

# 数字资源长期保存认证体系中用户相关指标分析

## Analysis of Criteria relating to Users in Digital Preservation

### Certification

张 玫<sup>1, 2</sup>

1 (中国科学院国家科学图书馆 100080)

2 (中国科学院研究生院 北京 100049)

**摘要:** 在 OAIS 参考模型的基础上, 对三个具有代表性的长期保存系统认证体系中涉及用户的指标加以比较, 分析它们在用户定义、对用户需求与知识背景变化的应对、数字对象的发现、传递及使用等方面的异同; 最后指出认证体系还需要进一步加强政策的适应性、对变化的预防及用户感知的服务效果等方面的指标。

**关键词:** 数字资源长期保存 认证 用户 目标群体

**Abstract:** Based on the OAIS Reference Model, the article compares and analyzes the user related criteria of three digital preservation certification in terms of user define, feedback to the changes of users' needs and knowledge base, and the discovery, transfer and usage of digital objects, points out the criteria concerned with policy adaptation, the risk prevention and the effectiveness of service need to be strengthened.

**Keywords:** digital preservation, certification, users, designated community

### 1. 数字资源长期保存认证体系概况

随着计算机技术的进步与网络的普及, 越来越多的资源以数字化形式出现。鉴于数字资源的自身特点, 如对应用环境的依赖性及其不稳定性, 加之一些经济、法律、政治及自然因素的影响, 数字资源持续可用性受到威胁, 而其长期保存也成为关注的热点。目前, 人们已经认识到数字资源长期保存的必要性, 并开展了大量相关实践。对长期保存活动而言, 最根本的宗旨就是保存人类的科学文化遗产, 并为公众提供持续服务, 因此, 如何保证长期保存系统的可信度便成为人们关注焦点之一。

早在 2000 年, 研究图书馆协会 (Research Libraries Group, RLG) 与联机计算机图书馆中心 (Online Computer Library Center, OCLC) 就在《可信赖数字仓储: 属性与职责》的报告中对可信赖长期保存系统的定义和特点做了详细阐述, 它指出可信赖长期保存系统 (Trusted Digital Repositories, TDR) 的目的在于为目标群体提供经过管理的数字资源的长期可靠的存取, 包括现在的存取和未来的存取<sup>[1]</sup>。

在此基础上, 一些机构或组织开展了对长期保存系统与服务可信度的认证工作, 并制定出具有普遍性或仅适合本机构具体环境的认证指标体系。其中比较有代表性的指标体系包括: RLG 与美国国家档案管理局 (National Archives and Records Administration, NARA) 根据不同国家对长期保存系统加以审计或认证的实践, 从中选取了一系列通用的核心标准与最佳实践来构建出一套具有普遍性的认证指标体系:《可信赖仓储的审计及认证: 指标与列表》(Trustworthy Repositories Audit & Certification: Criteria and Checklist, TRAC)<sup>[2]</sup>; 德国网络信息计划 (Deutsche Initiative für Netzwerkinformation, DINI) 制定的《DINI 文档与出版物服务认证》主要针对德国机构仓储的开放存取, 旨在为各大学机构仓储间的标准化及基于互操作的信息服务与交流提供建议<sup>[3]</sup>; 由德国联邦教育科研部 (Federal Ministry of Education and Research, BMBF) 资助的“德国数字资源的长期存储与长期存取专业网络” (Network of Expertise for Long-term

Storage and Long-Term Availability of Digital Resources in Germany, nestor) 下属的可信赖仓储认证工作组制定的《可信赖数字仓储的指标体系》则是一套基于德国国情的认证体系, 它比 DINI 认证更为全面, 其基本框架与 TRAC 保持一致<sup>[4]</sup>; 此外, 英国数字长期保存中心 (Digital Curation Center, DCC) 针对 TRAC 与 nestor 的认证指标, 开发出一套“基于风险管理的数字仓储审计方法” (Digital Repository Audit Method Based On Risk Assessment, DRAMBORA), 其目的是帮助各长期保存系统收集相关证据, 并以此来判断自身是否符合上述两种认证的要求<sup>[5]</sup>。

根据《可信赖数字仓储: 属性与职责》所述, 长期保存系统的可信度可从以下三方面衡量: (1) 保存机构如何获得目标群体的信任; (2) 保存机构如何信任第三方提供者; (3) 用户如何信任由保存机构提供的资源<sup>[6]</sup>, 其中第 1、3 两点直接与用户关联。多数长期保存系统的认证体系将其认证对象主要划分为机构的组织框架、数字对象管理以及系统框架与安全等三部分, 尽管没有把用户因素作为一个独立部分进行评估, 但在其具体的各项指标中也设立了许多与用户相关的指标, 鉴于用户因素对评估长期保存系统可信度具有不可忽视的影响, 因此, 本文拟从用户角度入手, 对长期保存系统认证系统中的相关指标进行分析, 一方面有利于深入了解用户对长期保存系统的影响, 另一方面, 对于我国今后建立的认证体系也具备借鉴意义。

## 2. 用户对数字资源长期保存系统的影响——以 OAIS 参考模型为基础

### 2.1 长期保存系统对“用户”的界定

在上述四种认证体系中, 除 DINI 认证外, 其他三种都声称其指标支持开放档案信息系统 (Open Archival Information System, OAIS) 参考模型, 该参考模型根据不同的功能, 将与系统相关的人或组织划分为生产者、管理者及消费者等三种角色, 其中, 消费者是指与 OAIS 服务交互以获取特定保存信息的人或系统, 它也包括“目标群体”, 即事先定义好的一组潜在的消费群体, 他们具备对特定信息的理解能力, 并且可能由不同的用户群体组成<sup>[7]</sup>, 而本文对“用户”的定义也将沿用 OAIS 参考模型中“消费者”的概念。

### 2.2 长期保存系统涉及用户的功能

OAIS 参考模型框架定义了 6 个功能模块, 即: 摄入、档案存储、数据管理、管理、保存规划及访问<sup>[8]</sup>。其中, 与用户相关的模块及其具体功能主要有以下两部分:

(1) 保存规划模块: 负责监测 OAIS 的环境, 并提供相关建议, 以确保 OAIS 存储的信息在经过长时间之后, 即使原有环境发生变化, 仍然能够被相应的目标群体访问, 其涉及用户的具体功能主要体现在监测目标群体的服务需求及其知识背景的变化上;

(2) 访问模块: 用于支持用户, 使其能判断、了解、定位并存取 OAIS 中存储的信息, 确保他们具备请求并接受信息的能力。其中涉及用户的具体功能主要包括接收来自用户的请求, 判断用户的访问权限并以此作为对其请求响应程度的依据, 根据用户需求生成响应并且将这些响应最终传送给用户。

## 3. 数字资源长期保存认证体系中与用户相关的指标分析

鉴于 DINI 认证主要针对机构仓储的开放存取, 其指标设立的出发点与其他三个认证体系存在较大差异, 因此本文将在保存规划与访问两大模块及其具体功能的基础上, 对除 DINI 认证之外的其他三个认证体系中涉及用户的指标进行对比分析。

### 3.1 保存规划中涉及用户的指标分析

#### 3.1.1 对用户的定义

在进行对比的三个认证体系中, 都具备与用户定义相关的指标。其共同点主要包括:

(1) 定义用户: 即确定潜在用户的身份, 在此基础上还应定义用户的知识背景;

(2) 定义用户需求: 用户需求不仅包括用户所需的服务内容、服务水平及信息访问途径,

还包括用户对信息的理解程度，这都会对长期保存系统中的数字对象管理与技术框架造产生影响，因此，一个可信赖的长期保存系统必须在细分用户群体的基础上，分析其各自的信息呈现方式。

(3) 定义的开放性：即在定义好用户身份及其需求后，还应公开此类文档，以便公众，特别是利益相关群体能准确了解系统的目的，同时还能起到监督作用。

除上述共同点外，三个认证体系也存在一些不同之处：

(1) 用户参与：**DRAMBORA** 提出需要就长期保存系统所理解的用户需求与用户进行交流，并征求其反馈意见，以保证系统能正确定义用户需求；

(2) 政策保证：**TRAC** 与 **DRAMBORA** 在其指标中都提出制定相关政策来说明长期保存系统提供的服务是如何实现目标群体的需求，以便从制度上保证用户需求能被准确执行。

### 3.1.2 对用户需求及知识背景变化的应对

鉴于用户需求可能随许多因素发生变化，长期保存系统的用户定义也应随之发生变化。针对这种情况，三个认证体系的指标基本相同，它们都主要围绕以下两方面展开：

(1) 建立监控机制：长期保存系统可采用组织的、技术的手段定期检查用户需求、期望及知识背景等是否发生变化。此外，**TRAC** 还特别强调需要监管硬软件技术变化的通知，因为这些变化都可能对用户服务产生影响；

(2) 建立反应机制：当用户需求发生变化时，系统需分析这些变化对服务的影响，同时还要建立一种灵活的反应机制，以便及时修正由上述变化带来的一系列系统内部的变动。

## 3.2 访问中涉及用户的指标分析

### 3.2.1 数字对象的发现

长期保存系统提供的服务建立在用户对数字对象发现的基础上，而三个认证体系对此都制定了相同的指标来确保用户能顺利发现或确定其感兴趣的内容，即定义适当的元数据体系：长期保存系统首先应根据目标群体的具体需要，确定其特定的资源发现与检索程度，在此基础上结合数字对象类型来定义元数据的范围、结构与内容等。相对于其他两个认证体系只是笼统地设定了判断长期保存系统是否设立了适当的元数据体系的指标，**nestor** 则进一步划分出关于描述元数据、技术元数据、管理元数据、起源元数据等更为详细的指标。

为数字对象分配适当的元数据体系从技术上保证了用户对数字对象的发现，此外，**nestor** 还要求系统能保证元数据的持续可用性，即采用外部可见的永久标识符来描述数字对象及其组成部分，再把数字对象与元数据封装起来；同时还要能为用户提供清晰的界面，以帮助他们直观地了解或接受本系统的服务。

### 3.2.2 数字对象的传递

数字对象的传递主要是指系统在收到用户请求后，根据具体的访问政策，判断用户具体的访问权限，从保存信息中提取出用户需要的部分，最后以一种用户可以理解的方式提供给用户。在这一过程中，三种认证体系都注意到了以下两点：

(1) 访问控制：长期保存系统可信度包括如何获得资源提供者的信任，这就要求系统按照资源提供者的要求提供访问控制，因此，首先需要定义访问的各种条件并加以记录，同时还应确保访问管理系统符合访问政策，从而保证只有授权用户才能享用对应的服务。在此基础上，**TRAC** 还提出系统应记录所有访问管理失败的情况，并鉴别其中是否存在不适当的“拒绝访问”事件；

(2) 响应用户请求：系统应能保证用户的所有请求，无论其是否符合访问政策，都能收到一个接受或拒绝的回应，此外，系统还应该采用适当的用户反馈机制，如电子邮件、网页表单、电话等，以使用户的意见能及时得到处理。

除上述共同点外，这三种认证体系还存在着一些差异：

(1) 定义发布信息包 (Dissemination Information Packages, DIPs): 作为系统传递给用户的信息单元, nestor 规定 DIPs 应根据存档信息包 (Archival Information Packages, AIPs) 与目标群体的使用环境来加以定义, 当这些影响因素发生变化时, DIPs 则能重新生成新的版本;

(2) 对传输途径的保障: 数字对象从保存系统传递到用户的过程中需要可靠的保障, DRAMBORA 提出需要具备健壮稳定的系统与技术来保证服务的传递, 而 TRAC 则设立了具体的指标来判断长期保存系统是否制定了相关政策来确保它对所有满足需要的访问操作加以记录。

### 3.2.3 数字对象的使用

为确保数字对象能为用户所用, 系统必须保证其完整性与真实性, 而 TRAC 与 nestor 对此都制定了相应的指标来判断系统的可信度:

(1) 数字对象的完整性: 数字对象的完整性可以从其各方面的属性来衡量。对用户来说, 若他请求的是某个数据集, 那么系统就应提供整个数据集。为达到该目的, 长期保存系统可以首先定义数字对象在传递过程中所能达到的完整性程度, 然后为用户提供判断数字对象是否完整的机制, 最后还应该能对数字对象呈现信息的质量加以分析, 并把分析结果返回给用户;

(2) 数字对象的真实性: 真实性是指数字对象应包含那些它本该包含的信息。数字对象在整个操作过程中可能会出现一些变动, 而长期保存系统应加以记录, 并以适当的方式 (如起源元数据) 向用户展示其变化的过程, 从而帮助他们判断该对象的真实程度。TRAC 特别指出, 当用户获取的数字对象不是原始拷贝时, 系统应确保该数字对象能通过某种方式回溯到其原始形式。

此外, TRAC 还提出了衡量数字对象正确性的指标, 主要用于针对数字对象在传递或使用过程中存在格式转换的情况, 具体可通过系统向用户展示转换的方法及其依据的标准等信息来判断。

## 4. 小结

通过分析上述三种长期保存系统认证体系中涉及用户的指标的异同, 我们可以发现它们目前仍然存在一些缺陷, 主要包括:

(1) 保障政策的适应性: 在三种认证体系中, 分别设立了一些关于制定政策来保证相关操作正确执行的指标, 其中 TRAC 更是要求长期保存系统从多个方面来考虑制定政策, 这一点尤其值得肯定。然而, 在相应的指标中, 都只是简略地说明政策的基本内容, 却忽略了这些政策需要与系统的外部大环境相适应。因此, 可以考虑设立相应的指标以避免出现本地政策、规范等与外部环境相悖的情况。

(2) 加强对变化的预防: 在 TRAC 与 nestor 中, 都有指标来衡量系统在面对用户群体或其需求发生变化时能否有效地收集到反映此类变化的信息, 之后又能否对此做出快速反应, 这将有利于系统及时纠正、修改已经存在的问题, 从而保证其服务的适应性。但是, 该机制仅适用于事后处理, 却忽视了预防这种局面出现的手段。DRAMBORA 由于采用了基于风险分析的方法来评估长期保存系统, 因此它考虑到了风险的预防, 这点将值得其他两个认证体系借鉴。

(3) 注重用户感知的服务效果: 目前, 认证体系主要从系统角度来判断服务的好坏, 而服务质量的本质却应该以用户实际感受到的服务效果与预期服务的差距来衡量, 体现用户感知的服务效果的指标在上述认证体系中都鲜有出现, 这从一定程度上阻碍了系统及其服务应以用户为中心的基本思想的贯彻, 因此, 认证体系在这方面都还有待进一步加强。

---

[1, 6] Research Libraries Group. RLG/OCLC Trusted Digital Repositories: Attributes and Responsibilities [EB/OL]. (2002-05) [2007-12-10].

---

<http://www.rlg.org/legacy/longterm/repositories.pdf>.

[2] The Center for Research Libraries, Online Computer Library Center, Inc. Trustworthy Repositories Audit & Certification: Criteria and Checklist [EB/OL]. (2007-02) [2007-12-10]. <http://www.crl.edu/PDF/trac.pdf>.

[3] DINI Working Group "Electronic Publishing". DINI—Certificate Document and Publication Services 2007 [EB/OL]. (2003-10) [2007-12-10]. <http://nbn-resolving.de/urn:nbn:de:kobv:11-10075687>.

[4] nector Working Group Trusted Repositories – Certification. Catalogue of Criteria for Trusted Digital Repositories [EB/OL]. (2006-12) [2007-12-10]. <http://www.ils.unc.edu/tibbo/JCDL2006/Dobratz-JCDLWorkshop2006.pdf>.

[5] Digital Curation Center, DigitalPreservationEurope. Digital Repositories Audit Method Based on Risk Assessment [EB/OL]. (2007-2-18) [2007-12-10]. <http://www.repositoryaudit.eu/download/>.

[7, 8] The Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System [EB/OL]. (2002-01) [2007-12-10]. <http://public.ccsds.org/publications/archive/650x0b1.pdf>.