

电子文件 长期保存：

理论与实践

Long-term Preservation
of Electronic Records:
Theory and Practice

肖秋会 著



社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

电子文件长期保存：

理论与实践

Long-term Preservation
of
Electronic Records:

Theory
and
Practice

肖秋会 著



社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

图书在版编目(CIP)数据

电子文件长期保存:理论与实践 / 肖秋会著. —北京:社会科学
文献出版社, 2014. 11

ISBN 978 - 7 - 5097 - 6620 - 0

I. ①电… II. ①肖… III. ①电子档案 - 档案管理 - 研究
IV. ①G276

中国版本图书馆 CIP 数据核字 (2014) 第 237110 号



电子文件长期保存:理论与实践

著 者 / 肖秋会

出 版 人 / 谢寿光

项目统筹 / 桂 芳

责任编辑 / 桂 芳

出 版 / 社会科学文献出版社·皮书出版分社(010)59367127

地址:北京市北三环中路甲 29 号院华龙大厦 邮编:100029

网址:www.ssap.com.cn

发 行 / 市场营销中心(010)59367081 59367090

读者服务中心(010)59367028

印 装 / 三河市尚艺印装有限公司

规 格 / 开 本: 787mm × 1092mm 1/16

印 张: 17.5 字 数: 296 千字

版 次 / 2014 年 11 月第 1 版 2014 年 11 月第 1 次印刷

书 号 / ISBN 978 - 7 - 5097 - 6620 - 0

定 价 / 78.00 元

本书如有破损、缺页、装订错误,请与本社读者服务中心联系更换

版权所有 翻印必究

目 录

第一章 电子文件长期保存的相关概念及理论基础	001
第一节 电子文件的概念、特点及类型	001
第二节 电子文件长期保存的概念及认识误区	005
第三节 数字保存的发展历史及现状	008
第四节 电子文件长期保存的理论基础	014
第二章 电子文件长期保存的法律基础	025
第一节 电子文件的证据力及立法保障	025
第二节 数字信息长期保存中的版权问题	040
第三节 电子文件在保存中的版权问题	059
第四节 电子文件在保存中的隐私权问题	071
第三章 电子文件长期保存的标准保障	075
第一节 电子文件长期保存系统标准	075
第二节 文件管理国际标准	096
第三节 电子文件管理与长期保存元数据标准	100
第四节 电子文件长期保存的格式标准	109
第五节 电子文件长期保存相关标准的选择及使用问题	114
第四章 电子文件长期保存的管理策略	117
第一节 国家层面的电子文件管理战略	117
第二节 组织机构电子文件长期保存政策和策略	142
第三节 组织机构电子文件长期保存成熟度模型及评价指标体系	164
第五章 电子文件长期保存的技术策略	169
第一节 电子文件长期保存的层次	169

第二节	电子文件长期保存技术策略及其选择	170
第三节	电子文件长期保存的元数据策略	182
第四节	电子文件长期保存的格式问题及对策研究	190
第五节	电子文件长期保存的存储媒体及其选择	201
第六章	可信数字仓储与电子文件的长期保存	205
第一节	可信数字仓储及其审计和认证研究进展	205
第二节	可信数字仓储的属性特征、责任框架及评价指标体系—— TRAC	208
第三节	欧美可信数字仓储审计和认证标准对我国数字档案馆认证 评估的启示	215
第四节	可信数字仓储与电子文件的摄取和长期保存	219
第七章	电子文件长期保存的合作机制及实现策略	224
第一节	电子文件长期保存合作的必要性和重要性	224
第二节	不同层次和范围的数字保存合作项目及联盟	226
第三节	我国电子文件长期保存的合作机制及实现策略	243
附录一	机构数字保存能力评估指标体系	253
附录二	TRAC 可信数字仓储的评价指标体系	258
参考文献		263
后 记		276

第一章 电子文件长期保存的相关 概念及理论基础

第一节 电子文件的概念、特点及类型

人们对电子文件（electronic records）一词的称呼和使用由机读档案（machine-readable archives）演变而来。20世纪50~60年代，随着计算机技术在军事和科技领域的应用，记录于计算机磁带上的一类新型档案——机读档案由此而产生。根据我们的文献调研，从20世纪60年代直至80年代中后期，欧美的政府机关及企业使用“机读档案”一词较多，“电子文件”的使用并不多见。80年代末90年代初，随着文件管理工作在现代计算机环境中的作用日益突出，“电子文件”一词开始流行，并逐步取代“机读档案”。

电子文件的产生有两种方式，一种是原生性的，直接产生于电子办公环境、电子政务、电子商务或者其他电子化业务活动及网络环境，另一种则是对传统介质档案做数字化处理以后形成的数字化副本。电子文件既具有文件的基本属性，又具有数字信息的特征。因此，从概念隶属关系来看，电子文件可以视为“文件”与“数字信息”这两个概念集合的交集。国际档案理事会对文件和电子文件的定义可为我们理解和全面认识电子文件提供有益的参考。国际档案理事会电子文件委员会《电子文件管理指南》（1997年）和在此基础上编制的《电子文件：档案工作者手册》（2005年）对文件的定义非常宽泛：“文件是由机构或个人在其活动的开始、进行和结束过程中产生或获得，由内容、背景和结构构成并足以证明该活动的信息记录”，该定义对文件这一“信息记录”并没有限制格式或者载体形式，适用于所有类型和载体形式。但是，该定义准确而清晰地指明了文件最基本的属性，即在文件形成主体的活动中产生并具有证明该活动的证据价值。电子文件是文件的种概念，其定义建立在文件的基础之

上,根据《电子文件管理指南》,电子文件是“通过数字计算机进行操作、传递和处理的文件”。文件必须与某一组织或机构的活动相关,该活动或者事务是文件产生的来源,同时,文件又为该活动提供证据,电子文件也不例外。因此,所有的组织都应该收集和保管文件,既为满足组织自身业务活动的需要,也为法律取证的需要,文件产生和保存的主要目的是对组织的运作或对法人及个人活动的问责提供证明。

可以按照两种不同的标准对文件进行分类。一种是根据文件的功能来分类。即根据文件在不同的办公职能活动或事务中所发挥的功能对其进行分类,如:案卷卷宗、法院文件、人事文件、关系数据文件、网页文件等。另一种则是根据文件的格式和表现形式来分类。包括:字处理文件、数据库文件、超文本文件、图像文件、电子表格、电子邮件、语音邮件、视频,等等。

内容、结构和背景是构成文件的基本要素,从这三个构成要素来全面理解和认识电子文件尤为重要。文件的内容是指文件所要表达的事实或者信息,是文件的核心。文件的结构是指文件的记录组织方式,包括文件使用的符号、文件的组织布局、文件的格式、文件的载体等。对于电子文件而言,有物理结构和逻辑结构之分。与传统文件固定而直观的物理结构不同,电子文件的物理结构是可变的且依赖于硬件和软件,电子文件的逻辑结构是指其各个组成部分之间的关系,使得文件可理解。文件的背景包括如下三个要素:第一,文件所包含的背景信息,如文件执行者的签名;第二,文件与同一全宗内其他文件之间的关系;第三,形成文件的活动信息。背景信息使文件本身与产生该文件的行政管理(事务)活动或其他文件发生关联,其目的在于:为完整而充分地理解文件提供必要的信息;为完整而充分地理解与文件相关的活动或事务提供必要的信息;为文件的处置(鉴定、迁移、移交等)提供必要的信息;为有效地管理和长期保存文件提供必要的信息;为有效地检索和利用文件提供信息。背景信息同时也能为文件的真实性、可靠性和完整性提供保障,这对于电子文件十分重要。文件形成机构的业务活动、功能和文件管理系统是文件背景的构成部分之一。电子文件的背景信息可以通过其内部元素如附件、附加信息、链接、数字和参数代码等保存,也可通过外部元素如元数据进行保存。元数据是用于描述文件的背景、内容、结构及管理活动的信息。元数据对于电子文件非常重要,因为它建立了文件与其形成的职能活动和行政管理背景

之间的联系。为了使电子文件可理解和可用,元数据几乎包括了所有类型的信息。元数据可用于不同的目的,如文件的检索、利用、真实性保障、可靠性保障、维护、保存和鉴定,等等。从文件形成和管理机构角度看,元数据可分为两种类型:提供业务活动背景信息的元数据和反映文件捕获及存储的管理元数据。

国际档案理事会关于文件定义、文件特质及文件构成的指南性意见对于我们准确认识电子文件的本质特征和构成要素具有积极的参考作用。此外,美国国家档案与文件署认为,电子文件是以计算机可处理的形式存储的文件。澳大利亚国家档案馆认为,电子文件是由计算机交流与维护的文件。而加拿大 InterPARES 项目组认为,电子文件是以电子形式形成(制作、接收及为今后的活动和参考而保留下来)的文件。

我国国家标准《电子文件归档与管理规范》规定,电子文件是指在数字设备及环境中生成,以数码形式存储于磁带、磁盘、光盘等载体,依赖计算机等数字设备阅读、处理,并可在通信网络上传送的文件。此外,《电子文件管理暂行办法》(中办国办厅字〔2009〕39号)规定,本办法所称电子文件,是指机关、团体、企事业单位和其他组织在处理公务过程中,通过计算机等电子设备形成、办理、传输和存储的文字、图表、图像、音频、视频等不同形式的信息记录。

为了支持组织的业务活动及提供证据,根据 ISO 15489-1 和《电子文件:档案工作者手册》,文件必须具备如下四个特质:

第一,真实性。指文件的内容、背景和结构能长久保持其产生时的原貌。一份真实的文件应符合如下条件:文件与其制作的目的相符;文件的形成和发送与其既定的形成者和发送者相吻合;文件的形成或发送与其既定的时间一致。为了确保文件的真实性,机构应执行并记录文件管理方针和政策,便于控制文件的形成、接收、传递、保管和处置,从而确保文件形成者是经过授权和确认的,同时,能够防止未经授权对文件进行增、删、改、利用和隐藏。

第二,可靠性。文件能够被作为可信的证据,强调文件作为证据的权威性和可信性。文件的内容可信,可以充分准确地反映其所证明的事务、活动或事实,为后续的业务或活动提供依据。

第三,完整性。文件是完整的且未经改动的。应防止文件未经授权而改动。在文件管理方针和程序中应明确:文件形成之后可对文件进行添加

或注释；在何种条件下、授权谁来负责添加或注释；任何授权的注释、添加及删除都应明确注明并可跟踪。

第四，可用性。文件能够被定位、检索、呈现及理解。可用的文件应该能够直接表明文件与产生它的业务活动和事务过程。文件之间的背景联系中应该包含有助于理解文件形成及利用事务的信息，这些信息能够在更为广泛的业务活动背景和功能中确认该文件。文件之间的背景联系记录了活动的顺序而应该被保存。

具备了上述特征的文件将有充实的内容、结构和背景信息为相关的活动和事务提供完整的证明，从而反映组织机构的决策、行动和责任。如果这样的文件以易于存取、可读和可用的方式保存，它们将支持业务需求，并能长久用于问责的目的。

电子文件除了具有文件的四个基本特质之外，还同时具有数字信息的共性：由电子计算机等数字设备生成和处理，信息用二进制数字代码记录和表示。其基本特点包括：数字信息对载体及环境的依赖性、数字信息内容的非人工识读性、数字信息可在不同载体间的转换性、数字信息可在信息网络中的流动性、数字信息的不稳定性，等等。

电子文件类型多样，按照电子文件的格式对电子文件进行分类最为常见，可分为如下几种：

(1) 文本电子文件。通过文字处理技术形成，以 XML、RTF、TXT、PDF 为通用格式。

(2) 图像电子文件。通过扫描仪等设备获得，以 JPEG、TIFF 为通用格式。

(3) 图形电子文件。通过计算机辅助设计或绘图等软件获得。

(4) 音频电子文件。通过数字音频设备获得，以 WAV、MP3 为通用格式。

(5) 视频及多媒体电子文件。用数字视频技术、多媒体技术设备以及超媒体链接技术制作形成，以 MPEG、AVI 为通用格式。

(6) 数据库电子文件。是以数据库形式组织存储的具有文件属性的信息记录。在数据库中有很多条记录，数据库文件可以是其中的某一条记录，也可以是若干相关的记录集合。不同的数据库管理软件形成的数据库电子文件具有不同的格式。

(7) 计算机程序文件。是用计算机语言编写的程序，是一种计算机软

件,包括软件型号、名称、版本号和相关参数手册、说明资料等。计算机程序文件分为通用软件程序文件和专用软件程序文件。对于专用软件产生的电子文件,原则上一般应转换成通用型电子文件。如果难以转换,则必须将该专用软件一并收集。

另外,按照电子文件的功能,还可将电子文件分为主文件、支持性文件、工具性文件;按照电子文件的组成方式,可分为简单文件、复杂文件(如多媒体文件、超文本文件、超媒体文件等);按照电子文件的存储介质,可分为磁带文件、磁盘文件、光盘文件。

第二节 电子文件长期保存的概念及认识误区

一 电子文件长期保存的概念

电子文件是数字遗产的一种类型。联合国教科文组织《数字遗产保存指南》关于数字保存(digital preservation)的定义是“数字保存是对以数字形式存在的信息和其他类型遗产的保护过程”。^①英国国家档案馆数字保存联盟 JISC、澳大利亚国家档案馆对数字保存的定义相同,即“数字保存是指为了确保数字对象在未来能被我们获取和利用而对其采取的主动管理过程”。^{②③}以上对数字保存的定义覆盖了所有类型的数字对象,其中,电子文件与其他数字保存对象相比,具有对形成机构和组织的凭证价值,因此,对于电子文件而言,文件的真实性是对其进行长期保存的基本前提。ISO 15489-1 对文件保存的定义如下:“文件保存(preservation)是确保真实性文件的技术和内容的长期保存的过程及操作”,即文件的长期保存涉及两个方面:文件形成的技术要素(如存储载体、存储格式等)和内容要素。传统纸质档案和电子文件在技术和内容方面的长期保存存在较大的差异性。纸质档案的内容、结构(格式)和载体是固定的,一旦形成不可转换或改变。因此,纸质档案的长期保存主要是针对其载体和内容的原生

① <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>.

② http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_plfinalreport.pdf.

③ <http://www.naa.gov.au/about-us/organisation/accountability/operations-and-preservation/digital-preservation-policy.aspx#section2>.

性或原貌的维护和保持。而电子文件长期保存的技术要素和内容要素较为复杂,传统的“原生性”或“原始记录”概念不再适用于电子文件,因为“原始”的电子文件是以计算机编码形成和组织的,而呈现在显示器上的我们可以理解和识读的电子文件信息内容则是经过计算机软件“编译”而成,已经不是其“原件”。而且,电子文件可以在不同载体间转换,可以有若干数字副本,同时在不同的客户端显示和操作。因此,确切地说,从“原生性”或“原始记录性”角度来理解和实现电子文件长期保存不是很适用,电子文件长期保存应该是对其真实性、完整性、可靠性和可用性的长期保障。为此,在电子文件长期保存过程中,必须全面考虑文件的三个基本构成要素,将电子文件的背景、结构与电子文件的内容一起保存,为电子文件长期保存奠定基础。澳大利亚国家档案馆《电子文件长期保存指南》(2002)分析了电子文件与传统档案的不同特点,并对电子文件长期保存的方式提出了建设性意见,值得我们参考和借鉴。^①

二 电子文件长期保存面临的挑战及认识误区

电子文件长期保存面临严峻挑战,最主要的威胁来自两个方面:技术过时和存储介质物理性能的退化。传统纸质档案如果保管条件得当可以留存几个世纪,但是电子文件对载体及所生成的数字环境依赖性很强,而计算机技术发展日新月异,由于技术更新、软硬件环境的变化和载体老化而导致电子文件内容无法读出的状况令人担忧。具体而言,电子文件长期保存面临如下挑战:①电子文件对机器的依赖性。电子文件在数字设备和环境中产生,必须依赖特定的计算机硬件和软件才能阅读和处理。②技术过时是电子文件长期保存的最大威胁。技术更新的周期越来越短,可能仅有2~5年,这要求人们必须在下一个技术更新周期来临之前就采取对电子文件的保护行动,否则难以保证电子文件内容能够在不断更新的技术环境中长期可读、可用。③存储介质的脆弱性。即使在存储介质并无受损的情况下,介质本身的不稳定性和不适宜的保管条件也会使其物理性能快速退化。④为了应对上述技术过时和载体退化的挑战而采取的一些措施又可能会对长期保持电子文件的完整性和真实性产生影响。⑤应尽早采取电子文

① National Archives of Australia. An Approach to the Preservation of Digital Records. http://www.naa.gov.au/Images/An_approach_Green_Paper_tcm16-47161.pdf.

件长期保存行动，否则数字信息内容会在极短时间内丢失或无法读出。

⑥应采取基于文件生命周期的管理方法对电子文件进行维护。⑦此外，电子文件长期保存还面临着法律、制度及经济问题。为了实现电子文件长期保存，应该在其设计和产生阶段就主动实施长期保存方案，而这就会更多地牵涉机构内部及机构之间的利益以及利益相关者的角色变化。

可见，电子文件的长期保存涉及很多因素，需要采取各类方法和途径。一个普遍的共识是，数字信息长期保存最经济有效的办法是及早、主动地实施长期保存方案。电子文件长期保存越早开始越能节约成本，最好是在它们产生之初就主动实施覆盖其整个生命周期的长期保存计划。恢复或者再现已经丢失或无法读取的电子文件信息比一开始就实施对其的长期保存的成本要高出很多，更何况电子文件一旦丢失或无法读取，就很可能没有了再现和再造的机会。

人们在电子文件长期保存的活动中，存在一些常见的认识误区，例如，关于电子文件长期保存的期限问题，不少人对此并不清楚，有人悲观地认为只有数年，有人以为与纸质档案一样可达数个世纪。其实，电子文件难以持久存取，根据美国国家档案与文件署及国际公认的期限，电子文件保存 50 年以上即可认为是长期保存。

此外，Adrian Brown 从一般的组织机构层面，对数字信息长期保存的认识误区进行了剖析，他在这个问题上独到的个人见解虽然可能存在偏颇，但能在很大程度上增强中小型组织机构对电子文件长期保存的信心和勇气。

误区之一，电子文件长期保存只有国家机构才能解决，小型机构对此无能为力。国家机构无疑能够引领电子文件长期保存的重要实践和原则，但是，根据已有的成熟经验、经济基础、实践工具和服务手段，对于任何类型和大小的机构而言，电子文件的长期保存不仅是现实可行的，而且势在必行。

误区之二，电子文件长期保存需要巨大的资金投入才能实施。人们很容易想到的一个典型案例是，美国国家档案与文件署为了长期保存联邦政府所产生的电子文件，投入 3 亿美元建设电子文件档案馆（ERA）。但这并不意味着不同规模和不同类型的机构都需要花费巨资才能实施其电子文件长期保存计划。比如，与 ERA 的巨资投入相比，英国考古学遗产中心（English Heritage Center for Archaeology）仅以数百英镑的成本投入就可以

建设其数字保存库。当然，这需要合理规划，及充分利用已有的工具、资源和系统。

误区之三，电子文件长期保存需要高深的技术才能实施。技术领域无疑占据了数字信息长期保存的研究前沿，但是，数字保存实践并不要求十分高深的技术背景。在信息管理人才培养方案和培训计划中，数字保存作为重要的职业技能之一，已经成为人才培养和职业培训方案的重要内容。因此，来自档案馆、图书馆和 IT 行业的档案保管者具有很多不同渠道的学习和培训机会，对于他们而言，数字保存职业能力并不是难以获得或不可提升的。

误区之四，电子文件长期保存的时机还未成熟，是以后或未来的任务。如前所述，电子文件长期保存越早开始越能节约成本，越经济有效。如果等到电子文件载体老化或信息丢失、无法读出之时才考虑其长期保存问题就为时已晚了。Adrian Brown 指出，对于一个组织机构而言，如果它们要实现巨大的利益，而且需要避免潜在的法律、财务、管理和声誉风险以及具有重大历史价值和经济价值的信息丢失，那么，电子文件的长期保存是迫切需要着手处理的问题。当然，这并不是指必须立即解决电子文件长期保存领域所有的问题或者与其他组织机构具有完全一样的数字保存需求。对于一个尚未实施电子文件长期保存的组织机构而言，现在就是开始实际行动的最好时机。^①

第三节 数字保存的发展历史及现状

一 欧美数字保存的发展历史及现状

1994 年 RLG（研究图书馆组织）对企业、博物馆、档案馆、图书馆、出版社、学术团体和政府机构等各类组织机构的数字信息保存状况展开全面调研，完成了《数字信息归档报告》并于 1996 年发布，该报告揭示了数字信息长期保存所面临的组织、技术、法律、经济等关键问题，分析在解决每一类问题时所遇到的主要障碍，并对消除每一种障碍因素提出了行

^① Brown Adrian. *Practical Digital Preservation: A How - to Guide for Organizations of Any Size* (London: Facet Publishing, 2013), 5 - 6.

动建议,研究技术更新的替代方案。^①该报告对于数字信息长期保存各种问题的揭示引起了政府及文化遗产部门的广泛关注,使人们深刻认识到数字信息的脆弱性及难以持久保存,报告还首次提出了数字信息长期保存的核心概念和术语,如“可信数字保存系统的认证”、“格式登记”、“成本模型”,以及“完整性”和“真实性”等基本术语,为数字信息长期保存理论与实践的进一步发展奠定了十分重要的基础。

从1996年至2014年,数字保存领域的发展十分活跃。数字保存理论及标准是基础,而在实践中不断变化的保存策略及保存工具又使理论及标准得到了进一步的拓展。国际标准OAIS参考模型在数字信息长期保存领域发挥了重要作用。OAIS最初由美国空间数据系统咨询委员会(CCSDS)于20世纪90年代开始研制,1996年发布了建议草案并作为业内标准得到迅速广泛的应用,最终于2003年正式成为一项国际标准(ISO 14721:2003),并于2012年更新。OAIS不仅提出了数字信息长期保存的功能模型和信息模型,为构建可信数字保存系统的基本框架及研制数字保存元数据方案提供了重要的参考,而且还构建了一套概念模型,规范了数字保存的基本术语,比如,OAIS对于“数字迁移”的规定与很多人对“迁移”的理解有所不同。OAIS认为,由于技术发展和存储介质物理性能的老化,无论今天数字资源维护得多好,随着时间的推移,最终都会将大部分的资源转移到不同的媒体或不同的硬件或软件环境中去,以维持它们的可存取性,即“数字迁移”。按照信息丢失风险依次增加的顺序,OAIS将数字迁移划分为更新、复制、重新包装和转化4种方式。在数字保存标准体系中的另一类重要标准是元数据标准。数字保存元数据标准在被广泛采用的元数据标准框架如METS(2001)和PREMIS(2003)的基础上得以研制,而OAIS所提出的“信息包”概念及对提交信息包、档案信息包和分发信息包的不同分类,为数字保存元数据标准的制定提供了重要的依据和参考。

组织机构通过构建自己的数字保存方案及其合作研究积极地推动了数字保存实践的发展。例如,CEDARS项目和丹麦国家档案数字保存测试平台(Dutch National Archief's Digital Preservation Testbed)都运用了严格的科学原则发展和测试了多种数字保存策略。最早的可信数字仓储一般由国

^① Garrett John, Waters Donald, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information Commissioned by the Commission on Preservation and Access and the Research Libraries Group*. (Washington, D. C.: Commission on Preservation and Access, 1996).

家文化遗产机构建立，如：澳大利亚国家图书馆、荷兰国家图书馆和英国国家档案馆相继在 21 世纪初建立了数字库。随着高质量的数字保存系统软件的出现，数字保存不再局限于国家机构层面，也使中小规模的机构如大学图书馆、地方档案馆及企业档案馆建立自己的数字保存库成为可能。大量的开源软件如 Fedora（1997）、EPints（2000）、DSpace（2000）不断涌现，商业软件如 Safety Deposit Box（2003）和 Rosetta（2008）也进入了市场，最新的基于云计算的 DuraCloud（2011）和 Preservica（2012）提供了基于云服务的数字保存模式，这些都为小型机构建设数字保存系统提供了条件。同时，可信数字保存库管理软件和开发工具如 PRONOM technical registry（2002），JHOVE characterization tool（2003）以及 DROID format identification tool（2005）在很多组织机构的数字保存系统开发中发挥了重要的作用。另外，对于网页的长期保存实践几乎与网络相伴而生，从 1996 年 Internet Archive 和 1997 年 Nordic Web Archive 的最早建立到地方、国家和国际的 Web 档案馆项目都推动了网页信息长期保存的实践发展。

2000 年以来许多数字保存研究项目得到了发展，如美国著名的 NDIIPP（2000）和欧盟委员会的各类数字保存研究项目包括 ERPANET（2001）、DELOS（2004）、Digital Preservation Europe（2006）和 Planet（2006）等。数字保存领域的联盟和基金会在数字保存中发挥了重要的推动作用，如英国的数字保存联盟 Digital Preservation Coalition（DPC）和数字管理中心 Digital Curation Centre，丹麦国家数字保存联盟 Dutch Nationale Coalitie Digitale Duurzaamheid，以及国际开放星球基金会 Open Planets Foundation。上述联盟组织和研究项目的开展标志着以数字保存业界联盟为基础的、在国家层面和国际范围内的合作得到了深入发展。

在数字保存早期，人们对数字信息的长期保存常常依赖于对实物的保存，着眼于媒介储存，强调发展更多耐用的媒介以及复制更多耐用的存储媒介，将数字信息打印在纸上就是一个显著的例子。而新的保存模式强调覆盖文件生命周期整个过程，在数字信息产生之时就应该主动保存，而且与更多的信息保管者进行合作并获得其支持尤为重要。2000 年以后，转换、更新和技术保存等依然是最基本的保存策略，但是具体实现方式更为丰富而且不断发展，如：数字考古、文件格式标准化、保存大量复印件和发展可长期保存的元数据、比特流复制、通用虚拟计算机的运用、大容量

存储系统,甚至包括所谓的“无作为”策略等等,上述数字保存策略的结合使用比单独使用其中的一种更为有效。近些年来,数字保存的关注点已经由技术角度转换到数字保存的可持续性问题。其中,两种关键的认识使人们开始转换新的保存模式并建立新的保存策略和实践:一种认识是,将数字内容与技术分离;另一种认识是,应当使更多的机构和组织参与数字保存过程,并在相关政策的引导下协调它们的活动。

2004年,联机计算机图书馆中心和图书馆研究组织(OCLC/RLG)公布了一项覆盖13个国家的国家图书馆、大学图书馆、研究性图书馆、企业、档案馆、博物馆等各类机构数字保存实践的国际调查。在48个被调查者中,92%正在实现(或准备实现)标准化、迁移或升级,同时,大多数被调查者表示他们拥有多种保存策略。最受欢迎的策略就是比特流(位流)保存,有85%的被调查者实施了该策略,紧随其后的策略是限制访问和标准化(这两种都是控制版本过多的方式)以及迁移。^①2011年,ARL(研究图书馆协会)对其72个成员的数字保存情况调查表明,人们对于数字保存政策重要性的认识增强了,超过50%的受访者表示他们已经有了相关政策。在这项调查中,主要的数字保存策略包括:备份(93%的实施率),存储于安全系统中(76%),校验(63%),迁移(50%),更新(47%)以及升级(7%)。许多机构在数字保存过程中采用了相关软件,而DSpace是最为流行的一种。^②

2008年9月,Boyle, Eveleigh 和 Needham 对英国地方的档案数字保存状况展开了一项较为全面的调查,共收到38个机构的反馈问卷。^③调查结果表明,绝大多数英国地方档案馆对数字保存的重要性有基本认识,74%的档案馆了解一些基本的数字保存资源如英国数字保存联盟DPC,50%左右的档案馆知道数字信息长期保存国际标准OAIS以及大英图书馆和英国

① OCLC/RLG PREMIS Working Group, *Implementing Preservation Strategies for Digital Material: Current Practice and Emerging Trends in the Cultural Heritage Community* (Dublin, OH: OCLC), 2004.

② Li Yuan, Banach Meghan, *Institutional Repositories and Digital Preservation: Assessing Current Practices at Research Libraries*, *D - Lib Magazine*, 2011, 17 (5/6), <http://www.dlib.org/dlib/may11/yuanli/05yuanli.print.html>.

③ Boyle Frances, Eveleigh Alexandra, Needham Heather, *Preserving Local Archival Heritage for Ongoing Accessibility* [EB/OL]. [2013 - 08 - 12]. <http://www.ariadne.ac.uk/issue58/boyle-et-al>.

国家档案馆在数字保存领域的重要举措。但是，有三分之二的档案馆不知道其他国际标准如 PREMIS 和 METS 以及重要的数字保存项目如“东英格兰数字档案馆区域试点（East of England Digital Archive Regional Pilot）”项目和 Paradigm 项目。在所有被调查对象中，约 47% 的档案馆已经制定了数字保存政策，但是，只有极少数档案馆在此基础上引进了相关标准及采取实际行动。大多数（79%）档案馆认为自己是被动地满足数字保存的需求，而不是主动地加强自己的数字保存能力。虽然档案馆都拥有接收进馆的原始数字资源，但是它们通常缺乏对这部分数字资源的详细了解，比如具体的案卷或文件规模，而且普遍存在各种已经过时的文件格式。由于缺乏对馆藏数字资源基本状况的调查和信息集中，大多数档案馆无法执行任何形式的数字保存计划。多数档案馆采取了备份、服务器存储等数字保存方式，87% 的档案馆拥有一些存储在 CD 或 DVD 等光学介质上的数字信息。42% 的档案馆只是简单地将数字信息存储在其原来的介质上，只有极少数档案馆实行了更为复杂的操作，如生成校验或者转换成标准格式。在所有的调查对象中，只有一个档案馆应用了内容管理系统，一个档案馆采用了外包存储，接近一半的档案馆使用了电子文件管理系统。此外，三分之二的地方档案馆在数字信息的利用方面不尽如人意，它们没有开发规范高效的检索系统，只能在线提供图像浏览而不支持对其他类型数字资源的在线利用。

数字保存的最主要障碍是此次调查的核心问题。调查结果在意料之中：资金不足是最主要的障碍，其次是缺乏 IT 支持和技术准备，再次是缺少政策上的支持。除了上述三种主要障碍因素以外，档案馆工作人员的动力、领导能力、时间、战略合作伙伴缺乏等也是阻碍数字保存的次要因素。总体上，档案馆对于数字保存具有积极的愿望，但是资金和技术不足是最大的障碍，而加大财政投入和加强对档案人员的技术培训被认为是解决之道。总之，根据 2008 年的此次调查，英国地方档案馆在数字保存中的问题主要是资金和技术能力不足的问题，而不是缺乏数字保存意识和愿望。因此，在数字保存中试图采取一种完美的、一步到位的解决方案是不现实的，只有逐步积累实践经验才能有所突破。

二 中国大陆地区的数字保存状况

2006 年，武汉大学刘家真教授课题组对我国 57 个机构的数字资源保