

## 欧盟数字化长期保存研究态势分析<sup>\*</sup>

郭红梅 张智雄

**摘 要** 计算机技术和互联网的发展产生了大规模的数字化对象,如何长期保存这些大规模异质数字化资源,受到了欧盟国家的广泛关注。欧盟委员会每年投入大量经费资助国家级大型存储机构进行数字化长期保存研究。本文以欧盟委员会在第六和第七框架中资助的22项数字化长期保存项目为代表,从资源、研究重心、智能工具和方法等八个方面进行分析,揭示目前欧盟数字化长期保存研究的总体态势:资源类型增多且结构日趋复杂,研究重心从基础概念转向对智能工具方法以及语义技术的研究,从OAIS参考模型转向对工作流以及资源生命周期管理模型和框架的研究,测试平台涉及学科领域更为广泛,而且很多机构加大了对资源的审计研究。参考文献13。

**关键词** 数字化长期保存 欧盟委员会 第六框架 第七框架

**分类号** G203

## The Analysis of Research on Digital Preservation in the European Union

Guo Hongmei & Zhang Zhixiong

**ABSTRACT** The digital preservation of resources has gained much attention in the European Union. The European Commission invests much money to sponsor digital preservation projects. This paper analyses the 22 projects co-funded by the European Commission as part of the Sixth and Seventh Framework Programmes for Research and Technological Development from eight aspects, such as resources, intelligence tools and approaches and so on. Through the analysis, this paper sheds light on the overall trend of research on the digital preservation in the EU. The types of resources are increasing and the structures are more complex. The research focus is shifted from basic concept to intelligent tools, methods and semantic technology. The research on the concept model and system framework has shifted from OAIS reference model to workflow and life cycle management model and framework. The test platform covers more disciplines and many organizations pay more attention to research of resource audit. 13 refs.

**KEY WORDS** Digital preservation. European Commission. FP6. FP7.

### 1 前言

计算机以及网络技术的发展不仅促进了新媒介的产生,同时还促进了数字化资源的数量激增,如何对这些大规模的数字化资源进行表示、存储和

管理,以便未来使用,一直受到图书馆、档案馆等存储机构的关注。随着数字化长期保存研究的深入,越来越多不同性质的机构意识到资源长期保存的重要性;除了传统的机构外,不少公司企业也参与进来;而且数字化资源类型日趋多样,结构日趋复杂,如何有效保存这些大规模不同性质的数字化资

<sup>\*</sup> 本文系国家社科基金后期资助项目“数字资源长期保存的技术研究与实践”(项目编号:09FTQ005)的研究成果之一。  
通讯作者:郭红梅,Email: guohm@mail.las.ac.cn

源成为全世界共同的焦点。欧盟关于数字化长期保存的研究起步较早,资助了很多与其相关的项目,进行了大量深入的研究,是数字化长期保存领域中的典范。本文拟对欧盟委员会在第六和第七框架(FP6和FP7)资助的22项数字化长期保存项目<sup>①</sup>从数字化资源、方法工具以及系统模型框架等八个方面进行调研,以这些项目为代表总结欧盟数字化长期保存研究的继承演变以及总体现状,为其他机构进行相关研究提供参考。

## 2 欧盟数字化长期保存的资源类型和结构

在欧盟数字化长期保存过程中,资源类型变得越来越复杂,早期项目集中对单一文件格式的处理,比如办公文件或视听资源的保存。第五和第六框架中资助的项目比较关注办公文件的存储(以文本文档和图像为主),其中提出的一些基本概念和方法可适用于很多场景。PLANETS<sup>②</sup>、PrestoPRIME<sup>③</sup>、PARSE、Insight<sup>④</sup>、CASPAR<sup>⑤</sup>和SHAMAN<sup>⑥</sup>项目对视听资源和科学数据的保存进行了大量研究,其中SHAMAN是一个明确提出利用产品生命周期管理办法(PLM)对工作流保存的项目<sup>①</sup>。

第七框架中资助的数字化长期保存项目主要针对数据结构更为复杂的资源内容,如LiWA<sup>⑦</sup>、SCAPE<sup>⑧</sup>、BlogForever<sup>⑨</sup>、APARSEN<sup>⑩</sup>和ARCOMEM(ARChive COMMunities MEMories)<sup>⑪</sup>项目主要处理与科学社会网络相关的资源,解决了在常规格式文件存储中遇到的问题。其中LiWA是世界上第一个网络归档系统,作为网络归档的开端,其结果对该领域的实践者来说具有非常重要的价值,是网络归档技术最好的实践者,有助于网络归档技术在欧洲范围内的协调,对互联网的内容特征(如规模大小、分布、来源、结构等)具有深入揭示能力;SCAPE项目为复杂数字化对象大规模、异质数据集的存储提供半自动化的工作流<sup>⑫</sup>;BlogForever项目主要进行网络日志内容的保存;ENSURE处理来自卫生保健和金融方面的科学数据<sup>⑬</sup>;DURAARK项目主要对3D建筑学数据进行保存<sup>⑭</sup>;DIACHRON项目对关联开放数据进行存储<sup>⑮</sup>;FORGETIT项目对多媒体报道的个人事件资源以及机构信息进行保存<sup>⑯</sup>。

目前长期保存项目主要集中对交互对象、嵌入对象、本体以及短暂数据的存储。TIMBUS项目主要对进程和应用数据进行保存<sup>⑰</sup>;PERICLES项目通过资源内容生命周期的语义演化来提高信息的重用<sup>⑱</sup>;Wf4Ever项目关注对科学工作流的保存,为

① [http://cordis.europa.eu/fp7/ict/creativity/digicult-preservation\\_en.html](http://cordis.europa.eu/fp7/ict/creativity/digicult-preservation_en.html)

② <http://www.planets-project.eu>

③ <http://www.prestoprime.org>

④ <http://www.parse-insight.eu>

⑤ <http://www.casparpreserves.eu>

⑥ <http://shaman-ip.eu>

⑦ <http://liwa-project.eu>

⑧ <http://www.scape-project.eu>

⑨ <http://blogforever.eu>

⑩ <http://www.alliancepermanentaccess.org>

⑪ <http://www.arcomem.eu>

⑫ <http://ensure-fp7-plone.fe.up.pt/site>

⑬ <http://duraark.eu>

⑭ <http://diachron-fp7.eu>

⑮ <http://www.forgetit-project.eu>

⑯ <http://timbusproject.net>

⑰ <http://www.pericles-project.eu>

科学工作流的重用和有效检索以及存储提供技术框架<sup>①</sup>; DAVID 项目<sup>②</sup>主要对数字化视听资料的损害进行预防和修复,由此可看出长期保存更关注对物理文件的数据、结构和行为进行分析,改变了以往仅以数据为中心的现象<sup>[3]</sup>。

### 3 欧盟数字化长期保存研究重心的变化

早期数字化长期保存项目主要由数字图书馆和档案馆发起,集中于对基本概念、系统架构和存储方法进行定义探讨,常包含资源的选择和评估、元数据的定义、唯一标识以及特性工具的描述等,该阶段代表性项目有 ERPANET<sup>③</sup> 和 DELOS<sup>④</sup>。其中 ERPANET 是欧盟第六框架中资助的项目,它的主要目标是提高数字化长期保存的意识,对文化遗产和科学对象进行评价和存储以实现资源共享; DELOS 是与数字化图书馆相关的卓越网络,它将欧洲数字化图书馆研究集成在一起,通过可信的中间框架提供多种个性化服务,便于各机构之间进行交流沟通。

随后不少项目尝试将一些可用的、已存在的方法、工具和模型等嵌入到长期保存系统架构中,该阶段主要关注对数字化长期保存工具的扩展应用研究。存储模型与系统架构的集成促进了工作流中组件与其他系统的集成和重用,如 PLANETS 交互框架、CASPAR 的集成框架以及 SHAMAN 项目中利用网络技术的集成保存框架<sup>[4]</sup>。另外一些数字化长期保存项目关注新方法和工具及其在具体场景下的应用研究。如 PROTAGE 项目关注代理环境的运用<sup>⑤</sup>; ARCOMEM 项目关注社会网络的使用; Li-WA 项目关注网络资源的归档; KEEP 项目主要利用仿真技术进行归档<sup>⑥</sup>; DIACHRON 项目基于关

联数据和现代数据库系统技术对网络数据进行保存; ForgetIT 项目将管理遗忘模型、协同存储和场景记忆三个概念结合起来,以简化个人和机构对数字化资源的存储。

随着视听资源的激增,在技术陈旧、媒介退化、关键人物、进程和系统失效的情况下,如何保证视听资源的安全性是 DAVID 项目主要解决的问题,它不只是对视听资料已产生的损害进行修复,而且从未来的角度对可能产生的损害进行探测和预防以提高长期保存的质量<sup>[5]</sup>。

近期启动的 4C 项目<sup>⑦</sup>对长期保存过程中使用的工具和方法关注较少,而是从投资收益的角度对各数字化长期保存项目进行分析,帮助欧盟长期保存机构进行数字化资源管理和存储的成本效益分析,使其投入获得最大收益。因此在保存过程中应该包括风险、价值以及质量等相关概念的研究<sup>[6]</sup>。

### 4 欧盟数字化长期保存工具和方法相关研究

#### 4.1 长期保存工具和方法的可扩展性研究

已存在的数字化长期保存工具大多是为实现系统某种具体功能,并不适用于对大规模数据的操作。但在实际工作中我们常需要对大批量的资源内容进行处理,如网络资源的归档以及大型机构(如图书馆和档案馆)资源的仓储,这就需要我们开发一些可扩展的、可处理大量资源的工具和方法。

在 SCAPE 项目中,为实现可扩展性,除封装组件外还利用虚拟机映射来封装复杂的软件环境, IaaS(Infrastructure-as-a-Service) 服务提供了云托管模型,并利用云存储技术对存储结构松散的大量

① <http://www.wf4ever-project.org>

② <http://david-preservation.eu>

③ <http://www.erpanet.org>

④ <http://www.delos.info/>

⑤ <http://www.ra.ee/protage>

⑥ <http://www.keep-project.eu/>

⑦ <http://4cproject.net>

数据进行扩展存储和处理<sup>[2]</sup>; TIMBUS 项目利用云存储技术以及可视化保存技术进行扩展操作; 长期以来网络资源快照复制一直是“freezing”存储, LiWA 项目把这种单纯的快照复制存储转化为“living”的网络存储, 建立了大规模数据分析测试平台, 开发基于互联网规模的可扩展方法; FORGETIT 项目中通过智能保存过程和双向平滑转换来实现信息使用和保存管理之间的协同存储; DURARK 项目从不同语义层面进行数据的查找和使用; PERICLES 项目开发 WP3、WP4 和 WP5 三个工具包, 分别从概念模型、工具方法、过程规则进行研究, 以上项目通过使用和开发可扩展工具来实现机构长期存储的可扩展服务, 从而提高数字化长期保存的水平; ENSURE 研究生命周期工具在长期保存中的适用性, 提出了数据保存生命周期管理机制, 能对初始条件的变化作出动态响应, 比如规则、格式或者经济条件等; APARSEN 以 pay-per-use 模式提供服务, 运用云存储技术定义了 PDS 云模型, 是对 CASPAR 项目中的存储框架方法 PDS 进行扩展得到的, 并可与 ENSURE 项目中的案例进行集成<sup>[7]</sup>。

#### 4.2 长期保存智能工具和方法研究

早期数字化长期保存工作需要具有相关专业技术人员来实施, 而且存储过程中使用的方法也需要人为进行交互, 并不能实现智能存储和自动化处理。目前趋势是开发一些智能化工具和方法来辅助用户并支持决策过程, 开发和利用知识库以及创新方法和工具来支持大量异质数据在复杂场景中的使用。

Plato 是一种开源的、支持网络决策的工具, 其基本概念是在 DELOS 项目中提出的, 在实际应用中经过多次修正和凝练, 最终在 PLANETS 项目中提出了充分论证的、支持透明决策的 Plato 保存计划方法。使用 Plato 工具可以呈现一个完整的长期保存计划, 通过执行相应程序可实现知识库的

集成、自动化测度以及原始对象迁移前后的比较, 同时还可对保存计划进行监测<sup>①</sup>。SCAPE 项目以保存计划工具 Plato 为基础, 集中于对大规模资源内容的处理, 提供自动化的决策支持过程。

GRATE 方法是在 PLANETS 项目中提出的, 它不仅通过网络提供了专用软件的远程接口, 而且允许用户上传原始数字化资源, 这些资源可在类似的服务中(如长期保存项目 DROID<sup>②</sup> 和 PRONOM<sup>③</sup>) 进行自动化识别, 也可对已存储对象的元数据进行展示; GRATE 可将这些资源放到模拟环境下进行使用, 在必要的时候还可将它们复制到用户系统中; 此外 GRATE 方法可通过按键对模拟环境进行控制, 减少本地仿真器的构建, 可对模拟操作系统内容运行的软件进行自动化管理, 如连接网络共享或关闭模拟机器等。

在数字化长期保存方法的最新发展和自动化系统基础上, PROTAGE 项目提出了可解决大批量异质性资源存储的先进软件代理技术, 通过复杂和可扩展的软件代理实现资源的自动化长期保存。这些软件可与其他数字化长期保存系统集成交互, 支持数字化资源的提交、保存系统的监测以及知识库之间的转移等操作。PROTAGE 方法是一个灵活的、分散的软件代理商系统, 可独立运行但需要与其他资源合作来制定决策支持过程, 该方法在保存、检索和共享数字化对象过程中需要用户参与<sup>[8]</sup>。

PERICLES 项目的核心研究是 WP3、WP4 和 WP5 三个工具包的开发, WP3 基于关联数据原理定义了概念框架统一模型, 可对动态存储环境进行表示, 并可与长期保存系统中的描述语言和工具进行集成; WP4 用于研发识别、抽取、分析和封装数字化对象以及相关存储环境的分析工具和方法, 可对表示信息、语境信息、语义内容以及元数据进行描述; WP5 对已存在的生命周期存储模型进行扩展, 以适应存储环境的不断扩展演化, 主要关注技术的

① <http://www.ifs.tuwien.ac.at/dp/plato/intro.html>

② <http://digital-preservation.github.io/droid/>

③ <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

变化 制定一些具体的过程和规则 此外它也开发一些对存储环境进行管理的过程和工具 尤其是评估过程<sup>[9]</sup>。

## 5 概念模型和系统框架的发展

### 5.1 基于 OAIS 参考模型的发展

OAIS 参考模型以及术语是数字化长期保存领域中通用的基本概念和模型,但该参考模型并不能详述所有长期保存系统,在后来的数字化长期保存项目中逐渐得到完善和细化<sup>[10]</sup>。CASPAR 概念模型是 OAIS 框架模型最好的实践者,建立在 OAIS 模型之上并对其进行了扩展和修正,提出了一个可广泛使用的框架模型,可对保存元素进行识别标识,也包括对表示信息和保存描述信息进行详细描述的信息模型。CASPAR 把 OAIS 兼容模型中的概念进行形象化表示,提供了有助于人们进行信息和知识存储的指导方针和方法等,它通过概念之间的依赖性,定义易于理解接受的概念,除了提出正式的概念模型外, CASPAR 还发展了很多工具并将其应用在实际数据中。SHAMAN 项目最先从系统设计角度来解决数字化保存问题,它识别了系统的主要特性和需求,促使利用企业架构框架来解决数字化保存,建立一个开放分布的资源管理基础框架,实现基于网格的资源整合,利用 GRID 计算,联合数字图书馆和持久性存档领域中的虚拟和分布技术来改善和扩展 OAIS 模型。

### 5.2 有关 workflow 保存的概念模型和框架

SCAPE、ENSURE 和 W4Ever 项目主要解决 workflow 数字化对象的保存问题。SCAPE 系统提供了数字对象的管理和计算平台,可与其他子系统如计划和监测子系统等交互,其中计算平台可将数据存储与网络处理集合在一起,可对数据密集型资源进行计算,尤其支持 SCAPE 工具和工作流的部署、识别和并行操作,并可不同的数据源和数据接收器集成在一起。ENSURE 项目中 pay-as-you-go 框架是基于云存储技术进行构建的,不仅可降低数据丢失的概率,而且能支持系统的自动化存储。

W4Ever 项目为科学工作流的长期保存提供了新工具和方法,可解决数据密集型科学实验资源的长期保存,框架中制定了一些在科学工作流存储中用到的功能接口。在框架定义阶段确定了具体功能,并将其分为四类功能实体,在实体定义过程中提供更精细的属性和概念作为参考,以使该框架与标准与 OAIS 参考模型相一致,而且 W4Ever 工具包可实现数据生命周期内对数据进行存储、管理和分析。TIMBUS 项目框架主要为企业智能风险管理系统以及具有所有依赖性和语境业务流程的存储提供参考,它克服了以往数字化长期保存系统中存在的缺点,测试用例涵盖了商业和科学领域中复杂的业务流程。

### 5.3 基于资源生命周期管理模型的研究

PLANETS 项目提出了获取需求的概念模型和长期保存框架,它将风险与行为联系在一起,在参与者特定需要下进行迁移和表示,通过实施保存计划、个性化和跟踪行动,如迁移和仿真来保持对数字化对象的使用。KEEP 项目将进一步发展仿真服务以及提供兼容接口来对 PLANETS 项目中的框架进行拓展。PROTAGE 项目中利用多代理系统框架,在系统设计过程中使用了不同方法,该框架包括用于长期数字化保存和使用的网络服务和软件代理工具,代理商协同保存过程可通过自动化查找、选择和执行网络服务来获取信息,支持决策并执行保存工作,网络服务为代理和用户特定类型服务,如获取、服务知识库、病毒检查、迁移以及元数据抽取等。PERICLES 项目主要是让一些复杂、异质、高度相关、易于变化的数字化资源在整个生命周期内保持可用性和可信性,提出了基于关联数据范式模型来描述资源,包括内容、元数据、过程、用户以及规则等,以便在环境变迁的情况下对资源的依赖性和一致性进行管理,利用扩展保存和生命周期模型来解决数字化生态系统的演变,并开发相应的工具;开发识别和捕获数字化对象和保存环境相关信息,如语义、用户以及交互背景信息等<sup>[9]</sup>。FORGETIT 项目中提出了 Preserve-or-Forget 框架,主要包括托管忘记(Managed Forgetting)、内容管理(Content Management)和

保存计划(Preservation Planning)三个模块。托管忘记模块主要进行信息评价、浓缩以及协同存储等。保存计划模块主要进行信息的收集、归档保存、数据管理以及使用等。内容管理模块主要负责信息的使用和编辑。在这三个模块的构建过程中集中了来自各方面的专家。通过这三个模块可实现对资源的智能存储和管理。

## 6 数字化长期保存语义技术的发展

在欧盟数字化长期保存项目中,语义技术主要应用在 CASPAR 和 SHAMAN 两个项目中。由于数字化长期保存是一个动态的存储过程,软硬件载体、软硬件框架以及数字化对象都在发生动态演变, CASPAR 项目中利用语义网络管理平台(SWKM)来实现对数字化网络资源的动态存储<sup>[5]</sup>。SHAMAN 实现了基于网络资源整合系统的构建,利用 GRID 计算和长期保存领域存在的虚拟和分布技术来改善和扩展 OAIS 模型,使用本体来表示数字化对象属性之间的关系。DURAARK 项目除了利用目前元数据方案还支持从不同语义层面对资源进行检索和使用。PERICLES 主要关注动态、异质数字化资源长期使用,它不仅考虑到这些资源以及相互依赖性的动态变化,还考虑到了在资源整个生命周期中由于技术的陈旧或术语演化导致的语义漂移。

## 7 数字化长期保存中元数据研究

在数字化长期保存领域,已经构建了不少可广泛接受的元数据标准,如 MAB、MARC、Dublin Core、PREMIS 等,其中 PREMIS 是长期保存领域已构建的核心元数据标准<sup>[11]</sup>。但长期保存迫切需要可长期使用的、可扩展的元数据标准,不少机构和学者积极尝试对元数据标准进行扩展,以提出适用于不同场景、目标、对象和方法的元数据。CAPSPAR 主要研究对对象信息和表示信息描述的标准,提出了元数据自动化管理工具 PreScan,主要用于元数据收集阶

段,能自动将抽取的嵌入式元数据与人工提供的元数据和依赖管理服务的元数据集成在一起,同时也能将一些格式标识符和元数据抽取工具集成在一起,其工作原理与网页搜索爬行方法类似。

PLANTEETS 项目主要研究数据字典的高级特性,创建用于元数据概念的数据字典,这种潜在的概念模型可支持动态的存储过程,而不仅仅是特性和事件的静态记录。KEEP、LiWA 和 ARCOMEM 项目主要对元数据的具体应用进行研究,如网络归档、社会网络以及视听资源等的存储。PrestoPRIME 根据 OAIS 模型进行语义交互,利用具体的参考标准来分析资源内容,PREMIS 为元数据的存储提供技术和方法支持,在各种事件可控词表下利用 PREMIS 事件和媒介表示,来实现内容添加和事件修改。网络完全改变了我们收集和利用数据的方式,开放数据指数递增,目前我们使用的大量数据如人口统计资料、临床统计以及很多科学测量值不可复制<sup>[12]</sup>。DIACHRON 项目欲自动收集这些元数据、信息来源以及各种背景信息,以便这些信息加工定义后仍可理解和利用。

## 8 数字化长期保存实验测试平台研究

SCAPE 项目中的具体结果将在三个不同领域的大规模测试平台上进行检测:网络内容测试平台主要验证异质数据的收集和传递环境的快速变化,存储平台主要在法律和政策规定范围内执行一个机构背景下的资源长期保存,研究数据集测试平台关注科学数据长期有效和可用的需求。DIAHRON 项目成果在开放政府数据、大型企业数据内部网以及生命科学中的科学数据三个大规模用例中进行分析。PERICLES 成果在两个领域的数字化长期保存中进行检测,一是数字化艺术品,二是来自欧洲太空总署与国际空间站的实验科学数据。ENSURE 长期保存方案在卫生保健、金融和临床试验三个领域进行实验和验证。PLANETS 面向服务的架构方法可

2014年3月 March 2014

以在不同性质的平台和已定义的受控环境下运行, 这些工具可通过标准的词表在网络服务包中嵌入, 因此用户可通过 Testbed 网络接口进行使用。通过该平台用户可设计和执行各种实验, 比如迁移、仿真以及可执行的保存计划实验。为验证 LiWA 方法的有效性, 项目中构造了两个演示程序分别针对社会网络和视听内容进行归档。

## 9 数字化长期保存真实性、可信性以及法律审计研究

数字化对象的真实性和可信性是所有数字化长期保存系统的共同要求。数字化对象的真实性就是确保数字化对象的完整, 保证信息内容不可更改。在技术方面欧盟数字化长期保存项目提出了适用于不同对象的方法, 但是在语义层面如何确保数字化对象的真实性存储的研究并不太多。LiWA 项目中提出了自动处理语义演化的方法, 使保存内容具有长期可解释性, 克服了语义存储在语言处理面临的问题, 解决了语义保存在语言方面的挑战<sup>[13]</sup>。

可信性是长期归档中存在的另一个重要问题。DPE<sup>①</sup> 是第一个采取自我评价方法 (DRAMB-ORA) 来解决资源可信性的项目, 它支持机构在对隐式风险进行识别和管理评价之前对活动目标以及资源进行全面的自我评估。APARSEN 项目提出了通用的第三方组织资源长期保存可信性检定方法。但在法律问题上, 很多项目遗漏了对长期保存资源进行验检, 最早关注该问题的是 DPE 项目, 随后 KEEP 项目针对仿真法律问题进行了研究; CASPAR 项目在对 DRM 进行深入研究的基础

上提出了适用于多种法律系统的工具; TIMBUS 项目解决保存业务流程中的法律问题。

## 10 结论

早期欧盟委员会资助的数字化长期保存项目主要用于提高社会长期保存意识, 随着大量研究人员和机构的参与和研究的逐步深入, 长期保存研究重心发生转移, 目前主要针对结构复杂工作流、交互对象、嵌入对象、本体以及短暂数据进行存储; 开发一些可扩展的、智能化的工具和方法; 开发可与其他系统进行集成的模型和系统框架, 提高资源共享; 注重对语义技术的开发, 同时考虑资源之间的依赖关系以及在整个生命周期中由于技术的陈旧或术语演化导致的语义漂移; 不断加强元数据构建的研究; 为保证系统功能的准确完善而构建针对不同领域资源的大规模测试平台; 同时还从法律审计等方面确保数字化长期保存的真实性、可信性以及未来的可用性; 近期启动的 4C 项目从投入管理角度出发, 研究如何使长期保存达到成本效益最优化。

欧盟资助的数字化长期保存项目顺利完成, 取得了不菲的成果, 建立了大型的专业团队, 对目前存在的问题提出了有效的解决方法, 推动了欧盟区域内长期保存的进步, 引领世界长期保存领域的发展。在完成这些数字化长期保存项目过程中, 各参与机构也积累了大量宝贵的经验, 但随着计算机技术和互联网的发展, 我们仍需努力来应对未来大规模异质数据源以及自动化处理技术等带来的新挑战。

## 参考文献

- [1] Strodl S, Petrov P, Rauber A. Research on digital preservation within projects co-funded by the European Union in the ICT programme[R]. Vienna University of Technology, May 2011.
- [2] Chailloux J. ERCIM News[M]. France: ERCIM EEIG. 2010: 3-11.
- [3] Inmark ESTUDIOS Y ESTRATEGIAS S A. Socio-economic drivers and impact of longer term digital preservation[R].

① <http://www.digitalpreservationeurope.eu>



- D. 5 Final Report ,2009: 11 –60.
- [ 4 ] Agosti M. Experiences and perspectives in management for digital preservation of cultural heritage resources [J]. Communications in Computer and Information Science ,2013 ,354: 1 –3.
- [ 5 ] The future of the past-shaping new visions for EU-research in digital preservation [C]// The Information Society and Media Directorate General of the European Commission Workshop Luxembourg 2011 ,5: 21 –46.
- [ 6 ] Ross S. Digital preservation , archival science and methodological foundations for digital libraries [J]. New Review of Information Networking 2012 ,1( 17) : 43 –68.
- [ 7 ] Chanod JP. Automation in digital preservation [C]// Dagstuhl Seminar Proceedings ,2010: 18 –23
- [ 8 ] Neumann M , Miri H , Thomson J , et al. Towards a decision support architecture for digital preservation of business processes [C]// iPRESS 2012: 74 –81.
- [ 9 ] Barateiro J , Draws D , Neuman M A , et al. Digital preservation challenges on software life cycle [C]// Software Maintenance and Reengineering ( CSMR) . 16th European Conference 2012: 487 –490.
- [10] Lavoie B F. The open archival information system reference model: Introductory guide [R]. DPC Technology Watch Series Report. OCLC Online Computer Library Center , Inc. 2004: 28 –36.
- [11] Dappert A , Enders M. Digital preservation metadata standards [J]. Information Standards Quarterly 2010 ,22( 2) : 156 –166.
- [12] Alemneh D G , Hastings S K. Exploration of adoption of preservation metadata in cultural heritage institutions: Case of premis [J]. American Society for Information Science 2010 ,47: 501 –508.
- [13] Buneman P. Database preservation [R]. digital preservation Europe. 2009 ,5: 18 –35

郭红梅 中科院文献情报中心博士研究生。

通讯地址: 北京市中关村北四环西路 33 号中科院文献情报中心信息系统部。邮编: 100190。

张智雄 中科院文献情报中心研究馆员。通讯地址同上。

( 收稿日期: 2013 –06 –18; 修回日期: 2013 –09 –01)