

2013 年第 12 期（总第 33 期）

长期保存跟踪扫描

主办单位：中国科学院国家科学图书馆

2013 年 12 月

**为传播科学知识，促进业界交流，
特编译《长期保存跟踪扫描》，仅供个人学习、研究使用。**

目 录

【动态追踪】	1
蒙大拿州立图书馆长期保存介绍.....	1
考古数据服务的价值和影响.....	2
【重要文献摘译】	2
4C 报告:《经济可持续性参考模型草案》摘译.....	2
【资料推荐】	12
DCC:《国际数字保存期刊第 8 卷第 2 期》	12
《提供研究数据管理服务》	13
【专题报道】	13
最佳实践交流大会 (BPE)	13
第十届数字对象长期保存国际会议综述- iPRES 2013.....	14

【动态追踪】**蒙大拿州立图书馆长期保存介绍**

蒙大拿州立图书馆 (MSL) 在数字保存和监管中面临的两大挑战是发展库存系统, 评价和记录数据交换中心 25 年的数据。MSL 正在开发地理信息系统库存系统 GIS, 该系统不仅记录存档数据, 还同时记录数据的分发信息包 (压缩文件、网络地图服务和地图应用程序等) 以及它们之间的关系。对于数据记录而言, MSL 正在努力解决如何容纳 13 个使用实例 (数据形式和情况) 的问题, 包括容纳数据之间的父子关系。由于 MSL 需同时评估和记录存档前的数据交换中心的数据, 所以急于运行一个可持续发展的系统以及相应的数据发现工具。

MSL 已经有处理频繁变化的数据集 (框架数据) 的归档程序。然而, 数据交换中心现有的大量数据提出了一个更大的挑战。目前 MSL 正在组织数据交换中心已提供服务的数据, 以及储存在外部驱动器、职工的硬盘甚至 CD 中的数据。许多数据是副本, 而且很多原始数据集没有元数据。MSL 需要审查数据并记录数据和副本, 以确定哪些数据需要存档哪些数据需要遗弃。GIS 组织在开始保存程序时, 要做的最重要的事是组织并记录已有数据。

MSL 正在数字化处理其拥有的州出版物。这项工作的强大推动力是为公众提供更便利的公共访问。网络统计数据显示, 一旦数字化, 用户访问文档的次数可大幅度上升。

MSL 提供数字化的地理数据的在线访问已经有很长的历史。目前 MSL 正努力获得对这些数据的物理控制和智能控制, 以揭示长期未被发现和被替代的数据, 我们迫切希望将这些数据进行管理并提供永久的公共访问。

MSL 属于一个专家网络, 这些专家能够理解和评估 GIS 数据价值, 可以依靠专家来支持 GIS 数据保存工作。也就是说, 这些获得支持的州立机构和地方政府可以从不同定位思考完成自己的数据保存工作。如果没有财政支持和人力资源, 或者是没有机构层面的实施数据保存的政策和管理的支持, 数据保存工作会比较艰难。这也很可能是他们的业务需求都集中在当前问题的缘故。

编译自:

<http://blogs.loc.gov/digitalpreservation/2013/12/content-matters-interview-the-montana-state-library-part-two/>

(唐果媛编译, 王敬 吴振新校对)

考古数据服务的价值和影响

Charles Beagrie 有限公司的 Neil Beagrie 和战略经济研究中心的 John Houghton 发表来自于 JISC 研究中所支持的关于评测考古数据服务的价值和影响的最终报告。这项研究的目的是探索及尝试衡量考古数据服务的价值和影响。他们运用了大量的经济学方法来分析从网络调查、用户及保存机构统计信息中收集的数据,同时用这些方法来补充和扩展其他非经济价值认知。

这项研究改变了利益相关方的观念,普遍提高了他们对考古数据服务、数字保存及数据分享的认同度。其实绝大多数利益相关方已经意识到考古数据服务具有很高的价值,但是依然认为这项研究拓展了他们对这些价值的理解范畴及这些价值对其他利益相关方的重要程度。他们看好其在经济方面呈现的价值,而在以前这是完全没有考虑到或看到过的。

报告全文:

http://repository.jisc.ac.uk/5509/1/ADSReport_final.pdf

编译自:

<http://blog.beagrie.com/?p=794>

(陈瑶编译,唐果媛 吴振新校对)

【重要文献摘译】

4C 报告:《经济可持续性参考模型草案》

4C 作为欧盟资助的重要保存项目,提供了大量成果共享,本报告为其系列成果之一。

1 说明

数字保存和其它活动一样,都是人们为实现明确的目标而持续性地供应丰富的资源。这种资源分配策略必须依赖于对数字保存长期成本的深入了解。例如:必需的投资,同时还有数字保存的预期收益,如预期的投资回报。只关注成本而不考虑收益(或相反)的可持续发展战略实际上根本就不是可持续发展战略,它们很可能注定要失败。可持续的数字保存活动必须要对预期的成本与效益有一个清晰的认识,如果成本高于预期收益,那么产生的必然结果是数字保存活动就不应该开始,或若已开始,就必须停止。

这并不是说制定一个可行的可持续发展战略是容易的。相反,它往往涉及建立和维护各种保存动机不同或完全相反的利益相关者之间复杂而微妙的关系,它经常需要从多个资源中产生和维护资源流。一个关键问题是在预算紧张时如何确定数字保存的重点资源,保存机构必须对选择或者保存哪些数字资产做出艰难的选择,有时候,为了保存价值高的稀缺资源,

保存机构还必须放弃部分已保存的数字资源。这需要制定应急计划来解决在数字资源长期保存很多方面中的不确定性。一般来说，设计一个可持续发展战略可能被认为要以永久保存所有资源这个不可能的目标为初衷，但考虑到在特定的保存环境下普遍存在的情况和限制，所以应致力于一些在实践中接近最初理想目标的工作。

数字保存的经济问题是一个难以掌控的复杂局面。在这种情况下，参考模型是一个有用的工具，在一个复杂的问题空间中，它可以帮助降低问题的复杂性并且突出关键概念、关系和决策点。OASIS (<https://www.oasis-open.org/committees/soa-rm/faq.php>)，一个促进开放信息标准的组织，定义了一个参考模型：“一个为了理解某些环境中实体之间重要关系以及支持这种环境下统一的标准或规范发展的抽象框架”。

换句话说，参考模型是一个定义及组织组件的模板，这些组件必须是特定问题空间中方案的一部分。参考模型定义了可以促进对组件共同理解的概念和相关词汇以及组件之间的关系。为了实现上述目标，参考模型采用了一定程度上的抽象概念，使得其能抽象于特定环境或是具体的问题空间之上，从而提高了在许多情况下的适用性。

这篇文章中描述的 ESRM 的目的是实现数字保存的经济可持续性参考模型，就像 OAIS 为数字保存技术方面所做的工作那样。该模型提出了一个框架，该框架引导了数字保存经济可持续性的讨论方向。这些讨论可能在多种环境下发生：内部的（一个特定的保存项目），或外部的（涉及不同的利益相关者群体）；重点关注可持续性问题的特定方面（如成本），或者作为一个整体来考虑问题；从头开始设计一个全新的可持续性策略，或评估现有策略的优点，或比较多种策略的特点。在所有这些环境下经济可持续性参考模型将会是一个有用的工具，甚至意味着更多。同样，保存人员在这里可以将 ESRM 的预期使用效果与 OAIS 所成功应用的使用效果同等看待。

2 背景（略）

3 ESRM 的目的（略）

4 框架草案

数字保存活动采取技术策略来把保持数字资源长期可访问所需的技术流程和工作流协调地组合起来：例如，冗余存储、常规媒体更新、格式迁移计划等等。同样地，数字保存活动必须制定一项可持续发展策略来协调经济因素，这些经济因素是为保障保存活动有充足的资源满足其长期保存目标所必备的。

OAIS 参考模型采用“开放档案信息系统”作为一个组织性的概念去呈现数字保存仓储库主要技术组件的一般性描述。ESRM 围绕可持续性战略的概念组织了以数字保存经济可持

续性为主题的讨论。同样，数字仓储库的规划者构建开放档案信息系统的目的是解决数字保存的技术基础，他们同时也必须设计一个可持续性的战略去建立与数字保存技术基础平行的经济基础。可持续性战略由几个关键部分构成：

- 经济生命周期：保存活动产生时的经济“上下文环境”。
- 可持续性条件：条件必须能满足和实现可持续性。
- 主要实体：保存环境，包括数字资产；保存流程；以及利益相关者必须做出决定以保证可持续性的条件得到满足。
- 经济不确定性：在实现可持续性发展时形成的不确定性为其提出的挑战。

上述组件在下面章节中进行详细解释。

ESRM 为数字保存活动设计可持续性战略时描绘出了规划者所面临的问题空间的关键因素。参考模型专注于可持续性战略的一般性概念，把参考模型分解为关键部件，同时让规划者重视与经济可持续性最相关组件的属性。因为参考模型虽然不能“解决”全部数字保存活动的可持续性问题，但它提供了一个可以理解这些问题本质的框架和已设计出的合适的解决方案。

4.1 参考模型概述（略）

4.2 可持续性战略

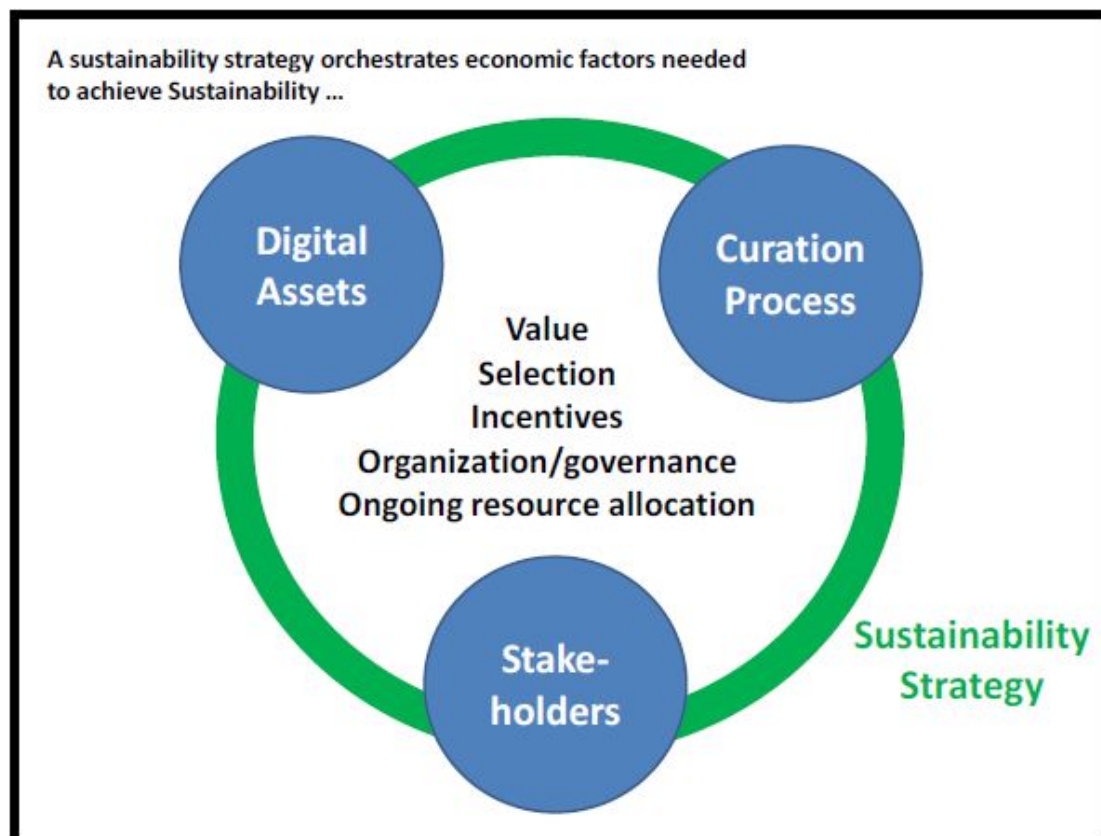


图 2: 可持续性战略组成

要点:

- 可持续性策略协调经济因素以确保保存活动在其整个经济生命周期中有足够的资源来满足它的长期目标。
- 达到可持续性意味着满足 BRTF (管制小组) 制定的五个可持续性条件。
- 设计一个成功的可持续性战略, 规划者必须理解关键实体的属性; 找出与属性相关的重要的经济不确定性; 同时制定出解决不确定性的适当补救措施。
- 没有哪个可持续性策略是完美的; 它仅能够最大化获得可持续性的前景, 但是并不能保证能够实现这一前景。
- 可持续性策略必须随着环境的发展而演变。

4.2.1 经济生命周期

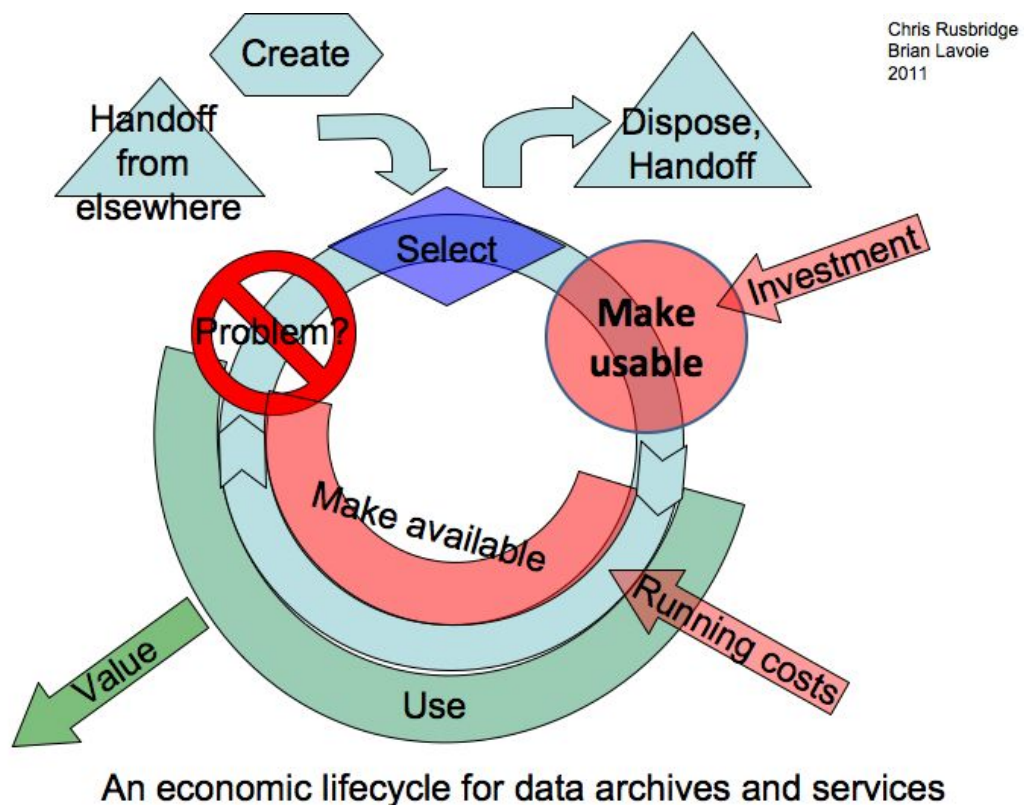


图 3: 数据存档与服务的经济生命周期模型

要点:

- 数字保存的经济决策是一个随着时间而演变的动态连续的过程，并不是一个一蹴而就的结果。
- 保存决策，尤其是有关可持续性方面，必须定期的回顾、重估，如有必要还需修改。
- 经济生命周期模型重点关注主要的经济决策时刻，这些时刻通常出现在数字资产从创建到可能的清除的整个“生命”过程中。
- 数字保存通常的决策模式包括两个端点——摄入数字资产时和从存档资源中删除资产时，这之间有一个周期性的时期，数字资产在这个时期内可被持续性地使用，直到周期发生中断。
- 持续可用性的常规周期的中断可以预料到，也可能预料不到。

4.2.2 可持续性条件

在讨论获得数字保存经济可持续性战略时，重要的是对“获得经济可持续性”真正意味着什么要有一个清晰的认识。BRTF（监管小组）总结报告列举了为最大化获得数字保存经济可持续性的预期收益所必须满足的五个条件。

4.2.2.1 价值

在特定的数字资产的长期保存中，这个条件需要认识到利益相关者的利益。如果没有人在保存的数字资源中看到价值，数字保存活动就不可能吸引到资金（从而支撑数字保存的运转）！但是远不止这些，还包括对保存数字资源感兴趣的利益相关者必须能够强烈地、具体地表达他们的利益或价值需求。哪些类型的有价值的活动能保证数字资源被永久性地获取？或者相反，哪些类型的有价值的活动不会导致数字资源丢失？是什么使得利益相关者愿意去支付资金来保障数字资产不丢失？由于花费的成本是资金但回报却不是资金而难以构建一个持续性的例子，所以可持续性受到了价值观念的极大影响。

4.2.2.2 选择

经济学的基本原则之一是资源具有稀缺性；我们经常不能获得对我们有用的所有资源。对于数字保管来说，这意味着“永久保管所有资源”是不可能的。保管可用的资源总会受到限制，因此我们必须优先考虑并做出选择：在可能的范围内，必须主动选择那些在长期使用中可能带来最大价值的数字资产进行保管。同时必须记住从两个方向进行选择，尤其是当持续保管的资产所产生的价值不能覆盖成本时，我们必须准备“取消选择”已保存的数字资产。

4.2.2.3 动机

利益相关者清晰的表明对持续获取的数字资源所带来的利益感兴趣是一回事（参见“价值”这一节），而进一步承担保存的责任则是另一回事。可持续性数字保存不仅需要利益相

关者认识到保存的价值，而且还需让他们乐意去赞助或者实施保存活动。简而言之，保存数字资源需要强劲的动机和激励。培养这些动机通常意味着识别和利用机构在保存活动中的自身利益：例如，可以把保存作为一个商业机遇、机构任务的一部分或完成任务的手段等等。通常，涉及到保存的利益相关者都有着不同的保存动机；这些动机需要在整个数字生命周期中进行清晰的说明。例如，当数字资产具有经济价值时，媒体公司可能会因为收益的动机而在有限时期内保存数字电影。当这段时期过去之后，激励机制需要以不同的动机把数字资产转移到另一机构，如图书馆或档案馆。

4.2.2.4 资源

经济可持续发展离不开资源。与其他活动一样，保存活动也需要丰富的资源才能实现长期目标。要满足这个条件往往归结为数字保存的发展机制，发展机制让从数字保存中获益并且愿意为此付费的用户提供资金和其他资源给愿意提供保存服务的机构。目前有各种市场和非市场机制，比如定价模式，强制收费或税收，志愿活动和慈善捐赠。无论选择哪种机制，都必须支持资源的持续流动，使得长期保存目标能够实现。仅仅使保存资源可用还不够，应当尽可能有效的使用这些资源。从这个意义上来说，效率并不意味着偷工减料，而是获得分配给保存活动的资源的最大价值。例如，我们应当努力通过更大量的保存活动来分散成本，以利用规模经济。我们也可以尝试通过相关的不同服务来分散成本，以利用规模经济，比如，在同一个资源平台上提供定位保存服务和终端用户访问服务。

4.2.2.5 组织

要点：

- 清楚的认识数字保存环境中可持续性的含义，对于设计可持续性战略至关重要。
- BRTF最终报告确定了五个可持续发展的条件：

(1) 价值：在对特定的数字资产进行长期保存的过程中，清晰明确地指出利益相关者的利益。

(2) 选择：资源是有限的，我们必须优先考虑并做出选择，在可能的情况下，应选择保存对未来有最大利用价值的数字资产。所作出的选择性决策应定期回顾和重新评估。

(3) 激励：可持续数字保存需要愿意赞助或执行保存过程的利益相关者，这依赖于利益相关者对保存活动的强烈动机和激励。

(4) 资源：保存活动需要丰富的资源来实现长期目标。需要建立机制，让从数字保存中获益并且愿意为此付费的用户提供资金和其他资源给愿意提供保存服务的机构。一旦获得保存资源，应尽可能高效的使用。

(5) 组织/管理：应选择适合预期操作条件的保存组织形式。管理机制也需要明确保存目标、制定实现目标的策略、分配保存责任以及评估结果。

4.2.3 关键实体

任何数字保存活动周围的经济环境都是复杂的，涉及到许多因素，这些因素在相互影响的过程中形成了必须运作一个特定活动所需的环境。了解实现可持续发展的环境以及影响，是具有挑战性的。幸运的是，可以将存在于所有数字保存环境周围的经济环境提炼出一些关键因素。全面了解这些因素及其属性是构建一个成功的可持续性战略关键的一步。

4.2.3.1 数字资产

要点：

- 数字资产被认为是在未来持续一段时间中有价值的数字对象。
- 数字资产具有几个共同的核心属性，这些属性会影响可持续发展规划。
- 因为数字资产具有长期使用性，但也会贬值，因此规划者必须考虑到保持数字资产在较长的一段时期内可使用的“总拥有成本”。可持续性保存活动需要流动的资金，而不是一次性大量的资金。
- 数字资产可以被一个利益相关者保存，但是可以被许多用户同时使用。因此，用户贡献保存的资源动机没有那么强烈。这种搭便车问题使得从受益者中收集足够的资源来维持保存活动变得很困难。一种解决办法是，如果受益者不贡献保存的资源，就不能访问数字资产。

4.2.3.2 保存过程

要点：

- 保存过程是指一系列保持数字资产在较长的一段时间内可使用的活动。
- 保存过程的价值来源于它能够提供数字资产价值的的能力。保存过程在经济方面的需求是一种派生需求，在这种情况下，它来自于保存数字资产的需求。
- 保存决策是路径依赖的。在经济生命周期的某一阶段做的决策会影响后一阶段的决策制定者的选择。

4.2.3.3 利益相关者

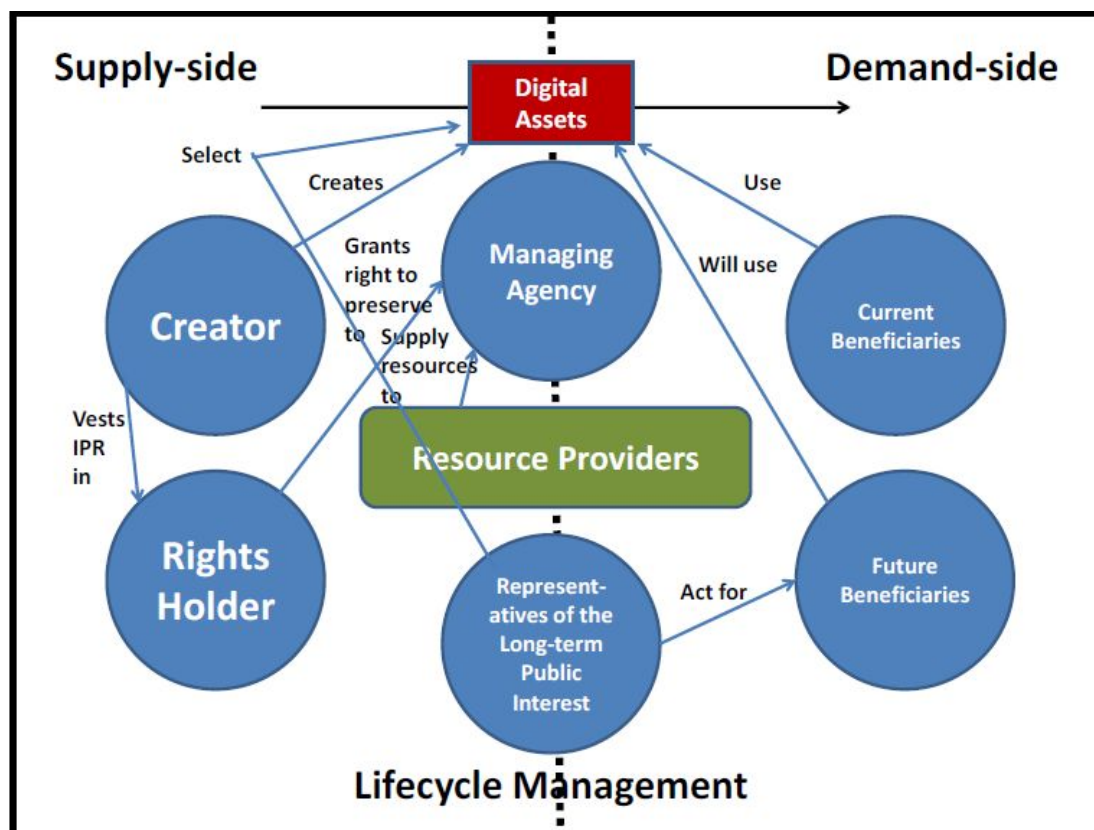


图 5-利益相关者的关系示例

要点：

- 和许多数字保存活动相关的广泛的利益相关者的利益，以及利益相关者之间的关系，深刻地影响了实现可持续发展的前景，同时也形成了最适合当前环境的可持续发展战略。
- 利益相关者的生态系统指明了在长期数字保存中利益相关者的主要角色，并阐明了他们之间的重要关系，因为他们与可持续性相关。
- 利益相关者可分为三大类：供应方（创建者；所有权者）；需求方（当前受益者；未来受益者）；生命周期管理方（管理机构；长期公共利益的代表；资源提供者）。忽视考虑利益相关者的利益和参与的可持续发展战略可能会在经济生命周期的某一时刻失败。
- 我们可以把利益相关者的生态系统理解成三个抽象的层次，综合起来就是：阐明了利益相关者在特定的数字保存环境下的组织形式，以及机构如何避开影响可持续发展的经济不确定性。第一层是与大多数数字保存活动相关的利益相关者的角色类型；第二层描述了在给定的数字保存环境中利益相关者之间现有的关系；第三层描述了在不同的个体和组织中利益相关者角色的分配情况。
- 认识到利益相关者作为不同的实体和角色之间的差别是重要的。一个组织（或个人）可以同时扮演多个利益相关者的角色。

- 利益相关者关系网中保存角色的分布对于识别固化在在利益相关者生态系统的某一种特定配置中的经济不确定性至关重要。

4.2.4 经济不确定性

4.2.4.1 规避

一些组织由于极其厌恶风险，所以不愿意承担任何重大风险。从这个意义上来说，他们可能会停止或不会首先尝试那些总被视为“风险”的活动。例如，孤儿作品是指那些无法查证版权所有者的作品。从法律上来讲，使用这些作品或者对这些作品采取一些保护措施都可能被视为侵犯版权，这可能就是一个经济不确定性。在这种情况下，将实施一种严格的规避政策，即只保存所有权状态明确允许可以保存的资源。不确定性可以通过只保存公共领域的作品或者不允许用户访问仓储库（如，消除存档资源的访问）来进一步降低。这个例子表明，为了规避不确定性，存在于规避策略中的一个重要权衡是：要么减少存储量要么减少利益。

4.2.4.2 缓解

当我们采用某项决策后，出现了一些严重的经济不确定因素并且难以避免这些不确定因素对机构的影响，那么这时就应该采取一些措施来减少或缓解这些不确定性。这里介绍两种有效的缓解措施，这两种措施最好一起采用，当然，如有必要，也可单独采用。

这两种方法可以用来降低不确定事件发生的概率，也可以减少当不确定事件发生时对机构的影响。当然，具体的方法还需要依赖于具体的环境，但是下面所列的这些例子还是可以说明这些方法的：

- 不确定性：用户失去了对机构的信任；缓解方法：改善保存流程，执行质量保证和认证过程（如获得“数据认可印章”）。
- 不确定性：用户对所保存的数字资源是否具有长远价值存在疑虑；缓解方法：可以提出一条这样的策略——使用户知道最初对数据的保存是轻量级的和廉价的，然而足以提供在未来需要时对数据做进一步的保存。
- 不确定性：机构对数据保存的权力存有争议或是没有保存的权力；缓解方法：在保存之前寻求相关的法律建议；办理一些许可证以及免责声明；对由保存/访问权力产生的关注点及时做出回应。
- 不确定性：保存活动所需的资金链条可能会中断；缓解方法：争取获得多重资金链条的供应，以此来减少当一条资金链条中断时对保存活动可持续发展的影响。

应该注意到在所有列出的这些例子中，无论是不确定因素发生的概率还是其带来的影响都不能完全消除。然而，这些减少不确定因素发生概率或影响的缓解措施反过来可以帮助机

构把对不确定性活动所带来的直观花费/效益比率调整到一个更为合适的程度。这些措施可以使不同的决策在具体的保存活动中继续实施，而不是采用回避策略。

4.2.4.3 转化

机构可以通过几条途径来实现转化策略。一种途径是通过类似于保险的政策，即需要通过高价的保费来换得，这样就可以使投保人的风险由保险公司来承担。当然，保险公司承接的许多风险实际上绝大多数都不会发生。许多保险公司覆盖（承接）了许多类型的风险，然而，不幸的是，我们并没有意识到任何专门针对数字资源保存的保险。

另外一种重要的转化方式是把风险或不确定因素简单地转移给其他的机构。例如，出版商与 Portico 或 CLOCKSS 签订的协议实质上是转移，因为这些出版商把数字保存的责任（与随之而来的不确定性）转移给了这些机构。那么为什么其他的机构愿意接受这种转移？比如，承担那些其他机构不愿担负的一些风险？

其中一个原因可能是不同机构有不同的任务，一个机构可能不把长期保存作为它的核心任务（或是商业模式），因此它就不愿意承担随着保存而出现的那些不确定性，然而别的机构可能把保存作为它的主要任务，并把随之而来的不确定性简单地当做是“做生意的成本”。

4.2.4.4 接受

要点：

- 在经济周期内存在无数的经济不确定因素。我们不可能消除所有的不确定因素，我们只能期望尽可能多地消除这些不确定性以及制定合适的计划与应急措施来缓解它们的影响，这些是所有可持续发展战略的必要组成部分。
- 对那些直观感知到的不确定性，有四种应对的方法：规避（即遇到不确定性，立即终止活动）；缓解（即减少不确定性发生的概率或是降低不确定性发生后其带来的影响）；转移（把不确定性转移到另外的机构，如保险公司）；接受（认识到了不确定性但是还是“忽视”它）。
- 一些规避风险的机构可能不情愿把他们暴露在任何具有风险的情形下。但是这些规避策略中往往存在一种重要的权衡，这些规避策略以减少容量和（或）利益的方式呈现，以达到规避不确定性的目的。
- 缓解策略能减少不确定事件发生的概率或是当不确定性事件发生时，减少对机构所带来的影响，或是两者都可。缓解策略不能完全消除不确定性发生的可能性或是其带来的影响，但却足以减少到使机构计划者能够接受的程度。
- 转移策略能够通过诸如保险公司来实现，或是通过把风险活动转移给其他的机构来实现。其他的一些机构由于这些风险活动是他们工作的重心而乐于接受这种转移。

- 接受策略指机构已经认识到不确定性的存在,但是没有采取任何措施来规避、缓解和转移这种不确定性。接受策略的优点在于不像规避、缓解、转移不确定因素那样需要额外的费用;其缺点在于如果那些不确定因素真的出现的话,那么这些不确定因素就会直接、完全地影响机构。
- 规避、缓解、转移策略是需要成本的,这就需要衡量其带来的效益。最佳做法包括在可持续发展战略中,需要解决哪些经济不确定因素,以及对采用哪种方法来解决它们。

5 ESRM在4C项目中的角色定位(略)

6 实践与工具(略)

7 往后的工作(略)

编译自:

<http://www.4cproject.eu/community-resources/outputs-and-deliverables/ms9-draft-economic-sustainability-reference-model>

(多人编译,多人校对)

【资料推荐】

DCC: 《国际数字保存期刊第8卷第2期》

DCC 机构的国际数字保存期刊第 8 卷第 2 期已经发布。这期内容包括来自于 2013 年国际数字保存会议上的 15 篇论文以及 3 篇同行评议论文。

这些文章侧重于从机构、学科、国家的视角来对进行研究数据管理(RDM)。有一些文章描述了英国、美国的大学是如何支持及实施研究数据的管理工作,这些工作会吸引其他机构也加入到相似的工作中来。还有一些文章描述了在德国、新西兰、美国有关学科数据中心的工作。另外的一篇文章则描述了在一些机构知识库的实体建筑及基础设施中如何形成一项基于主题的虚拟存档。有两篇文章谈到了当研究数据管理应用于艺术作品时应该如何审视,而艺术届的研究者并不考虑这些数据。最后三篇文章描述了国家概述,其中有两篇是关于英国的发展,一篇关于澳大利亚的发展概述。

研究数据管理的主题也在这些文章中有所体现,如其中一篇文章呈现了一项关于美国埃默里大学数据管理行为及期望研究的调查结果。另外,也有关于海量数据出处的探索研究,这些研究特别参照了一些“临床环境”。最后的一篇文章描述了一个机构如何运用 iRods、Cheshire3 及 Multivalent 来建立一项稳固的分布式的数据存储服务。

数字保存期刊第 8 卷第 2 期全文获取地址:

<http://www.ijdc.net/index.php/ijdc/issue/view/17>

编译自:

<http://www.dcc.ac.uk/news/international-journal-digital-curation-volume-8-issue-2>

(陈瑶编译, 唐果媛 吴振新校对)

《提供研究数据管理服务》

随着全球主要的研究机构的资助者要求研究机构展示他们是如何保存和分享研究数据, 研究环境正在发生改变。然而, 结构化研究数据管理实践还只得到初步发展, 而且现在这些服务建设还处于一个实验性的阶段并需要一些方法的模型和标准。本书作为开创性的指南, 会指引研究者、机构和政策制定者完成研究数据管理服务的创立及高效地运行。

本书主要主题:

- (1) 变化的混合体: 提供研究数据。
- (2) 提供研究数据管理服务的选择与方法。
- (3) 谁在处理数据? 有关角色、职责和能力的范围。
- (4) 研究数据服务可持续发展的途径: 从范围界定到可持续发展。
- (5) 研究数据管理基础设施与服务的范围和组成部分。

《提供研究数据管理服务》这本书于2013年12月20日出版。

编译自:

<http://www.dcc.ac.uk/news/new-book-delivering-research-data-management-services>

(陈瑶编译, 唐果媛 吴振新校对)

【专题报告】

最佳实践交流大会

NDIIPP 一年一度的最佳实践交流会(BPE)于美国当地时间 2013 年 11 月 13—15 日在犹他州美丽的盐湖城举行。今年的会议提供了充分的分享和讨论的机会, 将犹他州档案和记录服务局和犹他州周围的同仁们聚集在一起。

BPE 参会人员主要集中在州和地方政府图书馆、档案馆和记录管理社区。因此, 会议往往侧重于解决实际问题的方法。这种实用的精神在前参议员 Robert F. Bennett 的开幕主题演讲中得到了体现, 他鼓励参会人员与他们的立法者和资助者密切合作, 以找出数字监管的解决方法。关于如何成为一个成功的倡导者, Bennett 提出了三个重要想法:

- (1) 永远不要让别人做不能实现他们最大利益的事;
- (2) 永远和善;
- (3) 不要与其他人的预算竞争。

实用建议随处可见。来自 Multnomah 县的 Jenny Mundy 描述了一个可协调的后续规划的工作, 以帮助他们解决在招聘过程中的关键需求。会议议题“使美国法律在现在和未来都可以利用”吸引了来自犹他州立图书馆、犹他州行政法规局的参会人员以及来自法律信息保存联盟的数字保存先锋 Margaret Maes, 他们热烈的讨论了保存数字法律信息的当前做法。

来自亚利桑那州立图书馆的 Linda Reib 谈论了他们在为州立档案馆电子记录资源库寻求可持续性资助时面临的挑战, 同时展示了他们可帮助人们理解陈旧格式的辅助视觉设备。北卡罗来纳州立档案馆讨论了他们在保存当选官员和国家机构的社会媒体中的工作。

第一天下午会议的主要议题为 2014 年国家议程, 内容为利用数字监管和头脑风暴方式来利用 BPE 的能量, 以支持国家数字监管联盟的工作。

美国电影艺术与科学学院的 Milt Sheftner 提出了有关电影制作所面临的数字保存挑战的个人见解, 并且展示了两个 NDIIPP 报告, 数字困境和数字困境 2, 报告中提出了主要制片厂和独立制片人在保存数字电影方面面临的主要挑战。

Sheftner 展示后, 播放了纪录片-奇妙的影像 (These Amazing Shadows), 影片讲述了美国国会图书馆国家电影注册处的历史和其重要性。

第一次最佳实践交流会议是 2006 年在卡罗莱纳州威尔明顿举办的, NDIIPP 一直参与其中, 从那时起, BPE 社区在解决数字监管问题上一直都有新的进展。在“在 BPE 发生的所有事情都保存在 BPE”, 继续向人们展示 BPE 社区的工作也是很重要的。

编译自:

<http://blogs.loc.gov/digitalpreservation/2013/11/the-best-practices-exchange-conference-is-no-secret/>

(唐果媛编译, 王敬 吴振新校对)

第十届数字对象长期保存国际会议综述- iPRES 2013

说明: 本文已经被“现代图书情报技术”杂志社所采用, 将于 2014 年第 3 期公开发表。吴振新、付鸿鹄、李文燕、陈瑶、孔贝贝参加了本文的编译整理工作。

iPRES2013 (数字对象长期保存国际会议, International Conference Preservation of Digital Objects), 即第十届 iPRES 会议与 DCMI2013 (Dublin Core Metadata Initiative), 即第 21 届国际会议, 于 2013 年 9 月 2-6 日在葡萄牙首都里斯本联合召开。

iPRES 作为长期保存领域学术水平最高的盛会, 自 2004 年在北京首次召开后, 每年召开一届, 会议吸引了众多的数据创建者、管理者、使用者以及从事数据保存流程开发和培训的研究人员, 对全球长期保存活动的发展产生了巨大的影响, 发挥了积极的促进作用。

本次联合会议共吸引了来自 37 个国家的 392 人参加。会议由三个阶段组成: 第一阶段 iPRES 和 DC 汇集了丰富实用的培训专题, 共计 223 人参加共 10 场培训教程; 第二阶段包括 3 天的主会议, iPRES 在为期三天的主会议中, 共提供了 15 个 session 共 38 个报告, 涉及网络存档、对象及数据保存、合作与共享、机构知识库、国家战略、评估、管理、技术、基础设施、数字化馆藏、保管 (curation) 等主题; 第三阶段共计 246 人参加 9 个主题的研讨会, 致力于解决实践和具体的问题。

元数据作为长期保存领域中一项非常重要的内容, 是很多保存专家关注的重点, 此次的联合会议非常有利于两个领域的研究人员相互沟通。作者按主题对会议内容进行了汇聚, 提供给大家参考。

1 Web Archives

来自 Tessella 公司的 Rory Blevins 等人在报告中描述了一种用于提高大型 web 存档可伸缩性的机制。他们在早期的摄入过程中收集 Web 归档对象的格式信息, 并将这些信息作为的容器对象 (如 web 存档生成的 WARC 或 ARC 文件) 的属性存储, 在执行保存操作 (如格式迁移) 时就可以只查询容器的属性, 不必深入查询容器中美国实体, 这样会大幅减少需要查询的实体数量, 从而提升大型 web 存档系统的性能和可扩展性。

来自希腊的 Vangelis Banosy 介绍了一个新的概念-Website Archivability, 分析了哪些因素应该被考虑用于计算整个网站可存档性, 网站属性如何与这些因素对应, 并给出了 web 存档网站可存档性即时评估方法的整个流程, 提出了一套用于评估的属性指标, 旨在利用这样的一整套机制促进 web 存档内容的自动化质量控制。

来自德国的 Hendrik Kalb 等人介绍了他们通过德尔菲法 (也称专家调查法), 调研 web 存档与数字图书馆系统间交互性以及分析结果。调研从用途 (集成检索、访问、交换、复制) 和应用范围方面 (国家、机构) 以及关注目标对象 (数据对象还是元数据) 的差异分析了互操作的不同目的; 从获得国际领域的关注、预算和技术、共享工具、遵循公共标准有益于资源的长久性等方面分析了互操作带来的好处; 同时分析了在标准、工具和实施、组织、法律以及处理互操作性的方法等方面遇到的阻碍, 发布了专家建议的解决及改进方案, 明确了由于内容互操作、新型复制媒体和资源等带来的挑战。

来自葡萄牙国家科学计算基金会的 DanielGomes 等人将他们的一系列的成果在报告中
进行描述, 他们开发了软件将多种格式的数据合并为统一 ARC 格式, 同时处理了与长期保
存相关的 URL 等问题, 通过完善和增强 wayback 和 nutchWAS 的功能来满足他们在效率和
功能上的需求, 同时基于 Hunspell 开发了查询推荐机制, 使用户获得更准确的查询结果,
利用负载均衡机制来提升查询能力, 利用日志分析来评估检索结果的准确性, 开发了用于研
究的数据分析和访问工具, 提供测试数据集和 API 接口供公众使用。

2 对象及数据保存

奥地利 AIT 的 Andrew Lindley 从技术角度分析比较了两种保存关系型数据库工具:
SIARD 和 CHRONOS, 通过在实践中运用这两种保存工具, 作者指出了在保存关系数据库
中遇到的一些问题和挑战, 并对十五个方面的评估结果进行了对比分析, 他们认为更具灵活
性的 CHRONOS 工具更适合数据库的长期保存, 同时还需要更多地获得与所存储资源有关
的环境信息。

bwFLA 项目按照分布式和共享的方式开发了用于长期保存的工作流和工具, 不同保存
机构都可以利用该项目的技术解决方案和分布式服务模式, 规划相应的措施来实现特色数字
对象的长期保存。作者^[7]以特定领域-CD-ROM 的数字艺术收藏品为实例, 介绍了采用 bwFLA
项目的仿真即服务 EaaS (Emulation-as-a-Service) 框架及其抽象的工作流进行大量复杂数字
对象保存和访问的工作。他们整合相应的工具, 提供了一个实际的工作流, 能够快速遍历大
规模未知数字对象, 创建用于呈现描述的技术元数据描述, 生成内容元数据, 进行复杂对象
的集合组织和提供访问。

BBC 目前拥有 15PT 的影音文件, 从 2007 年开始 BBC 研发部开始对这些数据进行数字
化归档保存。Thomas Heritage 介绍了 BBC 对录影带数据实施长期保存经验和教训。BBC 存
档系统的核心处理过程分为 5 步骤: 预处理、处理、克隆、集中克隆与临时存储、抽样检查
/完全质量控制。BBC 首先把存储在可能过时的 LTO3 磁带上的数据迁移到新的存储设备,
并用基于文件的方式存储从磁带上迁移过来的数据, 同时按应用目的划分了两个系统—TV
历史保存库和数字仓储库, 后者则用来管理数据质量、提供数据搜索和访问以及数据输入输
出等功能。

Monika Linne 介绍了社会科学研究数据的保存、开放、分享项目—datorium。作为一个
基于 Web 的数据分享仓储, 它可以为研究人员提供工具复用其他学者的数据, 同时也提供
了标准的元数据模式: DC 元数据和自定义元数据模式, 帮助研究人员按照标准的方式记录

研究数据和元数据, 并提供数据访问控制、核查以及保存服务。该项目一期, 研究数据长期保存在 GESIS Data Archive 中, 第二期将由其自身进行数据保存。

在 SCIDIP-ES 项目经验之上, Giuseppa Caruso 等人设计了一个新的长期保存数据价值模型, 旨在评估随着时间推移地球科学数据价值的价值变化。该模型在 5 个核心因素(属性)——可用性、可访问性、完整性、完备程度(completeness)和可用程度(usability)的基础上, 综合考虑长期保存过程中各种活动的影响, 用成本/效益分析的方法, 兼顾成本/效益正负两方面的因素来权衡评估数据在长期保存过程中不断变化的价值。

为了长期地维护数据密集型科研所产生的数据, TIMBUS 项目开发了“保存过程”框架, Stephan Strodl 等人通过保存音乐分类作为应用案例对该框架进行了详细介绍。该框架将保存过程划分为规划、保存和重新部署三个阶段, 通过分析执行环境、识别其中的依赖性, 为每个阶段提供指导方针和执行步骤。此框架基于风险管理的方法, 不依赖于任何具体的领域, 可以应用到不同的环境(如商业环境或 E-Science 领域)。

同样来自 TIMBUS 的 Elisabeth Weigl 介绍了“软件托管技术框架”, 从技术的角度介绍了如何在人工辅助下自动验证被托管的软件。传统的软件托管保存主要关注软件的源代码和文档的材料组成, 然而软件工程的内容却不仅仅限于此, 还包括不同的部件, 例如编译器, 文本脚本等。作者采用 TIMBUS 的保存过程框架, 按规划、执行、和重新部署这样三个阶段来实施保存。他们通过扩展一款开源的软件质量工具, 结合软件托管的具体情况, 开发了一个软件托管技术框架, 提供了包括组件完整性、文档质量、法律权利等多个指标, 成功地检查涵盖从源代码保存到文档存储各项内容, 并通过 aTunes 验证了其有效性和性能。

Isgandar Valizada 等人介绍了 bwFLA 项目开发的分布式 EaaS (Emulation-as-a-Service) 实施和服务模式, 并讨论了如何提供可扩展的仿真服务的问题。作者描述了作为核心部件的仿真组件(emulation-components)如何与各种模拟器进行交互, 并提供必要的 API 和数据输入/输出服务。用户可以选择两种不同的基础服务: 直接与原始环境交互或者设置复杂的保存 workflow。EaaS 使用了 MASS (Metal as a Service) 来管理仿真机器, 根据用户即时需求实例化仿真部件机器, 自动地配置所需要的服务。同时 EaaS 还配备了分布式的认证和授权系统。

3 合作与共享

Daniel Simon 等人分析了如何在商业场景中应用长期保存来减少企业的风险。他们从企业风险管理的角度出发, 分析企业的风险, 从中识别出长期保存的需求, 把长期保存作为一

种对推进企业商业过程大有裨益的商业决策而非仅仅是必须的政策方案来看待,充分发挥长期保存的优势,帮助企业提高知识产权的透明度,确保对用户的长期支持,提升企业竞争优势。作者把企业风险定义为识别风险、分析风险、评估风险、长期角度对待风险、检测风险、确定风险管理的角色和职责六个阶段,成功地将企业风险管理(ERM)融入了长期保存的过程中。

Michelle Lindlar 等利用 Goportis (The Leibniz Library Network for research information, 莱布尼茨图书馆研究信息网)联盟的长期保存协作的经验和教训,从三个方面总结概括了长期保存协作网络的原则和需要注意的问题:在地理位置方面,联合位于同一国家的组织开展长期保存可以简化立法问题,共享联合目录可以协同不同机构的工作流;在组织类型方面,需要协调不同组织的观点,对于类似的组织类型使用类似的词汇,同时要注意不同的组织文化可能会形成的“隐患”;在藏品方面,不同的组织拥有类似的集合材料可以减少协作的整体成本,而相似但主题不相同的藏品则可以提高长期保存的协作性。这些经验为其他不同机构共同操作数字化保存系统开展长期保存合作提供借鉴和指导。

Jochen Rauch 则介绍了 ENSURE(Enabling kNowledge Sustainability, Usability and Recovery for Economic value)项目的初步成果,该项目开始于 2011 年 2 月,现已开发了一个包括三层架构的原型系统:通过插件层提供特定的功能,如格式管理、合规性、完整性检查以及访问特定的存储云;在服务层,使用插件创建 OAIS 解决方案,提供运行时的面向服务的架构(SOA)框架,满足用户包括经济需求在内的多方面的需求;在第三层,通过使用成本/质量分析引擎,利用配置器和优化器,创建和评估保存解决方案。该项目基于生命周期管理,使用本体来识别和触发数据对象必要的转换,以确保它们的长期可用性,同时还使用了云技术以增强其灵活性、可扩展性和降低启动成本。

来自 APARSEN 项目的 Barbara Bazzanella 等人对与数字资源长期保存相关的包含有互操作问题的项目和倡议进行了分析,他们收集了 4 类共 13 个互操作场景,根据分析的结果定义了一套矩阵,它结合互操作性的不同层面(例如,语法,语义,组织机构)与数字保存的不同方面(例如永久标识符、元数据、起源)和相关的互操作对象和模型,提供了一个数字化保存的互操作性的概念框架,该框架可以作为一个方便实用的互操作性解决方案的起点,为长期保存设计具体的互操作服务。该矩阵描述了 58 个互操作解决方案,归纳为 8 类:永久标识符,起源,数据质量,元数据,元数据收割和信息交换,认证、授权、权利,保存模型和服务,研究数据保存、发现、获取、重用和引用。此外,还制定了 4 套建议以推动数字资源长期保存系统中互操作性的实现。

长期保存领域没有实际使用的受控词表,不足以支撑新型的可扩展的保存生态系统的需求。Hannes Kulovits 等人介绍了这样一个开放的生态系统的的核心需求和要素,运用关联数据原则建立一个由领域词汇、领域实体、策略实例、策略词汇构成的简单的核心模型,提出一个受控词表元素核心集,其中涉及保存质量、目标、政策、组件,并演示了怎样在 SCAPE 生态系统中应用该模型以改进信息共享、推理和发现。

Paul Wheatley 等人介绍了 SPRUCE 项目采用面向社区的方法支持和维持基层的数字资源长期保存活动。本文介绍了一些用于保存数字内容的相关软件工具,包括解决 PDF 格式保存问题、合并文件格式识别等相关工具。此外,SPRUCE 项目提供在线和远程协作。项目正在建设一个资源工具包以帮助管理者和从业者资助和维持数字资源长期保存活动。

4 国家战略

Aileen O'Carroll 等人通过访问、咨询爱尔兰数字仓储库的用户来了解数字仓储库存在的问题,这些问题集中体现在政策的不完善(如版权问题)及用户具体需求不明确。通过分析得到的结果,作者提出了一些关于数字仓储库运行理念和具体定位的改善建议。

法国国家图书馆开发了 SPAR 系统以取代多个存储系统,从而使得长期保存变得更为简单、经济、有效,并为第三方用户提供存档服务。SPAR 基于"路径-tracks"和"频道-channels"的概念来组织内容和管理异构集合,按照一定规则创建路径和其下属的频道,每个频道都有一整套服务层面的协议,详细定义摄入、保存、访问的条件,系统至少为每个向 SPAR 系统提交数据的第三方机构建立一个新的频道。为了使得摄入、数据管理、存储和访问模块更加规范化和易于维护,SPAR 开发了预处理模块,将各种提交数据转换为统一的 SIP,使得所有处理 SIP、AIP、DIP 的功能都能用于所有路径和频道,论文还详细分析了共享服务如何利用开发的的基础架构为第三方用户定义存档服务、价格和处理过程。

新西兰国家图书馆的 Jan Hutař 介绍了该馆如何将其保存的数据从 fodera common 系统迁移到 Rosetta 平台。工作包括分析哪些数据需要被迁移、迁移的数据准备、执行迁移和所使用的工具。

5 评估、管理、保管

Roman Graf 在报告中介绍了一种对不同文件格式保存风险的自动评估方法。这种方法将风险程度与对风险的自动计量纳入到风险因子的定义中,目的是把来自关联数据仓库的知识作为风险分析的基础并加以利用。这种自动评估方法可以帮助数字保存领域的决策制定。

报告还介绍了一种用于聚合大量可靠的不同格式文件的工具,此外,文中利用了各种技术来分析不同格式文件的风险。

Diogo Proença 介绍了评估数字保存系统保存能力的方法与具体步骤。首先,作者通过假设数字保存系统具有一般系统所具备的质量特征(源自 ISO25010),包括:功能适用性(Functional Suitability)、运行效率性(Performance Efficiency)、兼容性(Compatibility)、可用性(Usability)、可靠性(Reliability)、安全性(Security)、可维护性(Maintainability)、可移植性(Portability),作者从数字保存的角度出发,说明这些质量特征在数字保存系统中如何体现并将之作为评价一个数字保存系统保存能力的指标。基于上述质量特征,作者提出了用于评估数字保存系统保存能力的方法及具体的步骤,包括:启动评估、规划评估、评估概要、数据收集、数据验证、评估分析、公布评估结果,并详细介绍了各评估步骤要完成的具体任务,并通过一个系统实例验证了该方法的具有实用性。

保存策略指导着一个机构的保存活动,应该被定期检查和更新。欧盟所支持的 SCAPE 项目设计了一个政策框架来支持机构构建保存政策文档。该框架是一个三层策略模型,高层为指南性策略,描述机构对于数字集合的共同长期保存目标;中层为保存过程策略,描述机构为了实现高层长期保存策略目标所采用的方法,这些策略需要足够详细以便于输入流程和做工作流程的设计。底层为控制策略,针对指定集合、指定的保存动作和指定用户社区,可供人工阅读以及机读,以便用于自动的规划和监测工具,确保所选择的保存操作和工作流程满足指定集合和社区的具体要求。

同在 SCAPE 项目的 Michael Kraxner 等人则介绍了所开发的支持仓储库长期保存生命周期的规划和监测工具集,包括内容分析 c3po、保存规划 Plato 和监测 Scout。作者描述了如何使用这套工具在真实的环境中实施可扩展的长期保存监控和控制,使保存者以系统和半自动的方式来管理,并以流线行的方式持续执行长期保存进程(保存生命周期),以此减轻一些大型数字保存的有效性问题的。

Luis Faria 等人进一步描述了 SCPAE 项目的保存监测系统 Scout。作者所在的荷兰国家图书馆需要对出版商和期刊进行完整性监测,但这些知识主要在出版商的网站用自然语言描述,很难自动监测,他们利用信息抽取技术,采集保存相关的信息,汇集到中心知识库,转换成可机读和分析的结构化和规范化信息,通过评价这些信息以发现风险和实施监测,从而实现了基于本体知识库的半自动的 web 信息保存监测。

遗忘管理是来自德国汉诺威大学的 Nattiya Kanhabua 在本次会议中提出的一个基于人脑遗忘功能的创新概念。通过动态、多维度的信息价值评估,记录信息重要性的变化,实时地

触发“遗忘”行动,包括多项操作,如聚合和自动摘要,调整搜索和排名的行为、消除冗余以及删除等。通过将资源选择模型化,使遗忘管理成为一个动态调整关注度的功能。该项目整合了提供预处理功能的信息处理系统和后端的长期保存系统,中间以遗忘管理功能作为连接层,通过感知进展实现情境记忆,触发遗忘管理进而实现协同保存。

密西根大学的 Devan Ray Donaldson 介绍了在特定环境下保存库管理者认定保存库的可信赖性的相关定义、指标和验证过程。分别从传统角度和终端用户视角对保存库的可信赖性进行了分析,介绍了一个正在进行的研究项目,该项目通过专题小组调研特定群体对可信赖性的理解,探索建立、测试、评估一个可测量的可信赖度指标。

6 技术、基础设施

技术部分包括来自 BitCurator 项目的两个报告,Christopher A. Lee 介绍了开放资源平台 BitCurator environment 的设计和功​​能,该项目提供了一套开放源代码的软件,允许利用数字取证方法收集机构的数字资料,以推动各种数字资源的长期保存目标。Kam Woods 则介绍了通过扩展和重用数字取证工具,生成符合数字保存元数据模型的元数据。整个过程分为两步,第一步是从平台中抽取与修改元数据,包括捕获图像元数据和文件系统元数据,并加以分析;第二步是转换所得元数据,使之能支持数字资源的保存和获取。文中选取了该平台的数字取证元数据,将其中部分用于记录与保存和访问相关的 PREMIS 事件中。

在数字化工程中经常遇到图像质量的问题,来自英国国家图书馆的 Sean Martin 对高质量图像的噪音和经压缩和释放后的图像的噪音进行了比较,发现在成像过程中的噪音无损图像的质量,可以使用更紧凑的压缩图像来代替现行的无损图像文件存储,对图像质量没有影响。如果能予以应用,可降低 30-70% 存储成本,这对大型数字化项目是个很好的决策。

来自英国国家图书馆的 William Palmer 等人介绍了他们对几种图形迁移工具的性能比较。他们选择了排名前三的工具 Kakadu、OpenJPEG、JasPer,利用 26GB 的 TIFF 图片,按照一定流程进行对比分析,并从满足压缩比需求、多轮编解码过程的速度及图象质量角度给出了分析结论。

Barry M Lunt 等人介绍了一种通过改变记录层的材料而实现永久数据存储的方法,他们选用目前 DVD 的标准,采用无机的、稳定的材料层做为光盘的记录叠层,制造了与 DVD 兼容的光学磁盘 M-Disc--可被 DVD 驱动设备读取,它使用标准的光学记录与再现激光器以减少读取时的抖动,这种技术具有非易失性并可被无限次的记录,同时不使用任何粘合剂就

可把被激光检测到的东西粘到材料上,使记录层在聚酯基材上沉积下来,目前还未测试到能够把记录材料剥离的方法。

Tomasz Miksa 等人则介绍了 TIMBUS 项目的另一项成果--保存和重新部署过程的验证框架 VFramework。该框架具有独立性、高度抽象和可定制的特点。整个框架包含两个序列的活动:第一个序列在原始环境中运行,包括描述原始环境、为保存进行系统准备、描述验证集、捕获验证数据,每一步的执行结果都存储在 VPlan 中。第二个序列在重新部署的环境中运行,包括为重新部署进行系统准备、捕获重新部署性能数据、比较和评价,完成每一步骤必须的信息从 VPlan 中获得。该框架是为了验证重新部署的过程是否符合预期的目的。

Costis Dallas 等人详细介绍欧洲数字图书馆 MoRe 系统的主要功能,该系统混合了聚合、数字保管和长期保存的功能,采用有效的方式屏蔽了内容提供商的变化,聚合了很多考古内容社团的数据,并整合到一个通用的元数据架构中。该系统为元数据质量监控和丰富提供了增值的数字保管服务,用于保证元数据的可靠性,此外,它提供了保存的工作流程,保证有效的记录保存所有交易和资源库的当前状态。

来自丹麦的 Eld Zierau 介绍了围绕建设应用 OAIS 参考模型的分布式数字资源长期保存 (DDP) 框架所开展的相关工作。这一框架将有利于未来采用高度分布式方式进行数字化保存的存档库的分析和审计。作者探讨了两种基本模型:分布式数字保存模型和 IR-BR 模型。IR-BR 模型描述了在多个遵循 OAIS 的机构知识库 (IR) 之间如何建立共享位库 (Bit Repository),其他有价值的模型也可以被添加到该框架中。

加拿大的 Peter Van Garderen 介绍了 Archivematica 项目致力于通过匹配现有的开源软件工具和 OAIS 模型创建了一个功能完整而且免费的数字保存系统。该项目通过仓储库接口与现有系统集成,采用一系列方法方便用户个性化定制 workflow,采取格式策略注册机制管理格式策略,并通过 FPR APIs 实现策略共享,并提供灵活的安装和部署策略选择。Archivematica 试图填补现有的仓储库管理应用与数字保存服务之间的差距而不是替代或复制他们的功能。

数字化过程中多个版本的保存导致存储容量需求快速增加,根据集合管理计划和内容策略,KB 部署了一个差异化存储策略^[9],通过建立物理对象和数字副本之间的关系,对集合的不同部分进行价值评估和风险评估,之后分别赋予不同的保存层次(共 5 层),从而了解有效运用集合管理的方式,有效平衡大型数字化项目的访问与长期保存带来的费用问题。

结 语

长期保存作为一项复杂艰巨的长期活动,不断面临新的挑战。历届 iPRES 都对于创新给予了关注,本次会议依旧关注着新出现的挑战和方案,创新的技术和方法,如遗忘管理;管理作为 OAIS 的重要功能依旧得到深入探讨,如基于风险管理的理论、长期保存生命周期的管理等等;对于数字资源长期保存的商业模式和附加价值的探讨也在不断丰富;长尾数字资源的长期保存开始受到关注;人员的培训和教育一直得到重视并在不断加强教育。

作者在全面浏览会议内容的基础上,初步总结了会议中几个热点主题:

- 网络信息保存和利用越来越受到关注。在本届会议中, Web Archive 不但作为首个主题出现,还占据了大会的两个专题讨论中的一个,供来自各国的存档工作者自由交流,探讨潜在的发展和存档数据的更多功用,如存档资源的规范化、存档内容的自动质量控制、存档资源共享、存档资源的访问和分析、应对大规模存档的挑战等等。
- 研究对象发生转移。本次会议众多论文探讨了利用已有成果(模型、 workflow、工具)来应对特定领域(诸如文化遗产、技术和科学的流程和数据、工程模型和仿真、医疗记录、企业流程和记录保存、网页存档、个人存档、电子采购)和复杂对象(诸如关系数据库、存储在 CDROM 及录音带上的音视频数据、艺术类数据、社会科学数据、商业数据、技术软件等)的长期保存挑战;
- 持续关注如何应对可扩展的大规模保存。多篇论文涉及大规模保存的基础架构、系统、数据存储、访问呈现等问题,应对大规模数据的自动工具不断出现,同时在专题会议中汇集了来自全球多个主要保存系统的管理和研发人员,就大规模长期保存的可扩展性进行了为期一天的讨论;
- 合作与共享成为保存领域永久的主题。占据了两个时段,策略、技术、工具、平台、元数据等方方面面,努力探讨地域间、机构间、学科间、集合间、系统间的协作与共享;
- 持续关注技术及基础设施的发展和应用。多篇论文涉及保存基础架构、系统、工具的发展和应用,重用和开源成为保存领域工具的两大特点,可以看出长期保存给开源系统和工具提供了一个良好的生态发展环境,大量开源系统和工具以及模块以开源形式发布和共享,间接地实现了成本降低和提升了投资价值。
- 保存监测的需求不断增强。随着全球多个项目和系统的部署和运行,对于 Preservation Watch 的需求和功能需求不断增长,本次多篇论文从理论、架构、系统、工具等多个角度进行了探索。
- 评估和评价继续保持热点。数据价值评价、系统可信赖性评估、保存格式风险分析、机构与系统的保存能力评估,评价成为保存活动中越来越重要的一部分。

- 国际重要的合作项目呈现高产出、高共享的特点。如 Timbus、Scape、APARSEN、bwFLA 等。

随着数字时代的飞速发展，整个社会已经无可避免地堕入数字宏潮中，如何保存人类社会的数字文明，长期保存领域面临着越来越多的挑战和难题，iPRES 会议就如同晴雨表，准确反映了社会需求和保存领域实践状况，同时又为广大的存档人员提供了高水平的国际交流和共享平台，有效地促进了全球数字资源长期保存的发展，为人类文明保存和发展做着不懈地努力。
