



# Digital Curation 和 Digital Preservation 之概念辨析\*

张智雄 吴振新 刘建华 郭红梅

(中国科学院国家科学图书馆 北京 100190)

**摘要:**【目的】对 Digital Curation 和 Digital Preservation 这两个相关概念进行辨析。【文献范围】以历史上与 Digital Curation 和 Digital Preservation 相关的重要文献和重要报告为基础,并调研分析 DCC、JISC、ARL 等主要机构对两个概念的各种定义。【方法】基于两个概念形成的历史,分析两个概念产生和发展的深层原因;基于当前主要研究团体和个人对两个概念的多个定义,分析两个概念的不同;进一步从 8 个方面入手,多角度对这两个概念的不同点进行对比分析。【结果】Digital Curation 和 Digital Preservation 虽然都可用于数字保存和管理,但它们的确是两个不同的概念,在很多方面有较大的差别。【局限】尚需对 Digital Curation 和 Digital Preservation 所包括的具体工作内容进行较为细致的对比分析。【结论】Digital Curation 和 Digital Preservation 是两个不同的概念,但二者互补,图书馆应当采用一种更为积极主动的方式来推进数字保存和管理。

**关键词:** Digital Curation Digital Preservation 概念辨析

**分类号:** G250

关于数字保存,英文中有两个对应的词,即 Digital Curation 和 Digital Preservation。由于语言文化等因素,国内学者对 Digital Curation 和 Digital Preservation 二者的区分不像英语学者那样容易理解到位。什么是 Digital Curation? 为什么有了 Digital Preservation,还要创造出一个 Digital Curation? Digital Curation 和 Digital Preservation 有什么区别? 这是笔者经常被问到的问题。

本文以历史上相关的重要文献为基础,对 Digital Curation 和 Digital Preservation 这两个概念进行辨析,在分析相关定义的基础之上,从多个角度对 Digital Curation 和 Digital Preservation 的不同点进行了较为细致的对比,以期能够较为清楚地分辨出 Digital Curation 和 Digital Preservation 的区别。对于 Digital Curation 应当翻译成什么中文术语这一问题,由于不

是本文的重点,因此不进行讨论,有些地方为了行文的需要,权且译为“保存管理”或“保管”。为了避免不必要的争议,本文更多的地方直接应用 Digital Curation 和 Digital Preservation 指代两个概念,不进行翻译。

## 1 两个概念的主要形成历史

### 1.1 Digital Preservation 概念的形成过程

Digital Preservation 概念的提出主要是为应对数字信息容易被损坏、篡改,或因技术退化而变得不可用的问题。

数字技术为信息的利用带来了很大便利,诸如可复制、易传输、大量携带等,但由于数字信息本身具有脆弱性,如容易被删除、盗取、修改、失真;需要

收稿日期: 2013-11-29

\*本文系国家自然科学基金项目“数字资源长期保存技术的研究与实践”(项目编号: 09FTQ005)的研究成果之一。

依赖于其他对象或其他环境,如依赖于技术、系统、标准、软件、上下文、组织、经济等;并且时刻受到数字技术快速退化的威胁,如媒体、硬件、软件、格式的退化等。因此,数字信息与印本信息相比,更容易被损坏、篡改,或因技术退化而变得不可用。

自 20 世纪 90 年代以来,档案界和数字图书馆界已经认识到数字资源的保存和印本资源的保存相比,需要解决很多更为复杂的问题,一些研究人员开始着手从数字保存的挑战、概念、技术、方法等方面进行探讨<sup>[1-4]</sup>。

1994 年 12 月,美国保存和获取委员会(The Commission on Preservation and Access)和研究图书馆联盟(The Research Libraries Group)联合组织了一个专项工作组,研究分析在将来无限长的时间内,采用什么样的技术方法来确保能够持续利用电子信息<sup>[5]</sup>。

1996 年,上述专项工作组发布了名为保存数字信息(Preserving Digital Information)的报告<sup>[6]</sup>。由于这个报告较为系统地论述了数字保存的概念、需求及挑战等重要问题,因此很多学者都认为这一报告是 Digital Preservation 研究的奠基性文献<sup>[7]</sup>。

这份报告强调尽管信息的丢失与存储介质的损坏密切相关,但是软件和硬件技术的快速退化已经成为信息丢失的主要问题。报告指出数字存档系统与数字图书馆系统的本质区别在于数字存档系统能够长期地存储和访问所存储的资源。报告中对实现数字存档的数字仓储提出了明确的规范和要求,如数字仓储需要通过规范的认证程序认证,数字仓储具有失效保险机制来保证系统安全等,这在后人看来是可信数字仓储(Trusted Digital Repositories)概念的雏形。

报告认为数字信息的保存关键在于实现数字信息完整性的保存,并从 5 个方面对保存数字信息的完整性进行了论述。它们分别是<sup>[6]</sup>:

(1) 实现数字内容的完整保存。可进一步区分为比特层的完整、格式和结构层的完整、呈现层的完整和知识层的完整等。

(2) 保证数字信息稳固不变。要求被保存数字信息作为一个完整和独立的单元来保存,保证被保存数字信息处于稳固不能随意改变的状态。

(3) 数字信息可唯一标识。要求被保存的数字信息可以被唯一标识,能够被人引用。

(4) 数字信息的来源可以追溯。要求记录被保存数字信息的来源以及对这一数字对象所进行的一系列保存管理操作。

(5) 保存数字信息的存活环境。包括技术环境、与其他对象的关联信息、通讯环境和更广泛的社会环境。

报告认为,与“翻新”相比,“迁移”是一个更广泛和更富有内涵的概念来揭示数字保存的活动范围,并对长期保存的“迁移”的概念、策略和方法进行了深入的分析。

Preserving Digital Information 报告中的概念及相关建议,在数字保存领域产生了重要影响,推动了数字保存研究和实践的兴起。20 世纪 90 年代末期,一些重要的数字保存项目和原型系统在此后开始出现<sup>[8]</sup>。如 1998 年 4 月到 2002 年 3 月,英国的三个机构(Universities of Leeds、Oxford 和 Cambridge)进行了 CEDARS 项目研究<sup>[9]</sup>;欧洲国家图书馆联合会常设委员会在 1998 年和 2000 年间进行了 NEDLIB 项目研究<sup>[10]</sup>;荷兰国家图书馆也在 21 世纪初开始了 e-Depot 系统的研究和建设<sup>[11]</sup>。而在美国,2000 年 12 月,美国国会通过了关于发起国家数字信息基础设置和保存计划(NDIIPP)的法案,开始了美国国家层面的数字保存研究<sup>[12]</sup>。

2002 年 1 月,空间数据系统咨询委员会(CCSDS)发布了“开放存档信息系统参考模型”(OAIS)<sup>[13]</sup>,从整体框架上对开放存档信息系统所包括的功能实体、保存信息、信息包类型、组织管理、保存规划等进行了明确,成为后来数字保存活动具体实施的指南。2002 年 5 月,基于当时正在出现的 OAIS 标准,RLG 和 OCLC 发布了“可信数字仓储:属性及职责”的研究报告<sup>[14]</sup>,对大型、异构、要求能够持续发展的数字保存仓储的特征进行了描述,对可信数字仓储的概念进行了界定,明确了可信数字仓储在管理、组织、财务、技术流程、系统安全等方面的要求,成为后续数字保存仓储审计和认证的指导文献。

OAIS 参考模型和可信数字仓储的发布,意味着数字保存(Digital Preservation)的理论基础已经成熟,数字保存的概念在认识上基本得到统一。

## 1.2 Digital Curation 概念的形成过程

在 Digital Preservation 概念逐步成熟的时候,科技界中采用了“Data Curation”来泛指对科学数据的管

理。2001年一个新的术语“Digital Curation”被提出，它为 Curation 赋予了更多的含义。

术语“Digital Curation”首次出现在一个名为“Digital Curation: Digital Archives, Libraries and e-Science Seminar”的研讨会上。这一研讨会由数字保存联盟和英国国家空间中心组织，于2001年10月19日在伦敦召开<sup>[15]</sup>。据 Beagrie<sup>[16]</sup>认为，这次会议上出现的新术语“Digital Curation”是组织者经过精心挑选的结果。这一新术语不但意味着要对资源或数据进行保存和管理，同时也在一定程度上意味着对数据进行增值管理，实现知识增长。

此后 e-Science 界对“Digital Curation”这一概念进行了更为深入的研究讨论。2002年，时任英国联合信息系统委员会（JISC）研究支持委员会（JCSR）主席的 Tony Hey 教授组建了一个专项工作组，以构建英国原始研究数据（特别是科学数据）的保存管理战略。2002年11月26日，这一专项工作组举行了一天的头脑风暴会，重点对“Digital Curation”的相关问题进行探讨。从专项工作组提供的会议报告<sup>[17]</sup>来看，虽然在会上大家最终没有对“Digital Curation”的定义达成统一认识，但均对“Curation”这个新术语表示认可，认为这一术语囊括了 e-Science 界所要应对的问题，虽然这些问题与 Digital Preservation 所要解决的问题经常交叠。

根据会议报告，当时来自不同领域的会议专家从多个角度对“Curation”的定义发表了意见。例如，Tony Hey 沿用当时英国研究委员会主任 John Taylor 的提法，认为 Digital Curation 与 Digital Preservation 的区别在于 Digital Curation 活动目标是保证数字数据不仅仅按原来的方式被应用。Seamus Ross 从博物馆的角度出发，认为 Curation 包括三个核心概念，即保护（Conservation）、保存（Preservation）和获取（Access）。Alison Allden 认为 Curation 意味着要对数字信息进行积极管理，包括提前进行数据生成的有效规划，并认为实现数据的重用是问题的关键，如果数据需要能够被重用，它们必须经过专门的处理。Rolf Apweiler 则认为，Curation 是人们对数据进行增值的一个过程。Jeremy Frey 持同样的观点，主张 Curation 本质上是一个研究过程，它实现对数据的管理、提升和增值等。

2003年，由 Lord 等<sup>[18]</sup>为 JISC 研究支持委员会完

成的 e-Science Curation Report 发布。这一报告对 Curation 的概念进行了更为深入的阐述。

这一报告的作者认为，他们对术语 Curation 的理解是建立在英文单词“Curator”的基础之上的。

“Curator”是指那些为公共利益而保管某些东西的人，这些东西的价值通常需要通过 Curator 的工作而体现出来。报告的作者因此强调，采用 Curation 有两个重要的用意，一方面明确表达为了公共利益而推动数据共享的政策，另一方面认为数字保存管理员（Digital Curator）应当在保存内容的管理上持一种积极主动的姿态，以推进保存内容的增值。

报告也指出，在基因和蛋白质领域中，“Curation”一词已经作为一项重要的数据库管理工作而存在，被用来特指对相关数据进行标注以及对这些标注进行管理的工作。而报告的作者认为，他们现在所提的“Digital Curation”概念，不仅仅局限于对存档内容的标注和管理，而更在于体现在整个科研环境下实现对科研记录的保存管理，需要将原始研究数据的管理放在一个日渐递增的科研链中去考虑，甚至从科研的生命周期中去看科学数据的保存和利用。

由于 Lord 等也意识到“Digital Curation”在当时环境下还是一个新概念，因此在他们的报告中，对“Curation”、“Archiving”和“Preservation”三种主要活动提出了“实用性定义”（Working Definitions）。基本内容如下：

（1）Curation：是指在数据产生之时就进行的对数据实现管理和促进数据利用的活动，这一活动确保数据符合当前的应用目的，并且能够被发现和重用。对于动态的数据集，它意味着要不断对数据进行丰富和更新，使其能够符合应用目标。更高层面的 Curation 不但包括数据的管理，还包括对相关数据标注内容及其他数据内容关系的管理。

（2）Archiving：是一种 Curation 行为，它确保数据合理地选择和存储，能够被访问，并且数据的逻辑和物理完整性随着时间的流逝能够得到维护，保障数据的安全和真实可信。

（3）Preservation：是在 Archiving 中的一种行为，它要求对特定的数据进行维护，不论时间如何流逝，这些数据始终能够被存取和理解，并不会随着技术的更新而变得不可用。

这一实用性定义,虽然还有许多地方值得商榷,但 Digital Curation 的确以一个不同于“Digital Preservation”的概念被提了出来,并且得到了广泛的认同。2004 年 3 月,在 e-Science Core 项目的支持下,英国 DCC (Digital Curation Centre)<sup>[19]</sup>成立。2005 年,第一届国际 Digital Curation 会议 (International Digital Curation Conference)<sup>[20]</sup>召开。2006 年,国际 Digital Curation 期刊 (International Journal of Digital Curation) 发布<sup>[21]</sup>。近几年来,更不断有以 Digital Curation 命名的机构出现,如雅典研究中心在 2007 年成立希腊数字保管单元 (The Greek Digital Curation Unit, DCU)<sup>[22]</sup>;而在 2010 年,加利福尼亚大学成立了加州大学保管中心 (University of California Curation Center, UC3)<sup>[23]</sup>;约翰·霍普金斯大学成立了数字研究和保管中心 (The Digital Research and Curation Center)<sup>[24]</sup>;多伦多大学成立了数字保管研究所 (The Digital Curation Institute)<sup>[25]</sup>。

## 2 两个概念的定义辨析

随着研究的不断深入,相关学者不断从不同角度对“Digital Preservation”和“Digital Curation”两个概念进行定义,进一步丰富“Digital Preservation”和“Digital Curation”的理论研究基础。为了从多方面对“Digital Preservation”和“Digital Curation”两个概念进行深入辨析,现将一些有影响的概念定义分析如下。

### 2.1 英国 DCC 关于两个概念的辨析

英国 DCC 是第一个以“Digital Curation”命名的研究机构,它关于什么是“Digital Curation”和什么是“Digital Preservation”的提法,直接影响着这一领域的研究动向。

关于什么是“Digital Curation”,英国 DCC 有着多个版本。

Giaretta<sup>[26]</sup>对“Digital Preservation”和“Digital Curation”相关的多个概念进行了区分。他认为需要区分 Digital Preservation、Digital Information Preservation、Curation、Data Curation 和 Digital Curation 等多个概念。他认为 Digital Preservation 的目标应当是、也只能是简单保证数字对象的比特流和字节流在将来可以被访问; Digital Information Preservation (数字信息保存) 是依据 OAIS 标准的、不仅仅是比特流和字节流的保

存,更重要的是数字内容的保存,它按照 OAIS 要求,要保证数字内容可获取、可理解,并且可应用; Curation 是一个通用的概念,主要指关照某些东西; Data Curation 是指管理并实现对数据的增值; 而 Digital Curation 是指实现对数字数据的管理和增值,它隐含地意味着从现在的数据中生成一些新的数据,以使这些数据更有用并符合使用目标。

DCC 在其宪章及原则声明<sup>[27]</sup>中对“Digital Curation”的定义是被很多研究人员接受的定义。这一定义认为 Digital Curation 是对可信赖的数字研究数据进行维护并使其增值的系列活动,它包括了整个研究周期内对数据所实施的各项主动管理活动。

而 DCC 网站上“Digital Curation”栏目下关于“Digital Curation”的定义<sup>[28]</sup>,则相当于是对 DCC 宪章中定义的进一步说明和解释。这一定义认为“Digital Curation”包括了在研究数据的整个生命周期内,对研究数据进行维护、保存并且实现增值的一系列活动。对研究数据的主动管理可以排除威胁这些数据长期研究价值的各种因素,降低因数字技术退化而带来的风险。同时这些活动能够使保管在可信赖仓储中的数据更广泛地被英国的研究团体所共享;在减少重复生成数据的同时,这些活动能够提高现有数据的长期利用价值,使这些数据可以支撑未来高水平的研究。

2008 年 DCC 发表了一份名为“什么是 Digital Curation”简报<sup>[29]</sup>,进一步对“Digital Curation”进行定义解释。这份简报认为“Digital Curation 是对数字数据进行管理和保存以使其能够被长期应用的活动”。所有从规划数据产生就开始的数据管理,在数字化和数字内容组织整理方面的最佳实践,以及为了确保这些数据在未来可以获得,能够方便地被发现、被重用的活动都是 Digital Curation 的重要组成部分。

简报认为 Digital Curation 同样也包括对大数据集的日常管理,例如那些保证数据可以被检索和持续读取的各项工作。在简报中,作者认为, Digital Curation 是适用于描述从信息的产生到结束的整个生命周期内所涉及到的各种职业活动的概念,例如数字化加工人员、元数据创建者、投资人、政策制定者以及仓储管理员的工作都可以认为是 Digital Curation 工作。在简报中,虽然没有将 Digital Curation 和 Digital Preservation 进行对比,但作者明确地认为, Digital

Curation 工作具有短期的益处和长期的价值,即 Digital Curation 在确保数据长期可持续性的同时,能够为数据的创建者和用户带来即时的价值。

## 2.2 JISC 关于两个概念的辨析

英国联合信息系统委员会(JISC)是英国 e-Science 项目的主要支持机构,如前所述,它是 Digital Curation 概念的重要锻造者之一,也是 DCC 的创建者之一。对于 Digital Curation 和 Digital Preservation, JISC 有着较为清晰的区分。

2003 年 7 月,在关于征询成立 DCC 意见的 JISC 通讯中, JISC 对 Digital Curation 进行了定义,并阐述了其与 Digital Preservation 的不同点。JISC 通讯<sup>[30]</sup>认为“Digital Curation 是一系列为当前和未来的数据应用而在整个生命周期内维护和利用数字数据和研究成果的活动”。JISC 认为, Data Curation 可被泛指是对大量的可信数据进行维护,使这些数据能够反映当前相关领域的研究状况;而 Digital Curation 活动隐含地包括了数字存档和数字保存的意思,但更为重要的是,它包括了所有为更好地实现数据生成和数据管理而做的各类工作,特别是对数据进行增值以生成新的信息和知识的活动。JISC 通讯认为, Digital Curation 不仅仅是存档和保存数据源及其相关的元数据,它还需要很好地处理数据创建者、数据提供者、数据存档者,以及数据消费者之间的交互关系。

2006 年 11 月, JISC 发布了名为“Digital Preservation: Continued Access to Authentic Digital Assets”的简报(Briefing Paper)<sup>[31]</sup>。在这份简报中, JISC 较为明确地对 Digital Preservation 和 Digital Curation 进行了区分。

JISC 简报认为, Digital Preservation 是在任意长的时间段内,只要认为数字对象还有价值,就要确保这些真实可信的数字对象能够持续和可靠地被访问而采取的一系列活动。Digital Preservation 不仅仅包括技术活动,它同时还包括对数字资源的管理,让数字资源能够存活下去的所有政策和组织因素。JISC 简报认为, Digital Preservation 要解决的问题主要有:由于技术退化,数字对象不能访问的问题;由于环境不存在,数字对象不能呈现的问题;由于数字对象和环境的变化,导致数字资源的真实可信性和完整性受到损害的问题。因此 Digital Preservation 要解决数字对象的可靠

性、可信性以及可重用性问题。

在这份简报中, JISC 专门对 Digital Curation 进行了定义和解释。简报中的定义与 DCC 宪章中对 Digital Curation 的定义较为相似,但更为具体。它认为 Digital Curation 是所有为将来和现在的应用目的,而对一系列可信赖的数字信息进行维护和使其增值的所有活动,特别是在整个生命周期内,对数据进行的主动管理和评估活动。

JISC 简报认为, Digital Curation 的概念基于 Digital Preservation,但 Digital Curation 更强调通过标注和持续的资源管理而使数据增值和增知的优势。尽管两者都强调要对数字资源进行管理,使数字资源不随着时间的流逝而发生重大改变(或仅仅是受控的改变),但 Preservation 只是 Curation 多种活动中的一种。

## 2.3 ARL 关于两个概念的辨析

2011 年 3 月, ARL 发布了由 Walters 等<sup>[32]</sup>撰写的、名为“新时代的新角色:为了保存而进行数字保管”的报告。报告中对 Digital Curation 的定义吸收了 DCC 关于 Digital Curation 概念的提法,认为 Digital Curation 是在数字信息的生命周期内,对数字信息进行管理并使其增值的活动,它包括在数据内容创建的时候就采取的活动。而对于 Digital Preservation, ARL 报告采纳了 Digital Preservation Coalition 的提法<sup>[33]</sup>,认为它是一系列为保障在未来特定长的时间段内可以持续使用数字资源的管理活动。

ARL 报告认为, Digital Curation 和 Digital Preservation 是两个高度互补的概念,可以明确区分出来的是 Curation 活动是以满足当前和未来的应用需求为目标的。报告中多处强调要促进两者有交叉结合,特别是研究型图书馆应当通过 Digital Curation 活动,来促进共享数字记忆的保存。

ARL 报告虽然有意不对 Digital Curation 和 Digital Preservation 做严格的区分,但从整个报告来看, ARL 作者显然更多地利用了 Curation 这一词来描述图书馆中一系列更为积极和主动的活动,认为图书馆员应当在数据和信息生命周期的更早阶段就开始介入,图书馆员应当重点关注在知识产生的阶段对数字增值,而不应仅仅在生命周期的结束阶段对信息产出进行管理。报告中引用了 Anna Gold 的话来形象地说明这些新的 Curation 服务和活动,“用丰富和健壮的方式对数

据进行链接以支持数据的重用和集成,为了做到这点,此前需要记录数据的来源,开发知识本体,进行专业标注,并且对数据进行分析。进一步往下游走,还有很多通过可视化、仿真、数据挖掘和建模以及其他知识表现和知识抽取活动支撑的服务”。而对于 Preservation 这一概念,在 ARL 报告中,更主要地作为一个 Digital Curation 活动的目的而出现。引入 Digital Curation 活动的目的在于激活研究图书馆的 Preservation 角色,以达到可以长期存取数字信息的目标。

#### 2.4 Sarah Higgins 对两个概念的辨析

还有很多研究者对 Digital Preservation 和 Digital Curation 两个概念进行了辨析。Sarah Higgins 是其中的一位。

Sarah Higgins 认为, Digital Preservation 和 Digital Curation 是数字保存发展的两个阶段,在文献[7]中,对 Digital Preservation 和 Digital Curation 进行了对比。

他认为,数字保存最初的工作重点是 Digital Preservation,它保证有价值的数字资源能够在技术退化和组织失败之后仍然能够存活。在 Sarah Higgins 看来, Digital Preservation 意味着采用一种被动的方式工作,其中数字资源被封存在一种不可访问的“Dark Archive”(黑箱存档库)中,只能通过一些授权的用户来保证这些资源处于完整和真实的状态。

而近年来,数字保存已转变到了 Digital Curation 阶段。Digital Curation 的工作重心是确保数字资源在其生命周期之内都能得到管理,使其在被需要的时候可以访问到。在这一阶段,应用了元数据,可以使数字资源能够被存取和发现;应用了认证授权机制,创建了审计追踪功能,确保数字资源不被非授权用户存取和改变;数字资源被积极地保存;可以为新的目的应用和重用这些数据资源,并可基于这些数据来创建新的数据。

### 3 两概念的多角度对比分析

在对上述 Digital Preservation 和 Digital Curation 概念分析的基础上,本文从多个角度对 Digital Curation 和 Digital Preservation 两者的不同点进行较为细致的对比分析,概括为 8 个方面,具体如下

所述。

#### 3.1 应对的问题不同

Digital Preservation 的提出主要源于数字信息自身的脆弱性和保存有价值资源的需要。Digital Preservation 主要应对数字信息比纸本信息更容易被损坏、篡改,或因技术退化、组织失效而变得不可用的问题。Digital Preservation 需要将有价值的资源长期保存下去,使后来人可以使用这些资源。

Digital Curation 的提出源于 e-Sciences 应对科研数据洪流的需要。Digital Curation 主要应对如何对科研数据进行有效管理的问题。随着越来越多的科研工作需要基于科研数据完成,也随着越来越多的科研数据如洪流一样生成,e-Sciences 界普遍认识到,需要采用一种主动的管理方式来实现对科研数据的保存和管理,以保证科研数据能够被更广泛地获取和重用。

#### 3.2 行动的目的不同

Digital Preservation 以保证数字信息的生存为目的。Digital Preservation 通过一系列的保存活动,保证有价值的数字资源在未来可以获取、呈现、理解和应用。正如 Caplan<sup>[34]</sup>所认为,从下到上、从较低要求到较高要求,数字保存目的分别是保证数字的可获得性(Availability)、可识别性(Identity)、完整性(Integrity)、持久生存能力(Viability)、可呈现能力(Renderability)、被保存对象的真实性(Authenticity)和可理解性(Understandability)。

Digital Curation 以保证数据能够被科研利用为目的。Digital Curation 实现对数据的保存和管理,并通过主动的活动(如对数据进行标注、评价、选择、转换等)使数据增值。Digital Curation 需要在数据的整个生命周期内对数据进行管理,要在数据产生之前,就实现有效的数据规划,在生命周期的各个阶段主动使数据增值,最终目标在于使这些数据可以方便地被他人应用或重用。

#### 3.3 参与者与服务对象不同

Digital Preservation 活动的参与者一般可分为数据提供者、保存管理者和授权使用者(数据消费者)。数据提供者提供数据,数据保存者对数据实施长期保存,而授权使用者可以访问保存仓储中自己有权访问的数据,原则上讲这些参与者的角色不能互换。Digital Preservation 活动的服务对象是未来的使用者,让保存

在仓储中的数据能够被未来的使用者理解和应用。

Digital Curation 活动的参与者围绕着数据的生命周期可分为数据创造者、数据提供者、数据存档者和数据消费者。这些参与者的角色可以相互转换, 如一个数据创造者同样可能是已增值数据的消费者。Digital Curation 活动的服务对象同时包括当前和未来的数据使用者。Digital Curation 在遵守数据权益的条件下, 鼓励所保存数据更多更广泛地被共享和利用。

### 3.4 遵循的保存行为模型不同

Digital Preservation 活动的行为模型可以用 OAIS 参考模型来概括。模型定义了 6 项功能活动、3 类信息包、3 种角色。按照这一模型, 保存管理者从数据提供者那里获得以 SIP 包(提交信息包)封装的需要保存的数据信息, 实施 6 项保存活动: 保存规划、数据摄入(通过摄入数据, 将 SIP 包转化为存档信息包即 AIP 包)、数据管理、数据存档、经营管理和访问控制, 对被保存的数据进行管理, 并通过访问控制, 向授权访问的用户(消费者)提供数据的 DIP 包(数据分发包)。

Digital Curation 活动的行为模型可以参考 DCC 的 Curation Lifecycle Model 模型<sup>[35]</sup>来概括。这一模型根据数据的生命周期, 提出 4 个贯穿于全生命周期的活动、8 个在生命周期内顺序进行的活动和 3 个偶尔执行的活动。4 个贯穿全生命周期的活动分别是: 对数据及其呈现信息进行描述揭示、进行数据保存规划、吸收团队监管和参与、实施保存和管理操作。在每一个生命周期内, 需要顺序进行 8 个活动, 分别是: 进行构想、产生或接收数据、对数据进行评价和选择、数据摄入、开展保存、进行存档、数据访问(数据应用或重用)、数据转换。而在 Digital Curation 活动中, 可根据实际情况, 可能在数据的生命周期中实施 3 个偶尔执行的活动, 分别是: 丢弃数据、重新评价数据和数据迁移。

### 3.5 保存的周期和对数据内容的管理方法不同

Digital Preservation 活动从接收数据提供者提供的数据开始。它需要对提供的数据进行细致的规范处理, 提供详尽的数据描述、技术元数据、来源信息等记录, 之后实现保存管理, 一直到将来的某个时段保证数据的生存。保存服务机构不关注数据提交之前的状态。在数据的保存管理上, Digital Preservation 重点

在于保存数据的“原生态”。虽然为了防止技术退化, 需要根据技术发展的需要, 对数据的格式、组织结构、技术环境进行迁移, 但原则上不涉及对数据知识内容的增删处理, 以保证数据内容的真实可信。

Digital Curation 活动从数据产生之前的数据规划开始。Digital Curation 强调数据保管者(Curator)要参与到科研活动中去, 在数据的整个生命周期之内, 对数据进行规划、收集、保存、管理和提供应用, 对于一些没有价值的信息, 可以丢弃。在数据的保存管理上, Digital Curation 活动强调不仅仅让科研数据可以存活可用, 保证这些数据的访问处理技术不退化, 更重要的是对数据的知识内容进行保存和管理: 通过对这些数据内容的处理, 得到新的知识内容; 可以根据情况对数据的知识内容进行增删处理(当然一般需要记录规范的保存元数据并保存好数据的原始版本), 以更好地满足特定应用的需要。

### 3.6 保存的技术方法不同

Digital Preservation 的技术方法主要有翻新、迁移、仿真、内容保存、技术保存、服务保存等。这些方法又可划分为更多细小的类别。例如迁移可以更细分为硬件迁移、软件迁移、载体迁移、格式迁移、版本迁移和访问点迁移等。但无论方法如何多样, 这些数字保存的技术方法都有一个共同的特点, 即保证被保存数据(包括比特层、数字内容层和知识层)能够存活下去, 在未来有有效的技术、工具和方法能够像先前应用数据一样正确地访问、呈现和应用这些数据。

Digital Curation 采用了更为广泛的技术方法来对数据进行保存和管理。对数据进行人工或自动化处理的很多方法都是 Digital Curation 所采用的技术方法, 如内容标注、数据关联匹配、数据建模、数据检验、数值分析、新数据生成、数据发布等都是 Digital Curation 的技术方法。这些技术方法的重要特点是围绕着数据的增值和数据的应用而展开的, 目标在于让数据能更方便地为用户所应用。

### 3.7 保存中的主观能动性不同

Digital Preservation 通常被认为是一种消极被动的保存活动。在 Digital Preservation 活动中, 数字保存服务机构被动地接收数据提供者提供的数据, 数据被封装在特定的信息包中被动地被保存。除了计划中的稳定性、完整性、真实性检查测试, 以及必要的翻新、

迁移等工作之外, Digital Preservation 要求尽量减少对保存数据的存取, 同时对数据的访问进行严格的授权控制, 以减少数据被人为删除、修改或损坏的可能性。

相比之下, Digital Curation 的活动被认为是一种积极主动的保存活动。Digital Curation 围绕着科研过程主动进行数据的保存管理。它要求主动地去帮助数据生产者规划数据的产生和收集, 在数据产生之后积极主动地实现数据的增值。在整个数据生命周期之内, 不断对数据进行管理、评价、转换和应用, 使被保存数据能够不断地被应用和重用, 积极争取在数据的保存和管理过程中, 发挥数据的价值。

### 3.8 保存仓储的要求不同

虽然 Digital Preservation 和 Digital Curation 都要求保存仓储是可信仓储。但两者在具体细节上还是有明显的区别。

Digital Preservation 的保存仓储通常被认为是一种不可访问的“Dark Archive”。数据提供者将数据提交给保存服务机构之后, 保存工作在外人看不清里面工作情况的“黑箱”中进行。虽然每个数字保存仓储都在努力通过认证, 证明自己是可信的。但是这些可信的保存仓储给数据提供者的参与权限很少。数据提供者只能了解自己提供数据的保存情况, 进行有限的存档内容审计, 而无权直接对仓储内自己的数据进行更多操作。

Digital Curation 的保存仓储是一种更为开放的可信仓储。这类仓储的目的是为了促进数据的重用和共享, 仓储提供标准规范让数据生产者或数据提供者可以规范地去准备数据、选择数据处理过程、确定数据保存密级、实现元数据标注、进行数据发布、接受用户对数据的预订、进行数据传递、实现数据关联、产生新数据等。在 Digital Curation 的开放可信仓储中, 开放的数据发布、数据检索和数据传递是一项很重要的功能要求。

## 4 结 论

从形成历史、定义辨析和多角度的对比分析, 可以明显地看到 Digital Curation 和 Digital Preservation 虽然都可用于数字保存和管理, 但它们的确是两个不同的概念, 在很多方面有着较大的差别。在具体的应用环境中, 应当根据应用场合, 使用不同的概念。

也应当看到, 两个概念是互补的。虽然笔者不同意 Digital Curation 的目的是为了 Preservation (“Digital Curation for Preservation”), 但笔者认为将 Digital Curation 和 Digital Preservation 二者结合, 才是数字保存管理的全部。

本文除了从理论上剖析这两个概念之外, 另一重要的目的在于希望界内人士在明晰这两个概念之后能够对目前和未来的工作提出思考。ARL 在 2011 年发布的“新时代新角色”的报告, 讨论了研究型图书馆的新趋势, 提出 Digital Curation 和 Digital Preservation 是研究型图书馆在新时代有效服务于其用户的关键。ARL 希望研究型图书馆能够围绕着科学研究实施 Digital Curation 和 Digital Preservation, 主动进行科研数据的保存管理, 使其不断增值以产生新的知识。正如报告中所说的那样, “图书馆不能再期望研究人员和学者到馆咨询和请求援助, 而是应该发掘新的方式, 主动为他们提供没有地域限制的服务”。要保持活力, 图书馆必须要调整其服务提供方式, 以一种更为积极主动的态度来适应新的环境和新的需求。无论 Digital Curation 还是 Digital Preservation, 都应该成为当前图书馆重点投入的发展方向。

### 参考文献:

- [1] Conway P. Archival Preservation Practice in a Nationwide Context [J]. American Archivist, 1990, 53 (2): 204-222.
- [2] Conway P. Digitizing Preservation [J]. Library Journal, 1994, 119 (2): 42-45.
- [3] Conway P. Preservation in the Digital World [EB/OL]. [2013-11-25]. <http://www.clir.org/pubs/reports/conway2/reports/conway2/index.html>.
- [4] Hedstrom M. Digital Preservation: A Time Bomb for Digital Libraries [J]. Computers and the Humanities, 1998, 31: 189-202.
- [5] Garrett J R. Task Force on Archiving of Digital Information [J/OL]. D-Lib Magazine, 1995. [2013-11-25]. <http://www.dlib.org/dlib/september95/09garrett.html>.
- [6] Andre P, Besser H, Elkington N, et al. Preserving Digital Information: Report of the Task Force on Archiving of Digital Information [EB/OL]. [2013-11-24]. <http://www.clir.org/pubs/reports/pub63/watersgarrett.pdf>.
- [7] Higgins S. Digital Curation: The Emergence of a New

- Discipline[J]. *International Journal of Digital Curation*, 2011, 6 ( 2 ): 78-88.
- [8] 张智雄, 郭家义, 吴振新, 等. 基于 OAIS 的主要数字保存系统研究[J]. *现代图书情报技术*, 2005 ( 11 ): 1-13. ( Zhang Zhixiong, Guo Jiayi, Wu Zhenxin, et al. An Analysis of Major OAIS-based Digital Preservation Systems[J]. *New Technology of Library and Information Service*, 2005 ( 11 ): 1-13. )
- [9] CEDARS. *Curl Exemplars in Digital ARchiveS* [EB/OL]. [2013-11-23]. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part4/fox>.
- [10] The NEDLIB Project [EB/OL]. [2013-11-23]. <http://www.dlib.org/dlib/september99/vanderwerf/09vanderwerf.html>.
- [11] Oltmans E, van Wijngaarden H. The KB e-Depot Digital Archiving Policy [J]. *Library Hi Tech*, 2006, 24 ( 4 ): 604-613.
- [12] A Collaborative Initiative of the Library of Congress. *Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program*[EB/OL]. [2013-11-25]. [http://www.digitalpreservation.gov/documents/ndiipp\\_plan.pdf](http://www.digitalpreservation.gov/documents/ndiipp_plan.pdf).
- [13] Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System ( OAIS ) [EB/OL]. [2013-11-23]. <http://public.ccsds.org/publications/archive/650x0b1s.pdf>.
- [14] Beagrie N, Lupovici C, Doerr M. *Trusted Digital Repositories: Attributes and Responsibilities* [EB/OL]. [2013-11-22]. <http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>.
- [15] Beagrie N, Pothan P. *The Digital Curation: Digital Archives, Libraries and e-Science Seminar* [J/OL]. *Ariadne*, 2002 ( 30 ): 98-102. [2013-11-20]. <http://www.ariadne.ac.uk/issue30/digital-curation/>.
- [16] Beagrie N. *Digital Curation for Science, Digital Libraries, and Individuals* [J]. *International Journal of Digital Curation*, 2008, 1 ( 1 ): 3-16.
- [17] Macdonald A, Lord P. *Digital Data Curation Task Force: Report of the Task Force Strategy Discussion Day*[EB/OL]. [2013-11-24]. [http://www.jisc.ac.uk/uploaded\\_documents/CurationTaskForceFinal1.pdf](http://www.jisc.ac.uk/uploaded_documents/CurationTaskForceFinal1.pdf).
- [18] Lord P, Macdonald A. *e-Science Curation Report, Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision*[EB/OL]. [2013-11-26]. [http://www.jisc.ac.uk/uploaded\\_documents/e-ScienceReportFinal.pdf](http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf).
- [19] Digital Curation Centre[EB/OL]. [2013-11-23]. <http://www.dcc.ac.uk/>.
- [20] *International Digital Curation Conference*[EB/OL]. [2013-11-23]. <http://www.dcc.ac.uk/events/internationaldigitalcuration-conference-idcc>.
- [21] *International Journal of Digital Curation*[EB/OL]. [2013-11-23]. <http://www.ijdc.net/>.
- [22] *Digital Curation Unit*[EB/OL]. [2013-11-24]. <http://www.dcu.gr/>.
- [23] UC3 : *University of California Curation Center*[EB/OL]. [2013-11-24]. <http://www.cdlib.org/services/uc3/>.
- [24] *Digital Research and Curation Center*[EB/OL]. [2013-11-24]. <http://ldp.library.jhu.edu/dkc>.
- [25] *The Digital Curation Institute*[EB/OL]. [2013-11-24]. <http://dci.ischool.utoronto.ca/>.
- [26] Giaretta D. *DCC Approach to Digital Curation* [EB/OL]. ( 2004-08-15 ) . [2013-11-26]. <http://twiki.dcc.rl.ac.uk/bin/view/OLD/DCCApproachToCuration>.
- [27] *Digital Curation Centre ( DCC ) .DCC Charter and Statement of Principles* [EB/OL]. [2013-11-26]. <http://www.dcc.ac.uk/about-us/dcc-charter/dcc-charter-And-statement-principles>.
- [28] *Digital Curation Centre ( DCC ) .What is Digital Curation?* [EB/OL]. [2013-11-26]. <http://www.dcc.ac.uk/digital-curation/what-digital-curation>.
- [29] Abbott D. *What is Digital Curation?* [EB/OL]. [2013-11-26]. <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation>.
- [30] JISC. *JISC Circular 6/03 ( Revised ): An Invitation for Expressions of Interest to Establish a New Digital Curation Centre for Research into and Support of the Curation and Preservation of Digital Data and Publications* [EB/OL]. [2013-11-22]. [http://www.jisc.ac.uk/fundingopportunities/funding\\_calls/2003/09/digcentre\\_townmeeting.aspx](http://www.jisc.ac.uk/fundingopportunities/funding_calls/2003/09/digcentre_townmeeting.aspx).
- [31] Pennock M. *Digital Preservation : Continued Access to Authentic Digital Assets* [OL]. [2013-11-26]. [http://sitecore.jisc.ac.uk/publications/briefingpapers/2006/pub\\_digpreservationbp.aspx](http://sitecore.jisc.ac.uk/publications/briefingpapers/2006/pub_digpreservationbp.aspx).
- [32] Walters T, Skinner K. *New Roles for New Times: Digital Curation for Preservation* [M]. Washington, DC: Association of Research Libraries, 2011: 78-86.
- [33] Beagrie N, Jones M. *Digital Preservation Handbook* [M]. New York: Digital Preservation Coalition, 2008: 8-9.
- [34] Caplan P. *FCLA Digital Archive ( 2005 Joint Meeting of the CSUL Committees PPT )*[OL]. [2013-11-23]. [https://fclaweb.fcla.edu/uploads/Lydia%20Motyka/FDA\\_documentation/Publ](https://fclaweb.fcla.edu/uploads/Lydia%20Motyka/FDA_documentation/Publ)

ications\_Presentations/FDAJointMeeting2005.ppt.

[35] DCC Curation Lifecycle Model [EB/OL]. [2013-11-23].  
<http://www.dcc.ac.uk/resources/curation-lifecycle-model>.

### 作者贡献声明:

张智雄: 负责论文设计构思、资料收集和分析、整篇论文的撰

写、最后审阅及定稿;

吴振新: 补充部分资料、论文内容完善, 特别是结论部分的内  
容完善;

刘建华, 郭红梅: 补充部分资料、论文内容完善。

(通讯作者: 张智雄 E-mail:zhangzhx@mail.las.ac.cn)

## Analysis of the Difference Between Digital Curation and Digital Preservation

Zhang Zhixiong Wu Zhenxin Liu Jianhua Guo Hongmei  
(National Science Library, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** [Objective] To analyze the difference between Digital Curation and Digital Preservation. [Coverage] Based on the important historical documents and reports related to those two concepts, the authors also investigate the definitions of those two concepts released from major institutes in this research area, such as DCC, JISC and ARL. [Methods] Based on the analysis of history of those two concepts, the authors figure out the underlying causes of those two concepts. Based on the definitions of those two concepts from major research institutes and researchers, the authors analyze the difference between those two concepts. Based on the analysis of those two concepts from 8 aspects, the authors perform a multi-aspect comparison study of the two concepts to identify the difference more clearly. [Results] Although Digital Curation and Digital Preservation have similar meaning, they are quite different two concepts that have lots differences in many aspects. [Limitations] More detailed comparison on the tasks of the two concepts is needed to give more clearly explanation. [Conclusions] Digital Curation and Digital Preservation are two different concepts, but they are complementary. A more active way is needed for the library to carry out digital preservation.

**Keywords:** Digital Curation Digital Preservation Concept analysis