

## 荷兰数据认可印章科学数据仓储认证及启示

韩雪华<sup>1,2</sup> 王卷乐<sup>1,5</sup> 石蕾<sup>3</sup> 高孟绪<sup>3</sup> 王明明<sup>1,4</sup> 李舸<sup>1,4</sup> 王玉洁<sup>1,4</sup>

(1. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101; 2. 中国科学院大学, 北京 100049; 3. 国家科技基础条件平台中心, 北京 100862; 4. 山东理工大学建筑工程学院, 山东淄博 255000; 5. 江苏省地理信息资源开发与利用协同创新中心, 江苏南京 210023)

**摘要:** 科学数据的安全性和可信性对于面向用户的公共存储至关重要。我国国家层面的科学数据政策体系还不完备, 科学数据仓储建设还较薄弱, 在数据仓储评估认证方面还存在空白, 迫切需要了解国际的前沿动态。本文选取荷兰的数据认可印章(DSA)为调查对象, 详细阐述DSA的组织架构、标准架构、评估认证流程, 介绍DSA与国际科联世界数据系统(ICSU-WDS)联合提出的可信数字仓储的核心认证机制。结合DSA数据认证机构的经验, 提出对我国科学数据仓储认证和相关标准制定的思考与启示。

**关键词:** 科学数据管理; 数据认证; 可信性; 数据认可印章; 数据仓储

中图分类号: G250

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2018.01.003

## Identification and Its Inspiration for Netherlands' Data-Seal-of-Approval in Scientific Data Repositories

HAN Xuehua<sup>1,2</sup>, WANG Juanle<sup>1,5</sup>, SHI Lei<sup>3</sup>, GAO Mengxu<sup>3</sup>, WANG Mingming<sup>1,4</sup>, LI Ge<sup>1,4</sup>, WANG Yujie<sup>1,4</sup>

(1. State Key Laboratory of Resources Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101; 2. University of Chinese Academy of Sciences, Beijing 100049; 3. National Science and Technology Infrastructure Center, Beijing 100862; 4. School of Civil and Architectural Engineering, Shandong University of Technology, Zibo 255049; 5. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023)

**Abstract:** The security and trustworthiness of scientific data are of great importance to user-oriented public depository. Because the scientific data policy system at the national level is not complete in China, the development of scientific data repositories is relatively weak. There even exists gap in assessment and certification of scientific data repository, thus it is urgent to learn and understand more advanced world experiences in this

**作者简介:** 韩雪华(1992—), 女, 中国科学院地理科学与资源研究所硕士研究生, 研究方向: 数据共享与知识发现; 王卷乐(1976—), 男, 博士, 中国科学院地理科学与资源研究所研究员, 地球数据科学与共享研究室副主任, 博士生导师, 研究方向: 科学数据共享、地理信息系统与遥感应用(通讯作者); 石蕾(1982—), 女, 国家科技基础条件平台中心副研究员, 研究方向: 科技资源管理; 高孟绪(1982—), 男, 博士, 国家科技基础条件平台中心副研究员, 研究方向: 科技资源管理与共享; 王明明(1993—), 男, 山东理工大学硕士研究生, 研究方向: 地理信息系统与科学数据管理; 李舸(1993—), 男, 山东理工大学硕士研究生, 研究方向: 地理信息系统与科学数据管理; 王玉洁(1992—), 女, 山东理工大学硕士研究生, 研究方向: 地理信息共享。

**基金项目:** 国家科技基础条件平台专项课题“科学数据管理与开放共享的模式与方法研究”(2015DDJ1Z11); 国家科技基础条件平台专项课题“科学数据汇聚的模式与方法研究”(2017DDJ12215); 中国科学院学部咨询评议项目“前沿与交叉学科科研信息化发展战略研究”(O7M74940F4)。

**收稿时间:** 2017年7月31日。

field. The paper selects Data Seal of Approval (DSA) as research objective. DSA's organization structure, standard architecture, assessment process, core certification and the mechanisms for trusted digital repositories that set up by DSA and WDS are introduced and discussed. Finally, combined with the experience of DSA, the paper provides suggestions for certification for scientific data repositories and trustworthy standards in China.

**Keywords:** scientific data management, certification, trustworthiness, DSA, data repositories

## 0 引言

科学数据仓储是开展科学数据汇聚、管理和共享服务的重要基础设施。科学数据仓储的核心目标是确保数据的真实、可靠、完整和可用性，国际上将数据仓储具备的这4种特性称之为“可信”。可信数字仓储(Trusted Digital Repository, TDR)的概念最早由美国研究图书馆协会(Research Library Group, RLG)于1996年提出，2002年进一步明确了可信数字仓储的属性及特点。RLG和保存与获取委员会<sup>[1]</sup>将可信数字仓储描述为“一个可信数字仓储的任务为现在和将来提供可靠的、长期可访问的、在其管理社区内的数字资源”。而“认证”是指由国家认可的认证机构证明一个组织的产品、服务、管理体系符合相关标准、技术规范或其强制性要求的合格评定活动<sup>[2]</sup>。科学数据仓储的认证即指由认证机构证明数据仓储管理能力和可信性(trustworthy)，既要得到数据提供方的信任，获得存放数据的机会，提升数据仓储长期吸引、汇聚、保存数据资源的能力，又要得到用户对仓储的信任，能够放心地使用这些数据，提升该数据仓储的应用水平和社会影响力<sup>[3]</sup>。

目前，已经成立的可信数据仓储认证标准或机构有：荷兰皇家科学院下属的数据认可印章(DSA)、美国联机计算机图书馆中心(OCLC)发布的《可信数字仓储审核与认证：指标体系与核查表》<sup>[4]</sup>、ISO发布的数字档案馆认证国际标准《可信任数字馆藏的审计和认证》(ISO 16364)等<sup>[5]</sup>。2009年，国际科联世界数据系统与荷兰数据认可印章DSA合作建立了对可信数字仓储的核心认证机制(the DSA and WDS certification)<sup>[6]</sup>，共同开展对科学数据仓储机构的评估认证工作。

本文拟介绍国际数据机构推荐认可的DSA的发展历程，阐述其针对数据仓储的评估指南和认证体系，提出其理念、经验和发展特点，为我国开展科学数据仓储可信性标准及认证工作提供借鉴和启示。

## 1 DSA的组织管理系统

### 1.1 DSA概述

2005年，荷兰皇家科学院和荷兰科学研究组织共同建立了数据归档和网络服务并赋予该项目数据发展认可印章的任务，以确保存储数据能被发现、理解和使用。2009年，DSA被移交给一个国际理事会，即DSA Board，并由该理事会管理和继续发展评估指南和同行评审流程<sup>[7]</sup>。

DSA由DSA社区(DSA Community)、DSA联合会(DSA General Assembly)和DSA理事会(DSA Board)三部分组成。

DSA社区由拥有DSA认证仓储的所有机构组成，截至2016年12月31日，DSA共有60个社区成员，其中欧洲有50个，美洲有7个，非洲有1个，全球性组织有2个，如图1所示。目前，DSA认证机构主要分布在欧美国家，亚洲地区被认证机构为0。

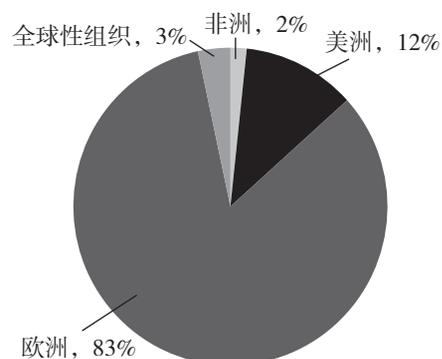


图1 DSA社区成员各大洲占比

DSA联合会由DSA社区中选举产生。联合会选举产生DSA理事会,在必要时向理事会提供建议。联合会成员每年最多进行3次同行评审,以确保DSA在团体主导下的可持续发展。现任DSA联合会由ICPSR、UKDA等机构的21个成员组成<sup>[8]</sup>。

DSA理事会处理DSA社区的日常事务、管理和监督DSA评估流程并负责召集联合会活动。DSA理事会由6~8个成员组成,每个偶数年从联合会机构代表中选举产生,任期两年。理事会在任期内须考虑DSA指南和规程的修订,所有变动在当前印章周期结束时提出<sup>[9]</sup>。2016年1月1日新任命的理事会成员分别来自斯特拉斯堡天文数据中心(Strasbourg Astronomical Data Center, CDS)、DANS、莱布尼茨学会社会科学研究所(GESIS-Leibniz Institute for the Social Sciences)、都柏林大学(University College Dublin)、芬兰社会科学数据档案(Finnish Data Archive, FSD)、UKDA、CINES和MPI。

## 1.2 DSA标准架构

### 1.2.1 标准指南概述

DSA利用可信数字仓储的核心认证机制对科学数据仓储机构进行评估认证,通过科学数据仓储认证证明数据仓储管理能力和可信任性。DSA主要是通过规定的评估指南来检验数据生产、存储、使用和复用方面的质量,将符合指南要求作为DSA理事会授予数据仓储数据批准印章的前提。评估指南内容是依照现有国家级或国际上的数字化数据存储指南来制定的<sup>[10]</sup>。评估指南的基本原则有以下5条:第一,数据可从网上发现;第二,在个人信息和知识产权相关法规的允许下,数据是可获取的;第三,数据以可利用的格式提供;第四,数据是可靠的;第五,数据可通过永久标识引用<sup>[11]</sup>。评估指南和5条原则主要针对3个利益相关者,即数据生产者、数据仓储和数据用户。数据生产者对数字化数据的质量负责;数据仓储对数据管理(数据保存及可用性方面的质量)负责;数据用户对数据使用质量负责。

### 1.2.2 指南详细内容

DSA评估指南主要包括16条,分为三大部分:一是针对研究数据的质量;二是针对数据仓储的机构与程序;三是针对研究数据的使用质量<sup>[12]</sup>(表1)。

### 1.2.3 DSA-WDS核心认证评估指导

世界数据系统(WDS)的主要任务是开展科学数据和信息的收集、交换和服务,同时应国际科学联合理事会(ICSU)的要求,承担了一系列重要的国际科学计划申明要保存的数据的管理<sup>[13]</sup>。WDS和DSA于2009年开始合作建立可信数字仓储核心认证机制(Core Trust Seal),该认证在简单性、健壮性和需要完成的工作之间取得平衡。2014年年初,成立了联合工作组来协调简化认证选项阵列,改进和激发科学数据的核心认证服务。WDS-DSA标准和认证理事会通过与欧洲研究数据联盟(the European plug-in of the Research Data Alliance)、数据归档和网络服务(DANS)的合作项目,正在研究建立一个评估指导系统,该系统依据16项指南,提出相应的16项评估指导。

评估指导1:仓储要把保存数据、保证数据可访问作为明确的使命。

评估指导2:仓储要制定并维护用于数据访问和使用的许可,并监督数据使用者遵从相关许可。

评估指导3:仓储要制定持续性计划,确保仓储内的数据保存和使用的可持续性。

评估指导4:仓储保证在数据的创建、访问、使用过程中,遵守相关法律法规和伦理道德。

评估指导5:仓储要具备充足的资金和专业的工作人员,制定明确的管理制度有效地践行使命。

评估指导6:仓储要采取相应机制确保专家指导和反馈的持续性。

评估指导7:仓储必须确保数据的完整性和真实性。

评估指导8:仓储保存的数据和元数据必须符合相关标准,确保数据对用户的易懂性和关

表 1 评估指南详细内容

序号	指南内容	指南侧重点
指南 1	DSA 要求数据生产者在将数据存入数据仓储时，提供充足的信息以供他人评估数据的质量并遵守相关的学科和伦理规范	第 1—3 项主要涉及研究数据质量。研究数据质量主要包括：科学与学术价值、研究数据及辅助信息的存储格式、针对研究数据的说明文档（元数据或是背景资料）
指南 2	DSA 要求数据生产者按照数据仓储推荐的格式提供数据	
指南 3	DSA 要求数据生产者按照数据仓储的要求在提交数据时一并提交元数据	
指南 4	要求数据仓储拥有并发布数字化存储领域的明确任务和使命	第 4—13 项规定了针对数据仓储的机构与程序的要求。数据仓储的质量取决于以下两点：一是数据仓储组织框架的质量；二是数据仓储技术基础设施的质量。DSA 认为，可信数字仓储机构应拥有健全、可靠的财政基础、组织基础和法律基础
指南 5	要求数据仓储进行尽职调查以确保符合法律法规和契约，必要时还应包括以人为对象的保护规定	
指南 6	要求数据仓储按照有据可查的流程和步骤管理数据存储	
指南 7	要求数据仓储有针对数字资产长期保存的规划	
指南 8	要求仓储根据贯穿数据生命周期的明确工作流程进行存储	
指南 9	要求数据仓储对数据生产者就数字对象的获取和可获得性承担责任	
指南 10	要求数据仓储使用户能够发现和利用数据，并能以一种永久的方式引用数据	
指南 11	要求数据仓储确保数字对象和元数据的完整性	
指南 12	要求数据仓储确保数字对象和元数据的可靠性	
指南 13	要求仓储的技术基础设施明确支持国际认可的存储标准（如 OAIS）中描述的任务和功能	
指南 14	要求数据用户遵守数据仓储制定的规程	第 14—15 项主要针对数据用户使用研究数据使用的质量
指南 15	要求数据用户遵守并同意相关领域普遍认可的交换与合理使用知识与信息的行为准则	
指南 16	要求数据用户尊重数据仓储有关数据使用的协议	

联性。

评估指导 9：仓储要采用文件化的程序管理数据的归档和存储。

评估指导 10：仓储要承担长期保存数据的责任，并制定相关的规划与文件。

评估指导 11：仓储要具备相应的专业知识来处理技术数据和元数据，并提供足够的信息供用户进行数据质量评估。

评估指导 12：仓储的数据归档要根据明确清晰的数据周期工作流进行。

评估指导 13：仓储允许用户发现数据，并以合理方式持久地引用。

评估指导 14：仓储允许重用数据，并确保有标准的元数据用于对数据的理解和使用。

评估指导 15：仓储在支持的操作系统和其他核心基础设施软件上可以运行良好，并使用适合的硬件和软件技术服务指定社区。

评估指导 16：仓储的技术基础设施要能够为其数据、产品、服务和用户提供保护。

## 2 DSA 评估及认证

### 2.1 评估标准

DSA 的 16 项评估指南参考以下 5 个级别（0—4）开展评定。0 代表不适用；1 代表未考虑；2 表示有理论性的概念；3 表示在实施阶段；4 表示完全遵守并实施。在当前的 16 项评估指南中，指南 1、指南 2、指南 7、指南 8、指南 10、指南 11、指南 12 和指南 13 等 8 项评估指南最低要求为“3”。另外的指南 3、指南 4、指南 5、指南 6、指南 9、指南 14、指南 15 和指南 16 等 8 项评估指南最低要求为“4”。如果一个数据仓储符合指南第 4 条到第 13 条，其数据生产者达到指南第 1 条到第 3 条的要求，数据用户达到指南第 14 条到第 16 条的要求，则被认定为可信的数字仓储（TDR）<sup>[14]</sup>。

### 2.2 认证流程

DSA 可信赖认证主要包括两种认证方式：自我评估和第三方认证（包括主管机构或资助

机构的审查)。通过认证后可在其网站上展示 DSA 标识。DSA 评估认证过程主要包括 3 个环节(图 2),即自评、同行评审和印章展示。DSA 提供相应的评估认证在线工具。

在自评阶段,申请者首先在指定页面提交申请表,收到用于登录在线工具的用户名、密码和登录链接后启动自评。申请者登录在线工具对 16 项指南逐项做出解释说明,并提供相应材料证明其达到了每项指南的要求并说明符合程度。自评通常需在 3 个月内完成。

自评提交后进入同行评审阶段。同行评审由 DSA 理事会指定人员完成,人员范围包括专业审核人员、理事会成员、相关领域内有资格人士。评估依据主要包括:自评中的陈述说明与指南是否符合、与指南要求的级别是否一致、是否有相应的说明文件等。同行评审通常需在 2 个月内完成。在适当情况下,理事会可采取相应措施保证同行评审在规定时间内完成。同行评审有以下两种结果。

结果一:申请方需补充更多材料。说明申请方自评已被审核,但现有证明材料未能达到 DSA 认证要求。申请方需再次登录,根据审核人员意见,提交修订后的申请和证明材料。修订版申请以及增加的材料继续被审核,直至达到 DSA 的认证要求。该过程可重复多次,若申请方与审核员之间发生分歧,可联系理事会进行沟通。

结果二:通过认证。申请方将收到授予 DSA

认证的系统通知和 DSA 提供的 HTML 代码。可将 DSA 的标识和机构链接展示在其网站页面。同时,DSA 将在官网上发布通知,认定新的 DSA 成员,并公开该仓储在申请过程中提交的证明材料。

DSA 的评估与认证还包括更新与延续环节。DSA 认证是无限期的,但如果要长期保持 DSA 的认证,数据仓储需要定期进行更新维护,以符合 DSA 发布的最新标准,并展示最新的 DSA 标识。

### 3 启示

自 2015 年 8 月 31 日国务院印发《促进大数据发展行动纲要》<sup>[15]</sup>以来,在各个行业领域加强了科学数据基础设施的发展,作为核心基础设施的科学数据仓储如何汲取国际经验,加强其自身的规范化建设是一个急迫的课题。我国科学数据仓储(数据中心)长期存在着分散薄弱、规范不一、难以持久等问题。与科学数据仓储建设薄弱相对应。DSA 作为通用基础的科学数据仓储认证机构,广泛适用于不同背景下的存储机构,包括图书馆、博物馆、档案馆、数字仓储、科学数据仓储等<sup>[16]</sup>,其认证理念、认证过程对于我国科学数据管理的标准化和规范化建设发展具有借鉴和启示作用。

(1) DSA 的分层认证、逐步完善的认证流程对我国建立自主科学数据仓储认证机制具有重要

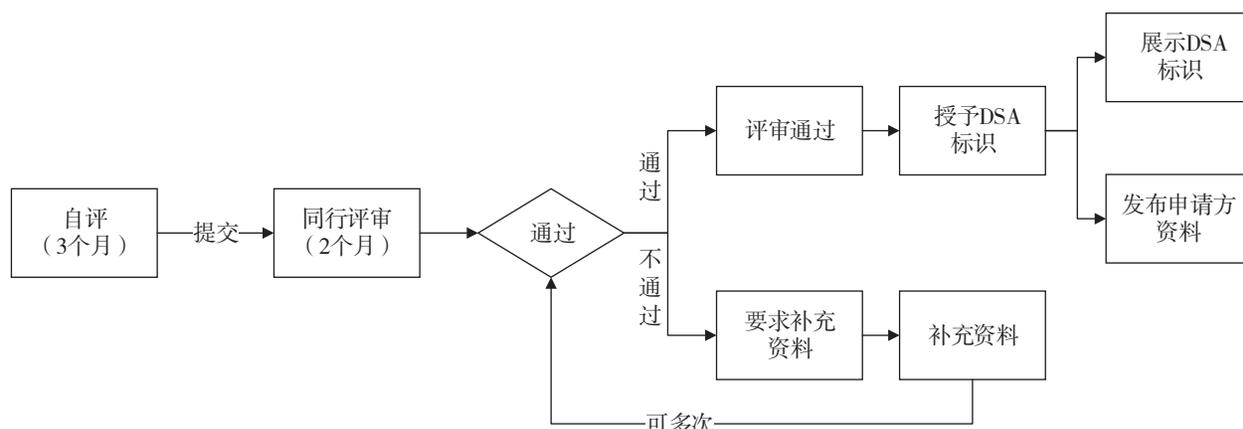


图 2 DSA 评估认证过程

的借鉴意义。DSA 认证过程从机构自评到全面的第三方认证,过程可多次重复进行,使申请方可以逐步提升可信任的等级。这种分层认证、逐步完善的流程在数据存储机构的评估中具有很好的合理性和可行性,有利于数据存储机构系统的逐步完善。可参考 DSA 认证的经验,建立我国自主的科学数据仓储评价机制和认证体系,不仅有利于我国科学数据仓储认证事业的发展,完善我国科学数据政策,而且可针对国内某些国际化程度还不高或者不宜国际化领域的科学数据仓储建设有可依据的认证标准、认证系统。

(2) 发展科学数据仓储认证与评估,可促进我国科学数据管理事业的国际化,提升我国在相关领域的国际影响力。目前 DSA 社区中还没有亚洲相关机构的身影,借助 DSA 认证经验,探索适合我国可信数据仓储建设发展的标准,一方面便于推动和提升我国可信数据仓储的建设发展;另一方面也使我国有更多的机会在国际相关领域发声和交流,促进我国科学数据管理的国际化与影响力。

(3) 科学数据仓储认证机构可有效促进科学数据的管理和发展。科学数据仓储认证机构作为第三方机构,拥有完善的组织架构与标准规范,可从数据质量、仓储机构、数据使用等多方面对科学数据仓储进行监督,加强数据的规范化管理,保护数据产权,有利于科学数据持续发展。同时,可提高数据存储运营的效率,完善存储数据的管理和维护,提升机构的保存能力。为自上而下的国家数据仓储和自下而上的学科数据仓储创造开放、公平的数据公共存储门槛。通过认证机构,可构建科研数据共享与交流的社区,社区内各数据中心共同关注数据中心可信任性的建设,共同推进科学数据管理的发展。

## 参考文献

- [1] OCLC R. Trusted digital repositories: attributes and responsibilities[J]. Mountain View Ca Retrieved, 2002, 24(1): 108-110.
- [2] 国际标准化组织.ISO/IEC17000合格评定.词汇和一般原则[S]. 2004.

- [3] 杨璐. 可信数字仓储认证: 必要性、标准与主体[J]. 北京档案, 2014(5): 19-22. DOI: 10.3969/j.issn. 1002-1051.2014.05.007.
- [4] OCLC and the center for research libraries: trustworthy repositories audit and certification: criteria and checklist (2007)[EB/OL].[2017-05-07]. [http://www.crl.edu/sites/default/files/d6/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf).
- [5] ISO. ISO 16363: space data and information transfer systems-audit and certification of trustworthy digital repositories[S]. 2012.
- [6] MOKRANE M, HUGO W, HARRISON S. WDS/DSA certification-international collaboration for a trustworthy research data infrastructure[C].EGU General Assembly, 2010.
- [7] 伏安娜, 张计龙, 殷沈琴. DSA 对我国科学数据共享中可信赖性标准制定的启示[J]. 图书馆杂志, 2016(10): 69-76.
- [8] GARRETT J R. Task force on archiving of digital information[M]. Virginia: Corporation for National Research Initiatives, 1995.
- [9] Data seal of approval: certification for sustainable and trusted data repositories [EB/OL].[2017-05-07].<https://www.datasealofapproval.org/en/news-and-events/news/2014/10/3/dsa-overview/>.
- [10] Community[EB/OL].[2017-05-07].<https://www.datasealofapproval.org/en/community/>.
- [11] DSA general assembly formed [EB/OL].[2017-05-07]. <https://www.datasealofapproval.org/en/news-and-events/news/2015/8/17/dsa-general-assembly-formed/>.
- [12] Data seal of approval (DSA): community & regulations [EB/OL].[2017-05-07].[https://www.datasealofapproval.org/media/filer\\_public/2013/09/27/dsa-regulations\\_2013.pdf](https://www.datasealofapproval.org/media/filer_public/2013/09/27/dsa-regulations_2013.pdf).
- [13] 王卷乐, 孙九林. 世界数据中心(WDC)回顾、变革与展望[J]. 地球科学进展, 2009, 24(6): 612-620. DOI: 10.3321/j.issn: 1001-8166.2009.06.006.
- [14] Data seal of approval: guidelines[EB/OL].[2017-05-07] [https://www.datasealofapproval.org/media/filer\\_public/2013/09/27/-guidelines\\_2014-2015.pdf](https://www.datasealofapproval.org/media/filer_public/2013/09/27/-guidelines_2014-2015.pdf).
- [15] 中华人民共和国国务院. 促进大数据发展行动纲要[J]. 成组技术与生产现代化, 2015, 32(3): 51-58.DOI: 10.3969/j.issn.1006-3269.2015.03.012.
- [16] 吴振新. 数字资源长期保存可信赖认证研究发展综述[J]. 中国图书馆学报, 2015, 41(3): 114-126.