

i2010:数字图书馆*

常唯 编译

1. 引言

2005年4月28日,欧洲六个国家和政府联合致信欧盟理事会主席和欧盟委员会,倡导建立欧洲虚拟图书馆,旨在实现欧洲科学文化资源的共享。委员会非常欢迎这一计划,并将启动龙头项目——i2010数字图书馆项目,以促进该计划的实现。

本报告概述了数字图书馆项目的远景及其主要建设模块,同时,指明了项目的第一要素——处理资源数字化、联机获取及数字文化遗产的长期保存问题。报告分析了展示欧洲全部文化和经济潜力方面所面临的挑战,就克服当前欧洲各方各行其是的现状提出第一套行动方案。

2. i2010:数字图书馆

数字图书馆项目的目标是在联机环境中更容易地使用欧洲的信息资源。它依托于欧洲丰富的文化遗产,以及多种文化、多种语言、拥有先进技术和新的商业模式的社会环境。

数字图书馆是有组织的数字内容集合,其内容可以被公众利用。它既包括数字化的资源,如图书馆、档案馆收藏的图书和其它实体资料的数字拷贝,也包括原生数字信息——这类信息在科学信息中越来越常见,如数字出版物和大量存储在数字仓库中的信息。该项目包含数字化资源和原生数字资源两部分。

需要充分挖掘数字技术的潜力以求更广泛、更容易地获取信息,在这方面将遵循三个主要原则:

- 联机可获得性:这是公众、科研人员及公司能够从信息中获取最大利益的前提;
- 模拟馆藏数字化,使之可以在信息社会中得到更广泛的利用;
- 存储并长期保存,以确保后代能够获取这些数字资源,避免珍贵资源的丢失。

数字图书馆项目将在几个建设模块中体现这些因素。当前报告主要针对欧洲的文化遗产,而下一个关键领域是科学信息。在该领域,由于需要处理和存储海量的数字数据且只以数字格式出版的出版物数量迅速增长,因此具有自己的特征和动态性。《展望2006》的报告将列举数字图书馆在这个领域所面临的具体挑战,包括要求有支持高容量的基础设施,以及在欧洲层面上采取行动。

数字图书馆项目将进行一场在线咨询会,以应对影响数字化、联机存取和数字资源长期保存的主要挑战。各方的回应将成为今后可能发布的有关数字化和数字资源长期保存建议书的重要依据,同时还将纳入欧盟的其它相关项目,如版权框架评论。

总的说来:

- 当前的报告“i2010:数字图书馆”重点强调文化遗产(2005);
- 将开展一场有关数字化、联机存取和数字资源长期保存问题的在线咨询会(2005),回应将作为数字化和数字资源长期保存建议书(2006)的依据,以及欧盟其它项目如版权框架评论(2006)的依据。

3. 实现欧洲文化遗产的联机可用:风险在哪里?

3.1 文化和社会方面

欧洲图书馆和档案馆拥有丰富的资源——包括图书、报纸、电影胶片、照片和地图,这体现了欧洲丰富的历史、多样性的文化和语言。这些来自不同文化背景、采用不同语言的资

* 本文为欧盟委员会致欧洲议会、理事会、欧洲经社委员会和地区委员会的报告全文。

源的联机呈现,将使得公众能够更容易的欣赏到本国和欧洲其它国家的文化遗产,并将之用于工作、学习和娱乐。这是对欧盟在文化领域行动的补充,有助于其目标的实现。

3.2 经济方面

图书馆和档案馆是投资和劳动力雇佣的主要部分。2001年,欧洲的各图书馆共雇佣了336673名全职工作人员,拥有超过1.38亿的注册用户,这些对经济的影响十分重要。

对上述资源的数字化将会大大增加这种影响。资源一旦数字化,欧洲的文化遗产将成为网络交流发展的驱动力。它将为一些部门(如旅游业和教育界)的增值服务和产品提供丰富的可重用的原始资源。如果能够被合理的长期保存,这些资源可以被一用再用。此外,数字化过程中公司开发的新技术将产生大量的副产品。

4. 数字化

欧洲数量繁多、种类多样的图书馆和档案馆给人以深刻的印象。例如,2001年欧洲图书馆馆藏图书和合订本期刊(bound periodicals)的数量为2,533,893,879册。

将这些资源数字化的主要动因是用户可以在联机环境中使用它们。但是,在某些情况下,数字化的主要目的并不是为了使用,而是为了长期保存这些内容。最显著的是视听材料,其模拟格式随时间不断老化,这会导致珍贵资料的丢失。

对十个主要广播电台档案资料的调查显示:它们共包括1百万小时的胶片,1600万小时的视频记录和2百万小时的音频记录。欧洲全部广播资料的馆藏约是它的50余倍。其中的大多数资料是原生的、以模拟形式存储。由于材质的腐蚀、破损或存储于孤立的媒体上,70%的资料处于损毁的风险之中。由于存储介质老化的原因,欧洲每年大约损失1万小时的视听资料。

当前,欧洲馆藏中只有一小部分实现了数字化。所有成员国都在进行数字化活动,但是各国在这方面的努力彼此孤立、支离破碎,进展相对很缓慢。Google开展的数字化欧美5个主要图书馆1500万册图书的项目推进了数字化进程。如果能够按计划实现,Google的这一项目将远远超过任何成员国在国家层面上所做的努力,也会推进世界其它地区的数字化进程。例如,印度和中国有雄心勃勃的数字化日程,其内容覆盖不同语种的资源。

欧洲的数字化进程面临大量挑战,包括:

经济挑战:数字化是一项劳动密集型工作,成本极高,需要大量前期投资,在许多情况下这超出了资源拥有机构的财力承受范围。同时,数字化所有相关资料是一项不可能完成的任务,因此需要确定哪些资料需要数字化,何时对之数字化。

组织挑战:“一次数字化、广泛使用”的策略可以使所有参加的组织受益。必须避免重复劳动——多次重复数字化相同的作品或馆藏,只有在国家层面或整个欧洲层面进行持续的协同努力才能避免这种重复。

要实现数字化,需要新的工作方式。获得私人组织的支持并与之合作有助于数字化的实现,但是这种合作目前发展的还不够好。此外,数字化方面的投入必须随相关组织内部的变化而变化,包括改善相关工作人员的技能。

技术挑战:关键的技术挑战之一是需要改善数字化技术以使得数字化(视听资料和图书)可以实现低成本、高效率,且成本可以承受。为了数字化印刷版文本,需要改进自动化的图书和文件输入设备,同时需要更高性能的用于非英语的光学和智能字符识别技术。

在过去的10~15年中,英语光学字符识别系统已经在大多数印刷字体上得到测试和使用,其识别能力通过自动或半自动的拼写校正算法得到加强。但用于其它语言的类似系统功能相对较弱,这导致识别的成本更高,但质量却在下降。

法律挑战:数字化意味着制作一个副本,从知识产权的角度讲这存在问题。《关于协调信息社会的版权和有关版权若干方面的指令(Directive2001/29/EC)》(欧盟于2001年5月22日通过)预见到在公共图书馆、教育机构、博物馆或档案馆所进行的具体复制中存在着

例外。但这种例外不是强制性的，成员国在具体实施时存在着差异。法律对数字化结果使用上的限制是数字化的另一个障碍。

5. 联机可获得性

传统的图书馆服务是基于对其拥有的实体资料的借阅，这种方式不容易转移到数字环境中。在当前的欧盟法律和国际协议的框架下，只有公共领域或获得版权拥有者明确许可的数字化资料才能联机获得，因此欧洲数字图书馆将以公共领域的资料为主。

对于文献，这意味着只有早于 1900 年的作品可以使用，因为根据拥有者的死亡时间来计算，这些作品基本没有版权限制。但是即使已经不受版权保护的作品，其使用前景也不总是十分明朗的。对于某种作品，即使其本身已不再受版权法的保护，但它的不同版本或出版形式可能拥有不同的权利，如导论、封面或印刷样式。

因此，如果版权法没有实质性的改变，或者与每一位版权拥有者逐个协商解决版权问题这种现状不改变，那么联机图书馆提供公共领域之外的资料是不可能的。

在这种背景下，作品版权状态的澄清和透明化非常重要。有时，确定作品知识产权状态的成本高于将其数字化并提供联机可用的成本。对于那些孤生作品——难以确定作品拥有者的作品——更是如此。在员工工作报告及与之相关的联机咨询会中强调这些要素。

联机可获得性方面的问题并不仅限于知识产权。将资料联机公布并不意味着它就能被用户容易的获取，更不意味着它就能够被检索和使用，要做到这些，还必须提供恰当的服务以帮助用户发现并使用资源。这意味着对资源内容结构化的、高质量的描述，包括馆藏信息和内容本身，以及支持使用的其它信息，如注释。

6. 长期保存数字内容

将作品或胶片数字化并不能确保其长期保存。所有的数字资源——包括数字化的作品和原生数字资源——必须被恰当的维护以确保其可用性。因此，缺乏合适的长期保存策略的数字化只是在浪费投资而已。

数字资源长期保存是信息社会的一个关键问题，在信息社会中，信息的供应呈指数型增长，动态性内容越来越多。现状是对数字资源长期保存方面的经验还不足，法律框架正在演变之中，资源稀缺，在长期保存方面所做的努力其结果还不确定。这一问题值得政府和相关机构给予关注并迅速解决。其影响远远超过图书馆和档案机构的范围，引起数字资源生产者 and 有兴趣维护其可用性的所有机构的关注。

导致数字内容损失的原因众多，首要的原因是软件的更新换代导致文档不可用。1985 年，为纪念“Domesday Book”出版 900 周年，为其制作了一个新的多媒体版本。2002 年，随着阅读这种介质的计算机越来越少，它似乎已无法读取。为了保存这一资源，开发了一个能够利用仿真技术存取这种介质的系统。有趣的是，尽管从 1986 年开始，在存取数字数据方面就开始出现一些困难，有着 900 多年历史的原始的“Domesday Book”仍旧被利用。

计算机程序的迅速过时和更替是另一个原因。除非数据被移植到当前的程序，或者认真保存了原代码，否则信息的检索即使可能，也可能变得十分昂贵。对于那些源代码不公开的封闭的数据尤其如此。数字存储设备（如 CD-ROMs）有限的生命周期是数字内容损失的另一个重要原因。

图书馆和档案馆已经开始在有限规模内处理数字时代的长期保存问题。跨地域协作也已经开始，但是从整体上讲，欧洲的行动还处于分散的状态。总体上，单个成员国对数字内容的长期保存还没有清晰的政策。即使是那些有长期保存计划的国家，其重点也是保存处于损毁风险中的模拟资源，而不是冒险进入数字资源长期保存领域。

已取得的大多数进步是在法定呈缴领域，一些成员国对数字资源引入法定呈缴义务，以确保资源被相关机构收集、保存。然而，当前不同国家呈缴内容的覆盖范围差异很大，例如，有些国家包括动态联机资源，有些则不包括。

数字资源长期保存所面临的基本挑战与数字化内容面临的基本挑战相似：

经济挑战：数字资源长期保存的实际成本还不清晰，这依赖于诸如移植的数量等因素。然而，由于经济能力有限，显然需要有选择的保存资源。

组织挑战：选择是必须的，但是由谁来选择？谁来承担长期保存的责任？保存哪些内容？在这些基本问题还没有回答清楚时，对使用何种方法以及如何实施可能还存在很大的分歧。在补充和交换实践经验的过程中可以获得欧洲附加值（European added-value）。数字信息长期保存也需要新的工作方式，包括改善员工的技能，在公共或私营组织之间进行更多的协作。

技术挑战：迄今为止，数字内容长期保存方面已经有了一些研究。一个主要的挑战是改善其成本效益比，使之更容易承受。进一步的理解，如何保存大量迅速变化的分布式信息是需要解决的另一个基本领域。相关主题（如自动分析或索引工具）的进步，将确保信息长期可获得或可重用。

法律挑战：因为数字内容的长期保存依靠复制和移植，因此必须遵守知识产权法。其它挑战与数字资源的法定呈缴相关：各成员国法律措施的制定、实施速度和覆盖范围之间存在着差异，这可能导致内容生产者在进行跨国活动时必须遵从不同的规则。在法定呈缴和知识产权的交叉点，阻止复制的技术保护措施以及限制对数字资源存取的数字权利管理系统的引入引发了一系列新问题。如果信息生产者不能进行不受保护的复制，那么法定呈缴框架就没有达到其目的。

7. 欧洲的反应

组织并资助文化资源的数字化及数字内容的长期保存是成员国的主要责任。然而，在某些领域，可获得较高的欧洲附加值，在这些领域，在几年前就已经开始了欧洲层面的工作，这些工作将进一步加强，另外还会开展一些新的活动。

7.1 进行中的项目

1. 欧洲层面的合作

欧盟启动 eEurope 行动计划，表明已承认欧洲文化遗产数字化的重要性。这引发了 2001 年的 Lund Principles 和相关的 Lund 行动计划，以及有关数字化的国家代表小组(National Representatives Group)的创建。至于数字内容的长期保存，2002 年的一个理事会决议强调了这一问题的重要性，呼吁成员国解决这一问题。

对于视听资源，与之密切相关的是议会和委员会有关电影遗产的建议草案及相关工业活动间的竞争，这呼吁成员国允许为长期保存目的复制这类资源。

2. 联合资助

欧盟研究基金已经资助了一些项目，旨在通过新技术更有效的获取欧洲文化遗产。在与欧洲图书馆项目（TEL）交换编目记录的过程中，已经发展起支持欧洲国家图书馆之间协作的项目，TEL 现已启动，提供访问欧洲国家图书馆馆藏的单一的门户。

研究计划中有几个项目开始处理数字内容的长期保存问题，还有一些项目专注于解决数字化问题。例如，在 PRESTOSPACE 项目(2004-2007, 900 万欧元的联合资助)中，数字化被用作保存模拟资源。

PRESTOSPACE 正在开发一个工具集，能够对视听资料档案进行数字化。这些工具可以降低转换和长期保存模拟视听资料的成本，并提高数字化质量。

7.2 新项目或改进项目

1. 与资助者的策略讨论

将以报告的附录 2 中所列的问题为基础，对有关数字化、内容的联机可获得性和数字内容的长期保存方面面临的组织、法律和技术挑战召开一个联机咨询会。咨询结果将作为数字化和数字内容长期保存建议的重要依据，同时也将应用于其它过程中，如欧盟 RTD 项目的

实施, 版权框架的修订等。一个数字图书馆高级专家组将就如何在欧洲层面上最有效的解决已知挑战向委员会提出建议。

委员会将加大努力, 推动欧洲在数字化和数字内容长期保存方面的协作, 旨在为成员国的数字化进程提供新的驱动力, 避免重复, 鼓励吸收有益的实践经验。在这方面, Lund 行动计划的更新以及在英国的任期内确定进度会发挥重要作用, 在 Lund 背景下使用定量指标将有助于测度整个欧洲的进步。

对于国家图书馆或保存本图书馆的工作将给予特别的关注, 因为它们是推动欧洲数字图书馆发展的主要因素。

2. 联合资助

将动员Research计划和eContentplus计划在欧盟层面上就数字化、数字内容长期保存和文化资源的可获得性问题采取行动。区域性基金已联合资助某些成员国的数字化项目, 这有助于数字化工作的深化。

在越来越多的案例中, 文化遗产的应用数字化已经成为“文化2000”计划下联合资助的合作项目的一个目标, 在委员会提出的“文化2007”计划中, 联合资助将可用于以改善文化作品跨国流通为目标的项目, 包括通过数字化或联机获取的方式实现。

3. 研究计划 (Research)

在Research计划中, 对于文化资源领域, 委员会将在正在进行的通用的与内容相关的研究(搜索引擎、语言技术等等)之后开展特定的项目:

信息社会技术计划于2005年9月21日结束的新一轮项目招标中, 已经拨款3600万欧元用于新技术的研究, 其中2200万欧元用于搜索和检索文化资源, 1400万欧元用于数字内容的长期保存。

在第七框架计划中, 委员会努力争取扩大其在数字化、数字文化内容的获取和数字内容长期保存领域研究的贡献。建议通过竞争中心网络刺激在这一领域的技术进步, 这可能成为欧洲数字化和长期保存的里程碑。欧盟将对项目提供联合资助促进这些中心的发展。

通过推荐选择的中心, 将主要参与为实现高质量的数字化和长期保存过程而开展的技术和法律方面的各种竞争。它们将整合并依靠技术公司、大学、文化机构和其它相关组织现有的know-how, 它们将:

- 研究并支持先进技术解决方案, 通过处理大量资源的实践来测试、检测这种方案;
- 推动这些方案对文化机构的适用性及其应用;
- 在用户机构中传递知识并支持技术开发。

4. 数字内容附加计划 (eContentplus)

eContentplus计划在2005~2008年期间将投资6000万欧元用于改善欧洲文化和科学内容的易获得性和使用方面的项目。其中, 实现国家数字资源和服务之间的互操作(例如, 通过公共标准)及推动资源在多语言环境下的获取和使用将是核心目标。

7.3 总结:

与资助者的策略讨论

- 数字化和数字内容长期保存的联机咨询 (Commission2005)
- 高级专家组的建立 (Commission2005)
- 推动提供数字化和数字内容长期保存方面的建议 (Commission2006)
- 相关项目的引入, 如版权规则的回顾(2006)和FP7(7th Research Framework Programme)的实施(2007)
- 更新Lund行动计划(包括定量指标)和理事会关于MS在数字化和数字内容长期保存方面的承诺 (MS/Commission 2005)
- 促进国家图书馆馆藏作品的使用。

联合资助

- 资助额达3600万欧元；
- 在FP7中增加数字化、数字内容长期保存和文化内容的获取方面的研究（2007）；
- FP7中数字化和数字内容长期保存的竞争中心（2007）；
- 对eContentplus资助6000万欧元（2005-2008）。

8. 结论

信息技术有潜力实现欧洲的科学文化遗产对当前和未来的用户可见并可用。这需要在成员国层面上和在联盟层面上的共同努力。然而，当前各成员国在数字化和数字内容长期保存方面的努力是分散的。对挑战的成功反应能够加速数字化，提高信息的可获得性并确保数字内容的长期保存。委员会希望欧洲议会和理事会批准提交的方案，并邀请成员国共同努力，在欧洲实现数字图书馆。

编译自：Communication from the commission to the European parliament, the council, the european economic and social committee and the committee of the regions. i2010: digital libraries. [2006-2-20].

http://europa.eu.int/information_society/activities/digital_libraries/doc/communication/en_comm_digital_libraries.pdf

（李丹）