准的协议形式出版访谈结果，那么，就需要科研人员的个人信息。）

通常情况下访谈过程需要两个小时，一般划分成两个阶段。但是，如果由于某种原因，科研人员无法或不愿意花费那么多的时间，访谈访问者应该准备专注于个人最重要的部分，必须包括：该研究的概述、按种类和阶段进行的数据分析、对数据组织方式的描述、科研人员对共享和获取的观点。当用户文档不包括其他部分（例如，使用工具、发现机制、互操作方式、知识产权等）时，至关重要的是在最低限度捕获所需的信息。访谈工作表是科研人



员填写的表格，提前填写或在访谈期间均可填写。这和数据保存用户文档一样重要，它是为了从研究者的角度捕捉信息。访谈手册有助于访谈双方顺利完成访谈过程。该手册建议增加有深度和广度的答案，从而形成一个丰富的用户文档。为了保证记录下访谈的详细细节，它建议在得到研究者允许的情况下，对访谈过程进行录音。当沉迷于数据讨论时，很难记录有效的笔记。

现场访谈之后再行转录或“索引”，将信息合成到用户文档模板中。目前，用户文档是半结构化的，还不能很灵活的表示科研人员对于多种数据属性的观点。（例如，所收集的数据可以被称为“初始的”或“原始的”，依据学科、实验室实践、特定项目，改动数据再分析可以称为“加工的”或“无显著个性特征的”等等。

最初设计用于审查具体项目的一般研究数据属性，用户文档也可以在其它情况下使用。一些问题有助于设计数据工作流程，解释了对保存、共享、保护“下游”的需求。或者，该用户文档有助于加强需要保存的数据集的管理要求（例如，对其他科研人员、出版商、或公众）。但迄今为止，所见到的最大好处之一就是提高认识和建立关系——通过提问正确的问题，科研人员对帮助他们的人表示感谢。

5.4 利益相关者用户文档、角色和方案

利益相关者用户文档

利益相关者用户文档设计用于更深入挖掘利益相关者已有的数据管理知识，以及它们是如何与数据相互作用的。该用户文档还可以帮助识别工具和服务如何更好的服务利益相关者，并洞察拓展的工具和服务，或提高利益相关者的数据实践。

构建利益相关者用户文档的第一步是要找出主要和次要利益相关群体以及它们彼此间的关系。了解利益相关者在帮助谁，工具和服务开发人员优先考虑资源的分配，以满足每个群体的需求。主要利益相关者是指正在开发的工具或服务所重点关注的对象。例如，在 DataONE 案例中，我们的目标是生物和环境综合研究奠定基础，因此，其主要利益相关者是科学家。

确定次要利益相关者更具挑战性，因为可能会出现各种各样的组织和个人，在研究过程中，他们经常与主要利益相关者互动。一个有效确定次要利益相关者的方法是，可视化呈现主要利益相关者的成功研究环境，然后确定次要利益相关者。正如上文第 4.3 节所述，DataONE 方法主要确定科学家曾工作的五个重点科研环境，然后找出这些研究环境中的其



他利益相关者。例如，在每个研究环境中，图书馆和图书馆员都被确定为重要的次要利益相关者。

接下来的步骤是了解有关的态度、做法、看法，和每个利益相关者群体的需求。有必要进行评估调查，以了解利益相关者群体当前的态度和做法，并且提供用于评估的基准线，采用怎样工具和服务，以及随着时间的推移，他们怎样调整研究和数据管理的做法。通过对各利益相关者群体的独立调查，可以构建丰富的场景——在整个数据生命周期的不同阶段，各利益相关者群体如何与数据进行交互。重要的是，它提供了判断不同态度、信念和习惯的能力。这使得工具和服务的开发者更好地识别并优先考虑需求。

如图 4 所示的 DataONE 数据生命周期是从数据角度发展而来的，我们期望多个个体在不同点进行交互，并且参与不同的活动。

在 DataONE 中，对科学家的评估在项目初期就完成，之后定期重复。对五个科研环境中次要利益相关者的评估，也是优先开发和部署的。通过这些评估，更完整的展现了复杂环境的科学研究过程，促进网络基础设施的开发工作，提升群体参与度。

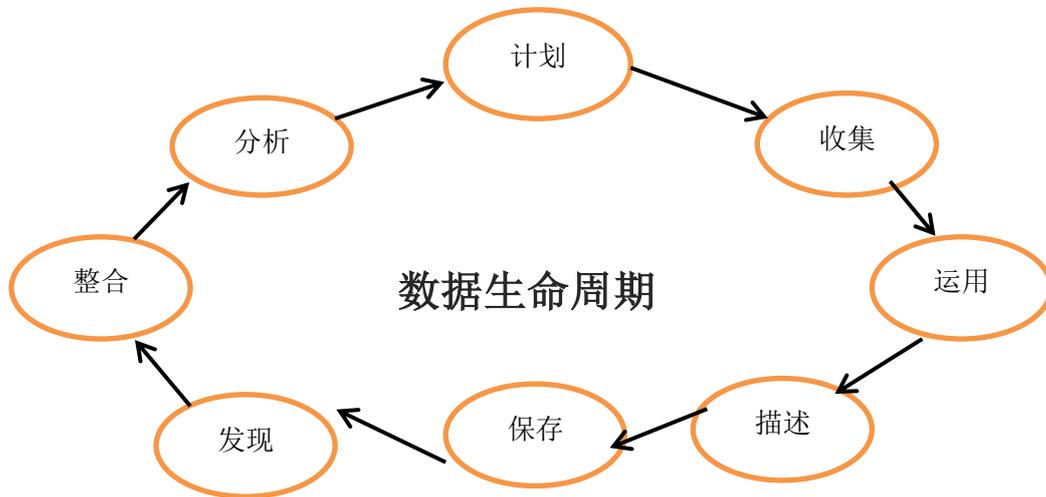


图 4 DataONE 数据生命周期

型人和使用情境



本作品采用[知识共享署名-非商业性使用-禁止演绎 3.0 中国大陆许可协议](https://creativecommons.org/licenses/by-nc-nd/3.0/)进行许可。

上文讨论的利益相关者评估，提供时下盛行的态度、观念、做法和要求。然而，在为数据服务和管理开发工具和服务的时候，了解各利益相关群体的成员如何参与研究过程，尤其重要。

以用户为中心的服务设计方法是在拟议的服务下，分析利益相关者的特征，创建“型人”的可能会影响其使用服务的关键特征，并制定较典型的情境场景。(48)在前面提到的设计阶段，利益相关者用户文档有助于在探索阶段确定用户需求范围。在设计阶段，型人有助于典型用户特征与新兴的设计理念保持一致。情境则更加详细地介绍：该服务将如何为用户(以型人来表示)提供服务。

DataONE 项目的设计过程完整体现了这些方法的使用过程(Michener 等人, 2012)。DataONE 网络支持科学家和其他利益相关者从事相关科学、数据和政策领域。这也有利于方便、安全、持久性的存储数据，并传播数据发现、分析、可视化和决策工具。正如 3.2 节所提到的，DataONE 以利益相关者的现行做法、观点、数据生命周期中各阶段的需求为“评估基线”，跟进其利益相关者分析。这些都是在线调查，与 DAF 调查及数据保存用户文档的问题范围类似。相对于它们，DataONE 对服务角色和利益相关者的人口统计学特征调查得更详细些。

根据调查结果和其他材料(如 DCP 的访谈记录)，型人用来描述主要和次要的“典型”利益相关者。其中包括：处于职业生涯早期、中期和晚期的研究科学家、数据建模者、数据图书馆员、民间高手、大学管理者等。给型人设定一组用例，代表 DataONE 工具和服务的预定功能，然后将其充实到情境之中，描述在生命周期的特定点(如跨多个数据源的搜索)，如何使用特定功能等。

重点关注用户所参与任务的用例和场景，可以阐明一个人在研究和数据保存周期中如何利用工具和服务。用例枚举了用户使用系统时会采取的操作，它有助于查找所需要的功能，并且是常用的软件开发工具。情境类似于用例，但是它是叙述性文字，是在寻找用户的潜在目标和动机，以描述系统是如何使用的。

用例和情境专注于用户承担的任务，而不是用户，因此，他们可以通过创建型人得到加强。这些都是典型用户的代表，包括他的要求，以及情感观点(如情绪)。型人通常由数据创建，而数据是通过评估、访谈或人种志方法，从用户那里收集来的。即使在不能直接与用户协商的情况下，也可以归纳整理这些数据，使型人的特色，进行项目开发，并确定价值主张和情境，以指导工具和服务方面的发展。例如，在 DataONE 中，工作组成员通过社会



文化和可用性评估确定为主要和次要利益相关者群体里个人，创建场景，然后列出科研人员所希望有的工具的清单，以及 DataONE 提供的增值项目清单。

5.5 发展研讨会——从编程马拉松到混搭

情境、用例或轻量级语句的功能需求在敏捷开发方法中仍然相关。像 SCRUM 一样，其摒弃了传统方法，取消了独立“需求”阶段，以免导致冗长的规范。在任何软件开发方法中，压倒一切的目标是，使产品或服务的“主人”清楚地阐明需要什么功能，并定义对他们来说，高质量意味着什么。（50）通过开发一些作品，而目标用户能够对如何改进给出反馈意见，敏捷开发的重点是围绕有用的产品或服务迅速转动。

研讨会是快速确定发展优先事项及原型设计解决方案的另一种方法，尤其适用于开发者、用户和其他利益相关者之间结构松散的密集合作活动。Web 开发催生了日益广泛的“非传统会议”，数字监管和保存群体已经迅速采用和适应它们。

CURATEcamp 就是一个例子（51），其灵感来自于类似 Hackfest（52）和 BarCamp（53）的活动格式。他们的共同目标是鼓励多元化群体出席，并通过分享想法和方法，参与解决问题。相较于软件制作，它更强调互相学习。CURATEcamp 活动有助于建立一个围绕“监管微服务”方法的社区（28）。活动通常涉及：

- 期望每个人都作出演示，进行谈话，推动讨论，或者参加小组或圆桌会议
- 开放的平行会议议程，以确定主题或“项目”（不同群体在该活动过程中可能会进行处理）
- 参加者用实际行动表态，参与任何他们认为可以促进的群体
- 会议反映活动如何做得更好
- 注重社交和共餐

这种风格的活动已有更广泛的吸引力，起初更倾向技术导向。例如，从 2009—2012 年，DevSci 系列研讨会（54）围绕英国机构知识库，建立了一个软件开发社区。近来发生了更多具有相同目标的活动，引起了对数字内容、保存机遇、工具或数据的再利用等问题的广泛讨论。例如，AQUA 和云杉数字保存项目举办为期 3 天的“混搭”活动。从业人员和开发人员被邀请参与。在混搭研讨会论坛上，从业人员作为“集合业主”进行短期发言，展示数据示例，以及它们的保存目标和问题。从讨论中发现任务和挑战，主持人提供 wiki 模板记录整个活动过程中是如何得出解决方案的。



这种方式可以超越保存议题而有所扩展，例如，类似方式已被使用于生命科学数据管理项目 BRISKit 中。在“社区满足和黑客”活动围绕着临床数据管理和治理问题，以寻找适当匹配的开源软件解决方案（57）。

数字管理中心采用“巡回推广”活动召集利益相关机构。这些结合了科研人员主导、影响他们项目的数据管理问题的案例研究，以小组会议形式确定优先事项，以支持改进。虽然没有制订“黑客日”形式的具体解决方案，但这些都是面向规划的机构 RDM 服务。然而，一些机构发展特定学科的 RDM 平台，作为其战略的核心部分。例如，Monash 大学为确保机构 RDM 的可持续性，对科研团体进行嵌入开发。

活动汇集了特定科研单位和服务提供商，以了解他们数据的问题，并找出可行的解决办法。它们可能成为必不可少的，以确保采取了机构 RDM 服务。

6. 下一步计划和未来的挑战

6.1 管理要求

识别提供和开发 RDM 服务所存在的长期挑战，必须确定哪些已确定的要求是通用的，哪些在不借助外部基础设施和服务的情况下跨机构支持，哪些还不够具体不足以为机构提供本地支持。一般来说，这些决定是优先事项和资源的平衡考虑，通过前瞻战略和路线图，促进 RDM 服务向前发展。若干请求可以通过现有专门知识和指导资源，在不变更服务的情况下，就可以被满足；而对那些确实要求服务变更的请求，有些可能意味着需要重新设计软件平台的使用，有些可能需要开发新的应用程序，而有些通过更新本地联机指南或者使用外部采购的工具和资源等方式解决。

在上述方面，RDM 与其他学术支持机构提供的服务没有什么不同。在那些机构用于管理变革的方法同样适用于 RDM。许多组织管理在线服务变革，并支持使用 ITIL（IT 基础架构库）方式进行 IT 服务管理。CMMI（能力成熟度模型集成）方法也广泛用于过程改进。CMMI 特别介绍了需求管理的步骤。商业上可用的工具支持需求管理，例如，协助确保用例和需求陈述，并且追溯到服务的变化等，反之亦然。

虽然 ITIL、CMMI 和其他通用的项目/流程管理办法提供的概念适用于 RDM 服务，这并不一定意味着，RDM 支持服务应充分与其合成，以提供 IT 支持、图书馆服务台，或任何其他“服务台”服务。可能存在有效率进行整合的案例，但同样重要的是，考虑到第 3 节所



本作品采用[知识共享署名-非商业性使用-禁止演绎 3.0 中国大陆许可协议](https://creativecommons.org/licenses/by-nc-nd/3.0/)进行许可。

提到的研究实践等各类机构活动之间的差异。项目经理需要考虑，当地研究文化是否会采取嵌入其他支持领域的服务，并评估一线支持人员无法从其他种类的支持请求中区分出 RDM 问题的风险。

6.2 规模和复杂性的挑战

“……全面的数据管理不仅仅是将数据存放到资料库中。它需要对数据整体语境的理解，与生产数据的以及提供服务的研究者接触，在更广泛的学术交流背景下，将数据作为主要的复合对象。”

无论采取什么方法发现 RDM 项目要求，该方法都需要定期审查，以确保它可以容纳规模和复杂性带来的新挑战。正如第 2 和第 3 节所讲到的，这些都是贯穿整个研究和数据生命周期的挑战。对研究者而言，因为大规模合作的研究机会，以及技术为他们实践带来的复杂性，一种更正规、更协调的研究数据管理方法的价值经常间歇性出现。

所以，机构可以适应他们的支持系统，以反映数字化研究环境的变化，这一点也很重要。这些措施包括学术交流服务的快速增长，以及体制界限外，网络基础设施和研究基础设施。

学术交流正在发生变化，以适应新的数据发布形式，受到同行评审数据所改变的需求所驱动，以及为数据挖掘而提供更多内容，正与新兴的跨学科“大数据”源集成。科研基础设施是重要的，因为他们规模增长以应对这些挑战，并提供新的工具和中介服务，以支持更广泛的访问更多种类的数据。

目前，管理传统学术成果的地方，尤其是期刊文章，一般采用机构 RDM 做法。在英国，研究理事会的政策鼓励这一点。科研质量评估过程继续专注于传统的学术产出，这就需要机构库来管理它们。通常，数据管理政策强调保持数据的真实性，如果没有合适的学科知识库，就用类似的机构知识库作为存放的地方。

这可能比我们想象中的改变得快。例如，数据挖掘可以突出数据知识库内容的价值，刺激机构知识库和基于域的知识库之间的存储用户争夺竞争。此外，已公布数据与软件之间的复杂关系必须解决，以应对日益增长的“重制危机”（62）。当前的 RDM 模型框架数据共享是选择最合适的知识库来存放。但是，我们可以看到：越来越多的要求出现，需要跨越多种知识库、已发表的文章和其他在线资源，来协调记录“什么数据已被用在什么地方”。假定研究结果的标准取决于良好有界的数据集这一做法已经在改变，例如，当科研人员从已经共享的、非常大的数据集得出调查结果时，不能存放在单一的知识库中，那么政策



也需要发生变化。例如，美国国家科学基金会（NSF）的政策已经改变，涉及所有的学术成果，包括数据集信誉良好度（63）。

需求发现的短期挑战是从“一次性、一刀切”模式转向侧重于研究者数据管理计划的访谈，或向更广的范围告知该机构的 RDM 政策。本指南中提到的方法都标识为“好问题”，并且可以在下文中找出他们所使用的例子。还有很大空间，可以进一步完善这些，以帮助研究支持专业人员轻松地识别适用于不同语境的问题。

RDM 支持专业人员需要良好的“要求问题”（该问题适合研究生命周期的不同阶段），并认识到生命周期中的变化。例如，RDM 咨询服务向科研人员提供数据共享支持，直到项目结束，这可能需要比写授权前数据管理计划时，询问更详细的问题。问题也可能不同，这取决于如何以及在何处提问（在线、一对一讨论、小组讨论等）。它们应该符合跨越不同领域的不同成熟级别支持。例如，该机构可能在特定领域中有优势，如元数据咨询，或者使用外部数据知识库的指导并且想要更详细地询问这些领域的要求。

无论形状怎样改变，RDM 环境的变化是必然的。本指南中描述的需求方法，可以帮助组织应对最基本的需求。要真正提供不断完善的服务，供应商需要倾听和应对研究生命周期中的变化，并逐渐适应更复杂的需求。

鸣谢

原著笔者感谢同事们的意见，尤其是 DataONE 的执行主任 Rebecca Koskela，Edinburgh 大学的 RDM 协调员 Kerry Miller。

DCC 的机构参与方案是由监委会资助（2009-2011 年，同时由英格兰高等教育拨款委员会资助）。根据合作协议（#083094），DataONE 是由美国国家科学基金会资助。

