

国家自然科学基金项目“数字保存的风险型元数据与风险监控研究”
(批准号71673255)研究成果

数字保存的风险型元数据与 风险监控研究

臧国全等 著



WUHAN UNIVERSITY PRESS
武汉大学出版社

国家自然科学基金项目“数字保存的风险型元数据与风险监控研究”
(批准号71673255)研究成果

数字保存的风险型元数据与 风险监控研究

臧国全等 著



WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

数字保存的风险型元数据与风险监控研究/臧国全等著. —武汉: 武汉大学出版社, 2019.6

ISBN 978-7-307-20842-1

I. 数… II. 臧… III. 数据保护—风险管理—研究 IV. TP309.2

中国版本图书馆 CIP 数据核字(2019)第 065841 号

责任编辑: 林莉 沈继侠

责任校对: 李孟潇

整体设计: 马佳

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮箱: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 北京虎彩文化传播有限公司

开本: 787×1092 1/16 印张: 19.25 字数: 456 千字 插页: 1

版次: 2019 年 6 月第 1 版 2019 年 6 月第 1 次印刷

ISBN 978-7-307-20842-1 定价: 58.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

前 言

本书是国家自然科学基金项目“数字保存的风险型元数据与风险监控研究”(批准号71673255)的研究成果。

由于数字资源的脆弱性、载体依赖性、保存格式易过时性等特征,对其实施长期保存是一个充满风险的过程。所以,识别风险是数字保存的一项常规工作,实施这项工作的基本方法是建立数字保存的风险型元数据,并在此基础上进行风险监控。这是本书选题的缘故。

除引言外,本书内容由四部分组成:数字保存的风险检测条件、数字保存的风险检测、数字保存的风险监控、数字保存的风险型元数据。其中,第一部分是风险检测的基础,第二部分是基于第一部分实施的风险检测,第三部分是基于第二部分实施的风险监控,第四部分是基于第二部分汇总得出的风险型元数据。

第一部分是数字保存的风险检测条件。数字保存的风险检测需要两个基础条件:一是检测项目,是对数字对象进行检测所需的风险识别列表,也即风险识别模型;二是检测对象,是对“数字对象的与风险相关的属性”的描述记录,也即用于风险检测的元数据。

在风险识别模型建立方面。首先,分析数字保存风险识别模型的基本要求:概念清晰性、风险专指度的合适性与一致性、全面性、易用性。其次,识别数字保存的核心职责:可用性、可识别性、持续完整性、可呈现性、可理解性、真实性。最后,针对每项核心职责,分析列出相应的风险,构建数字保存的风险识别模型。

在用于风险检测的元数据选择方面。首先,选择《PREMIS 元数据字典》用于数字保存风险检测的适用性验证的理由:保存型元数据是用于记录数字对象长期保存的相关操作信息,保存系统使用这些信息实现维护数字对象可用性、可呈现性、可理解性、真实性和可识别性等的核心保存职责;数字保存风险存在于数字对象的长期保存过程中;在业已存在的各种元数据中,保存型元数据最接近适合用于数字保存的风险检测;已有的保存型元数据方案有多个,其中《PREMIS 元数据词典》是应用最多的一个,且已成为国际上的“事实标准”。其次,验证《PREMIS 元数据词典》用于数字保存风险检测的适用性:设计检测方法,实施《PREMIS 元数据词典》与保存风险之间的映射,设计检测方案,进行检测实验。最后,得出结论:《PREMIS 元数据词典》可以用于数字保存的风险检测,但存在不少问题,有必要重新设计一套数字保存的风险型元数据。

第二部分是数字保存的风险检测。包括单风险检测和全风险检测。其中,单风险检测是全风险检测的基础,全风险检测是单风险检测的延伸。

在数字保存的单风险检测方面。首先,单风险检测准备,包括单风险检测的基本思路、术语界定、实验数据采集、单风险检测算法、检测代码编制说明。其次,各类单风险

检测的实施,包括可用性、可识别性、持续完整性、可呈现性、可理解性、真实性六种单风险检测,每类单风险检测实施分为单风险的含义与类型、单风险的风险型元数据、检测流程图、检测实验。

在数字保存的全风险检测方面。首先,全风险检测准备:在各单风险检测基础上实施,包括全风险检测的术语界定、实验数据采集、全风险检测算法。其次,全风险检测实验,包括单维度全风险检测实验、二维度全风险检测实验、三维度全风险检测实验、四维度全风险检测实验。

第三部分是数字保存的风险监控。包括数字保存的风险监控准备和风险监控实验。

在数字保存的风险监控准备方面。首先,风险监控的研究思路,实质是在时间维度上连续的风险检测,故在风险检测基础上实施。其次,设计风险监控算法,包括单维度单风险、多维度单风险、单维度全风险、多维度全风险等监控算法。

在数字保存的风险监控实验方面。包括单维度单风险监控实验、二维度单风险监控实验、三维度单风险监控实验、单维度全风险监控实验、二维度全风险监控实验、三维度全风险监控实验等六类。

第四部分是数字保存的风险型元数据。包括数字保存的风险型元数据建立、风险类型与风险型元数据元素之间的对应。

在数字保存的风险型元数据建立方面。汇总各个单类风险的风险型元数据,形成数字对象方面、保存事件方面、行为主体方面、保存政策方面、产权方面5种类型的风险型元数据。

在风险类型与风险型元数据元素之间的对应关系建立方面。在各个单风险的风险型元数据基础上,编制《保存风险类型与风险型元数据元素之间的对应表》。

本书由臧国全拟定大纲并负责组织撰写。博士研究生李哲、张凯亮,以及硕士研究生朱晓庆、刘田、臧弘毅、王秀娟、周丽媛、崔焱、周晓倩、贾瑞莹、张男等参加内容研究、书稿撰写和校对工作。

在本书内容研究和撰写过程中,项目组参考了大量的中外文文献,由于篇幅所限,未能一一列出。在此项目组一并对所有参考文献作者表示诚挚的谢意。

由于作者水平有限,不足之处敬请批评指正。

数字保存的风险型元数据与风险监控是业界一个崭新的研究领域,随着其理论探讨和实践的推进,本书的一些理论、方法和结论也需要进一步完善,为此,我们将在该领域继续探索。

作者

2019年3月

目 录

1 引言	1
1.1 本书背景与研究意义	1
1.2 国内外研究现状及发展动态分析	2
1.2.1 有关数字保存的风险型元数据	2
1.2.2 有关数字保存的风险监控	5
1.2.3 国内外研究趋势	8
1.3 研究思路	8
1.3.1 数字保存风险型元数据的研究思路	8
1.3.2 数字保存风险监控方法的研究思路	10
1.3.3 实验思路	11
1.4 研究内容	13
1.5 研究方法	13
2 数字保存的风险检测条件	15
2.1 数字保存的风险识别模型	15
2.1.1 模型的基本要求	16
2.1.2 模型的建立	20
2.1.3 模型与基本要求的相符性分析	27
2.1.4 模型的应用	28
2.2 《PREMIS 元数据字典》的风险检测适用性实验	29
2.2.1 实验方法的设计	29
2.2.2 《PREMIS 元数据词典》与保存风险之间的映射	30
2.2.3 保存风险的检测实施	32
2.2.4 实验	35
2.2.5 结论	37
3 数字保存的风险检测	39
3.1 单风险检测	39
3.1.1 检测准备	39
3.1.2 数字保存的可用性风险检测	43
3.1.3 数字保存的可识别性风险检测	63

3.1.4 数字保存的持续完整性风险检测	83
3.1.5 数字保存的可呈现性风险检测	100
3.1.6 数字保存的可理解性风险检测	116
3.1.7 数字保存的真实性风险检测	137
3.2 数字保存的全风险检测	151
3.2.1 检测准备	151
3.2.2 单维度全风险检测	154
3.2.3 二维度全风险检测	161
3.2.4 三维度全风险检测	172
3.2.5 四维度全风险检测	178
4 数字保存的风险监控	181
4.1 监控准备	181
4.1.1 风险监控实施的基本思路	181
4.1.2 术语界定	181
4.1.3 监控的数字对象样本获取	183
4.1.4 监控算法描述	184
4.2 单维度单风险监控	186
4.2.1 单维度可用性风险监控	187
4.2.2 单维度可识别性风险监控	188
4.2.3 单维度持续完整性风险监控	190
4.2.4 单维度可呈现性风险监控	192
4.2.5 单维度可理解性风险监控	193
4.2.6 单维度真实性风险监控	195
4.3 二维度单风险监控	196
4.3.1 {CD, RP} 二维度可用性风险监控	196
4.3.2 {CD, RP} 二维度可识别性风险监控	200
4.3.3 {CD, RP} 二维度持续完整性风险监控	203
4.3.4 {CD, RP} 二维度可呈现性风险监控	206
4.3.5 {CD, RP} 二维度可理解性风险监控	209
4.3.6 {CD, RP} 二维度真实性风险监控	212
4.3.7 {SD, RP} 二维度可用性风险监控	215
4.3.8 {SD, RP} 二维度可识别性风险监控	217
4.3.9 {SD, RP} 二维度持续完整性风险监控	221
4.3.10 {SD, RP} 二维度可呈现性风险监控	224
4.3.11 {SD, RP} 二维度可理解性风险监控	226
4.3.12 {SD, RP} 二维度真实性风险监控	230
4.4 三维度单风险监控	232

4.4.1	{CD, SD, RP} 三维度可用性风险监控	232
4.4.2	{CD, SD, RP} 三维度可识别性风险监控	234
4.4.3	{CD, SD, RP} 三维度持续完整性风险监控	235
4.4.4	{CD, SD, RP} 三维度可呈现性风险监控	236
4.4.5	{CD, SD, RP} 三维度可理解性风险监控	237
4.4.6	{CD, SD, RP} 三维度真实性风险监控	239
4.5	单维度全风险监控	240
4.6	二维度全风险监控	241
4.6.1	{CD, RP} 二维度全风险监控	241
4.6.2	{SD, RP} 二维度全风险监控	245
4.7	三维度全风险监控	248
5	数字保存的风险型元数据	250
5.1	数字对象方面的风险型元数据	250
5.1.1	数字对象标识符(Object Identifier)	250
5.1.2	数字对象类型(Object Category)	251
5.1.3	检索点信息(Retrieval Point Information)	251
5.1.4	文件系统信息(File System Information)	252
5.1.5	呈现环境信息(Browsing Environment Information)	253
5.1.6	保存级别(Preservation Level)	257
5.1.7	编解码信息(Encoding and Decoding Information)	258
5.1.8	存储信息(Storage Information)	259
5.1.9	对象的关系(Object Relationship)	260
5.1.10	保存策略(Preservation Strategy)	262
5.1.11	重要外观特征(Significant Appearance Characteristic)	264
5.1.12	存储位置(Storage Location)	264
5.1.13	固定性信息(Fixity Information)	265
5.1.14	签名信息(Signature Information)	265
5.1.15	文件大小(Size)	267
5.1.16	存储介质(Storage Medium)	267
5.1.17	重要属性(Significant Properties)	268
5.1.18	文件格式(Format)	269
5.1.19	创建程序(Creating Application)	270
5.1.20	呈现次序(Presentation Sequence)	271
5.1.21	数字对象内部结构信息(Object Interior Structure Information)	271
5.1.22	数字对象外部结构信息(Object Outer Structure Information)	272
5.1.23	整体式语义信息(Unitary Semantic Information)	273
5.1.24	知识点式语义信息(Knowledge Point Semantic Information)	274

5.1.25 还原信息(Recovery Information)	275
5.1.26 信息摘要(Message Digest)	276
5.1.27 修改历史信息(Modification History Information)	277
5.1.28 描述信息(Description Information)	278
5.2 保存事件方面的风险型元数据	278
5.2.1 去索引(Deaccession)	278
5.2.2 删除(Deletion)	280
5.2.3 介质刷新(Storage Medium Refresh)	280
5.2.4 病毒检测(Virus Check)	280
5.2.5 压缩(Compression)	281
5.2.6 加密(Encryption)	281
5.2.7 数字迁移(Digital Migration)	281
5.2.8 固定性检查(Fixity Check)	282
5.2.9 信息摘要计算(Message Digest Calculation)	282
5.2.10 解压(Decompression)	282
5.2.11 外部结构判定(Outer Structure Determination)	283
5.2.12 语义信息识别(Semantic Information Identification)	283
5.3 行为主体方面的风险型元数据	284
5.3.1 软件(Software)	284
5.3.2 组织机构(Organization)	285
5.3.3 个人(Person)	285
5.4 保存政策方面的风险型元数据	285
5.4.1 检索点(Retrieval Point)	285
5.4.2 文件系统(File System)	286
5.4.3 存储介质的使用寿命(Media Life)	286
5.4.4 介质刷新频率(Media Refresh Frequency)	286
5.4.5 病毒检测周期(Virus Check Period)	287
5.4.6 数字对象标识符框架(Object Identifier Frame)	287
5.4.7 数字对象的关系(Object Relationship)	287
5.4.8 安全备份政策(Safety Backup Policy)	287
5.4.9 数字迁移基准准确率(Digital Migration Benchmark Accuracy)	288
5.4.10 固定性检查周期(Fixity Check Period)	288
5.4.11 内部结构信息列表(Interior Structure Information List)	288
5.4.12 最佳浏览工具判定(Best Browsing Tool Determination)	288
5.5 产权方面的风险型元数据	289
5.5.1 版权信息(Copyright Information)	289
5.5.2 许可信息(License Information)	290
5.6 保存风险类型与风险型元数据元素之间的对应	292

5.6.1 保存风险类型至风险型元数据元素的对应表 292

5.6.2 风险型元数据至保存风险类型元素的对应表 293

参考文献..... 296

1 引言

1.1 本书背景与研究意义

本书缘起。人类社会进入了数字时代，数字保存已经成为人类社会的一项基本职责。宏观上，数字保存涉及技术和管理两大问题。作者多年来一直致力于数字保存的管理研究，在质量管理方面，主持完成了国家自然科学基金“数字保存系统质量保证标准与认证研究”(批准号 70673094, 2010 年结项); 在经济管理方面，主持完成了国家自然科学基金“数字保存经济要素与经济评价研究”(批准号 71173197, 2016 年结项)。本书内容也属于数字保存的管理研究，是数字保存风险管理的一个方面。因此，可以说，本研究是作者已有研究的延续。

风险型元数据含义与应用。数字保存需要周期性(或实时)的风险监控，以期发现保存的哪些数字对象出现风险; 就像人体需要周期性的健康检查，以期发现哪些器官出现异常。人体的健康检查是通过检测人体与健康相关的指标来实现。同样道理，数字保存的风险监控可以通过检查数字对象的与风险相关的属性来实现。这种“数字对象的与风险相关的属性”即是本研究的“风险型元数据”。因此，建立风险型元数据是风险监控的基础，风险监控是风险型元数据的应用。

研究意义。由于数字资源的脆弱性、载体依赖性、保存格式易过时性等特征，对其实施长期保存是一个充满风险的过程。所以，识别风险是数字保存的一项常规工作，实施这项工作的基本方法是建立数字保存的风险型元数据，并在此基础上进行风险监控。因此，研究意义可概括为：

(1) 有利于丰富数字时代图书馆学的理论与实践。传统馆藏资源建设是图书馆学的重要内容，数字保存是传统馆藏资源建设在数字时代的延伸，其风险监控也是传统馆藏资源管理与保存的理论与实践在数字时代的扩展。

(2) 有利于数字保存项目的可持续性运营。数字保存项目的整个生命周期都伴随着风险管理，风险监控是风险管理的一个重要环节，也是风险规避与风险排除的前提。因此，实施风险监控有助于减少数字保存项目风险可能导致的损失，提高数字资源的保存质量，延长保存项目的运营周期。

(3) 有利于数字保存行业的健康发展。风险监控在很多行业都有成熟的应用，借鉴其经验，建立数字保存行业的风险监控方法，实现对数字保存核心职责的各种威胁的预警，为其健康发展提供基础保障。

1.2 国内外研究现状及发展动态分析

1.2.1 有关数字保存的风险型元数据

据普查, 目前为止, 针对数字保存的风险型元数据未见报道。

数字保存的风险存在于数字资源长期保存过程之中, 也存在于整个管理过程之中。所以, 与风险型元数据密切相关的有保存型元数据和管理型元数据。因此, 本部分文献综述就针对上述两类元数据方案展开。当然, 风险型元数据还与其他类型元数据(如描述型元数据)相关, 但关联度相对较低, 鉴于篇幅, 不作综述。

(1) 保存型元数据方案方面。保存型元数据是实现和支撑数字资源长期保存的相关操作元数据。国外的保存型元数据方案已较成熟, 国内则大多包含在综合性元数据方案之中。

①国外研究现状。美国学者 Bearman 和 Sochats^①(1996)在《基于证据需求的元数据》中较早论述保存型元数据的概念与作用。之后, 出现的保存型元数据方案主要有: a. 美国 RLG^②(1998)的《数字资源保存型元数据》; b. CCSDS^③(1999)的《OAIS 参考模型》中针对数字对象建立的保存描述信息框架; c. 澳大利亚国家图书馆^④(1999)的《PANDORA 保存型元数据方案》; d. 网络化欧洲存储图书馆^⑤(2000)的《NEDLIB 长期保存的元数据模型》; e. 英国 Cedars 项目^⑥(2000)的《数字保存的元数据方案: Cedars 项目规范框架》; f. 康奈尔大学图书馆^⑦(2000)的《保存型元数据结构》; g. OCLC^⑧(2001)的《支持数字对象保存的元数据框架》; h. OCLC 和 RLG(2002)的《保存型元数据元素集: 定义与示例》^⑨和

① Bearman D, Sochats K. Metadata Requirements for Evidence [EB/OL]. [2018-07-01]. <http://www.archimuse.com/papers/nhprc/BACartic.html>.

② Research Libraries Group. Digital Resource Preservation Metadata [EB/OL]. [2018-03-01]. <http://www.oclc.org/research/preserv/presmeta.htm>.

③ CCSDS. Reference Model for an Open Archival Information System (OAIS) [EB/OL]. [2018-02-10]. <http://www.ccsds.org/documents/pdf/>.

④ National Library of Australia. PANDORA Preservation Metadata for Digital Collections [EB/OL]. [2018-03-11]. <http://www.nla.gov.au/preserve/pmeta.html>.

⑤ NEDLIB. Metadata for Long Term Preservation [EB/OL]. [2018-04-18]. <http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf>.

⑥ CEDARS. Metadata for Digital Preservation: The CEDARS Project Outline Specification [EB/OL]. [2018-05-01]. <http://www.leeds.ac.uk/cedars/MD-STR~5.pdf>.

⑦ Cornell University Library. Preservation Metadata Structure [EB/OL]. [2018-03-19]. <http://www.cornell.edu/academics/library.cfm>.

⑧ OCLC. A Metadata Framework to Support the Preservation of Digital Object [EB/OL]. [2018-01-03]. http://www.oclc.org/research/projects/pmwg/pm_framework.pdf.

⑨ OCLC, RLG. Preservation Metadata Element Set — Definitions and Examples [EB/OL]. [2018-06-20]. <http://www.oclc.org/digitalpreservation/archiving/metadataset.pdf>.

《数字对象的保存型元数据：现状综述》^①；i. 欧盟 Minerva 项目^②(2002)的《数字对象保存型元数据》。

2003 年至今，保存型元数据的研究与实践聚焦在 PREMIS 的《保存型元数据词典》上。

同年，OCLC 和 RLG 成立“保存型元数据实施战略”(PREMIS)工作组，对全球 70 个数字保存的研究机构和保存项目进行调查和深度采访，在其发布的《数字文化遗产的保存：现状与未来》^③(2004)中，全面考察了保存型元数据。基于该调研报告，参考 OAIS，该工作组(2005)编制了《保存型元数据词典》^④，包括 5 个实体(如，数字对象)、28 个一级语义单元(如，数字对象存储)、102 个元素(如，数字对象存储介质)。2008 年、2010 年和 2015 年，该词典分别进行了 3 次修订。

目前，PREMIS 的《保存型元数据词典》已经成为国际上的“事实标准”，广泛应用在众多数字保存项目中。

②国内研究现状。国内的保存型元数据方案大多基于 OAIS 参考模型，较有影响的有：清华大学图书馆^⑤(2001)的《馆藏数字资源元数据方案》中的“保存型元数据”部分；国家图书馆^⑥(2002)的《中文元数据方案》中的“保存型元数据”部分；国内军队院校数字图书馆系统^⑦(MDLS)中设计的《保存元数据方案》(2003)；科技部科技基础性工作专项资金重点项目“我国数字图书馆标准与规范建设”^⑧(2003—2009)的成果《基本元数据规范》和《专门元数据规范》中涉及的保存型元数据。目前，大型保存项目的元数据方案中都设置有保存型元数据部分(如知网、维普等)。

(2)管理型元数据方案方面。管理型元数据是对数字保存的管理政策与管理机制进行描述的元数据。国外的管理型元数据方案比较成熟，国内鲜有独立的管理型元数据方案。

①国外研究现状。美国国会图书馆^⑨(1998)在《数字仓储的核心元数据元素》中较早提出管理型元数据。之后，主要研究有两类：

① OCLC, RLG. Preservation Metadata for Digital Objects: A Review of the State of the Art[EB/OL]. [2018-07-30]. http://www.oclc.org/research/pmwg/presmeta_wp.pdf.

② Minerva. Preservation Metadata for Digital Objects [EB/OL]. [2018-09-01]. <http://www.ncecho.org/presmet/pmdo.htm>.

③ OCLC, RLG. Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community[EB/OL]. [2017-12-01]. <http://www.oclc.org/research/projects/pmwg/surveyreport.pdf>.

④ OCLC. PREMIS Data Dictionary for Preservation Metadata(V3.0)[EB/OL]. [2017-12-30]. <http://www.loc.gov/standards/premis/>.

⑤ 牛金芳. 清华大学图书馆保存元数据方案[J]. 大学图书馆学报, 2003(2): 22-25.

⑥ 国家图书馆. 中文元数据方案[EB/OL]. [2018-01-05]. <http://www.doc88.com/p-79421253027.html>.

⑦ 王松林. 一个基于 OAIS 和 DC 的保存元数据方案[J]. 图书馆学刊, 2005(1): 1-3.

⑧ 国家科技图书文献中心等. 中国数字图书馆标准规范建设[EB/OL]. [2018-07-20]. <http://cdls.nstl.gov.cn/>.

⑨ LOC. Digital Repository Development Core Metadata Elements[EB/OL]. [2018-07-01]. <http://www.loc.gov/standards/metadata.html>.

综合性管理型元数据方案。主要有：a. 都柏林核心元数据计划(DCMI)^①的《DC 管理型元数据元素集》(2003)及其修订版(2006)；b. 丹麦编目中心^②(2004)的《数字资源管理型元数据方案》；c. 美国 Bryn Mawr、Haverford 和 Swarthmore 等高校联盟^③(2008)的“电子资源追踪管理系统”中定义的“数字资源管理型元数据框架”；d. 美国数字图书馆联盟^④(2012)的《电子资源管理：ERMI 数据元词典》中的“管理型元数据元素”。

专题性管理型元数据方案。主要有：a. 产权管理方面，国际版权组织^⑤(2002)的《开放数字产权语言》中定义的数字资源产权管理元数据；b. 用户隐私管理方面，W3C^⑥的P3P工作组(2006)的《隐私偏好平台规范》中设置的隐私保护政策元数据；c. 数字资源服务方面，IBM 和 Oracle^⑦(2011)的《Web 服务元数据交换》中设置的服务型元数据；d. 数字资源安全方面，Mozilla 和 Google^⑧(2012)的《数字资源内容安全方针》中定义的安全元数据；e. 信息政策方面，国际结构化信息标准促进组织 OASIS^⑨(2013)的《可扩展存取控制标记语言》中定义的信息政策元数据。

②国内研究现状。目前还未见独立的管理型元数据方案的报道。现有的方案大多是在综合性元数据方案中设置一些管理型元数据元素或模块，主要有：北京大学数字图书馆《中文元数据标准框架》^⑩中的管理型元数据(2002)、清华大学《建筑数字图书馆元数据体系》^⑪中的管理型元数据(2004)、CALIS《重点学科网络资源导航库的元数据规范》^⑫中的

① DCMI. The A-Core: Metadata about Content Metadata [EB/OL]. [2018-05-20]. <http://www.ifla.org/udt/dc8/note.htm>.

② The Danish Bibliographic Centre. Denmark's Electronic Research Library management metadata description[EB/OL]. [2017-10-11]. http://www.dbc.dk/top/top_general.html.

③ The Tri-College Consortium. Electronic Resources Tracking System (ERTS)[EB/OL]. [2017-11-20]. <http://www.haverford.edu/library/erts/>.

④ Digital Library Federation. Electronic Resources Management: ERM Data Meta Dictionary[EB/OL]. [2017-12-22]. <http://www.diglib.org/pubs/df102/ERMFINAL.pdf>.

⑤ International Copyright Organization. Open Digital Rights Language (ODRL)[EB/OL]. [2018-05-19]. <http://www.w3.org/TR/odrl/>.

⑥ W3C. The Platform for Privacy Preferences Specification [EB/OL]. [2018-07-14]. <http://www.w3.org/TR/P3P/>.

⑦ IBM, Oracle. Web Services Metadata Exchange[EB/OL]. [2018-05-29]. <http://www.w3.org/TR/ws-metadata-exchange/>.

⑧ Mozilla, Google. Content Security Policy[EB/OL]. [2018-08-11]. <http://www.w3.org/TR/CSP/>.

⑨ OASIS. eXtensible Access Control Markup Language[EB/OL]. [2018-07-21]. <http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.pdf>.

⑩ 肖珑等. 中文元数据标准框架及其应用[EB/OL]. [2018-10-25]. http://www.lib.sjtu.edu.cn/chinese/virtual_reference_desk/metadata_framework.pdf.

⑪ 孙晓菲. 高校专题数字图书馆建设简论——以哈佛大学和清华大学专题数字图书馆建设为例[J]. 图书馆论坛, 2007(2): 19-22.

⑫ 张惠君. “CALIS 重点学科网络资源导航库”的标准与规范建设[EB/OL]. [2017-10-01]. <http://www.calis.edu.cn/calisnew/images1/neikan/5/3-1.htm>.

管理型元数据^①(2006)、上海交通大学图书馆制定的《教学参考书元数据标准》^②中的管理型元数据(2009)、国家科技图书文献中心^③研制的《热点门户管理型元数据》(2011)等。目前,大型保存项目的元数据方案中大多设置有管理型元数据部分(如知网、维普等)。

1.2.2 有关数字保存的风险监控

据普查,目前为止,针对数字保存的风险监控方法未见报道。

风险监控的实质是对实现数字保存核心职责的威胁进行监视。本研究设计的风险监控方法是:首先识别数字保存的核心职责,其次针对每种核心职责识别可能的威胁,再次针对每种威胁找出包含的风险型元数据元素,最后通过检查各项元素的内容实现对每种威胁的监控。因此,该部分文献综述也分为上述四个部分。

(1) 数字保存的核心职责方面。

①国外研究现状。Waters 和 Garrett^④(1999)在《保存数字信息》中较早论述数字保存核心职责,并将其界定为维护数字信息的完整性和长期可访问性。之后,主要研究有两类:

核心职责的类型。主要有:a. OCLC 和 RLG 工作组^⑤(2001)在《基于 OAIS 的支撑数字对象保存的元数据框架》中将保存职责界定为维护数字对象的持久生存能力、可呈现能力和可理解性;b. CCSDS^⑥(2003)开发的《OAIS 参考模型》中认为核心职责是维护保存内容的稳定性、参考性、来源性、语义环境、可理解性和可访问性;c. PREMIS 工作组^⑦(2007)把核心职责归纳为维护数字对象的持久生存能力、可呈现能力、可理解性、真实性和可识别性;d. Vermaaten 等^⑧(2014)在《成功数字保存的威胁识别:基于 SPOT 模型的风险评估》中对上述保存职责进行了综合分析。

① CALIS. CALIS 特色库管理型元数据规范[EB/OL]. [2018-09-21]. <http://www.lib.whu.edu.cn/calis2/files/ocv1.doc>.

② 上海交通大学图书馆. 教学参考书元数据标准[EB/OL]. [2018-01-09]. <http://cdls.nstl.gov.cn/mt/blogs/2nd/archives/docs/CDLS-S05-001.pdf>.

③ 国家科技图书文献中心. 热点门户管理型元数据[EB/OL]. [2018-03-09]. <http://168.160.1.114/SPT-AboutUs.php#元数据标准规范>.

④ Waters D, Garrett J. Preserving Digital Information: Report of the Task Force on Archiving of Digital Information[EB/OL]. [2018-09-02]. <http://www.oclc.org/content/dam/research/activities/digpresstudy/final-report.pdf?urlm=161430>.

⑤ OCLC/RLG Working Group on Preservation Metadata. Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects[EB/OL]. [2018-04-11]. http://www.oclc.org/research/project/pm_framework.pdf.

⑥ Consultative Committee on Space Data Systems. Reference Model for an Open Archival Information System (OAIS)[EB/OL]. [2018-07-26]. <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf>.

⑦ PREMIS. Introduction and Supporting Materials from PREMIS Data Dictionary for Preservation Metadata[EB/OL]. [2018-08-26]. <http://www.loc.gov/standards/premis/v2/premis-report-2-1.pdf>.

⑧ Vermaaten S, et al. Identifying Responsibility to Successful Digital Preservation: the SPOT Model for Risk Assessment[J]. D-Lib Magazine, 2012, 18(9/10).

核心职责的实证分析。主要有: Bradley^①(2006)、Caplan^②(2008)、Dappert^③(2009)和 Fojtu^④(2013)分别对不同类型保存系统的核心职责进行了案例分析。

②国内研究现状。对数字保存核心职责的专门研究未见报道。为数不多的涉猎文献(张智雄^⑤, 臧国全^⑥, 杨淑萍^⑦等)出现在论述其他主题(如 OAIS)时附带说明这种职责, 且多为对国外的介绍。

(2)数字保存的威胁方面。

①国外研究现状。Conway^⑧(1996)在《数字世界的保存》中较早探讨数字保存威胁。之后, 主要研究有三类:

专用型威胁模型。主要有: 数字对象文件格式的威胁(Lawrence^⑨, 2000; Arms^⑩, 2003; Stanescu^⑪, 2005; Rog^⑫, 2007); 保存介质的威胁(Wright^⑬, 2009); 特定类型数字资源(如 Web 数字资源)的保存威胁(McGovern^⑭, 2004)。这些模型适合于相应领域的威胁识别, 尽管具有一定的互补性, 但无法替代综合型威胁模型。

综合型威胁模型。结构上有: 等级型威胁模型(Barateiro^⑮, 2009); 同位列表型威胁

① Bradley K. Digital Sustainability and Digital Repositories [EB/OL]. [2018-07-19]. http://www.valaconf.org.au/vala2006/papers2006/45_Bradley_Final.pdf.

② Caplan P. The Preservation of Digital Materials. Library Technology Reports, 2008, 44(2): 156-177.

③ Dappert A. Significance is in the Eye of the Stakeholder [EB/OL]. [2018-07-23]. <http://dl.acm.org/citation.cfm?id=1812838>.

④ Fojtu A. Tools for long-term preservation of digital documents [EB/OL]. [2018-07-30]. http://nrgl.techlib.cz/images/Evskp_2012_prispevek.pdf.

⑤ 张智雄: 数字保存是数字时代信息中心的一项重要战略[EB/OL]. [2018-07-30]. http://www.edu.cn/li_lun_yj_1652/20091124/t20091124_424016.shtml.

⑥ 臧国全. 数字保存的经济属性解析[J]. 图书情报工作, 2010(17): 32-35, 81.

⑦ 杨淑萍. 关于数字资源长期保存风险管理问题的探讨[J]. 图书馆学研究, 2007(7): 83-87.

⑧ Conway P. Preservation in the Digital World[EB/OL]. [2015-06-19]. <http://www.clir.org/pubs/reports/conway2/>.

⑨ Lawrence G W. Risk Management of Digital Information: A File Format Investigation[EB/OL]. [2018-05-23]. <http://www.clir.org/pubs/reports/pub00/contents.html>.

⑩ Arms C. Digital formats: Factors for Sustainability, Functionality, and Quality[EB/OL]. [2018-05-09]. <http://www.imaging.org/IST/store/physpub.cfm?seriesid=28&pubid=692>.

⑪ Stanescu A. Assessing the Durability of Formats in a Digital Preservation Environment: The INFORM Methodology[J]. OCLC Systems & Services, 2005, 21(1): 61-81.

⑫ Rog J. Evaluating File Formats for Long-Term Preservation [EB/OL]. [2018-06-23]. http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/KB_file_format_evaluation_method_27022008.pdf.

⑬ Wright R. The Significance of Storage in the "Cost of Risk" of Digital Preservation[J]. International Journal of Digital Curation, 2009, 4(3): 20-32.

⑭ McGovern N Y, et al. Virtual Remote Control[J]. D-Lib Magazine, 2004, 10(4).

⑮ Barateiro J. Addressing Digital Preservation: Proposals for New Perspectives[EB/OL]. [2018-03-09]. <http://cs.harding.edu/indp/papers/barateiro7.pdf>.

模型(Clifton^①, 2005; Rosenthal^②, 2005; Thomaz^③, 2006; PARSE. Insight^④, 2010); 网状型威胁模型(DCC^⑤, 2007)。另外, 有些通用威胁模型也描绘了威胁、数字对象、保存环境之间的关系(Dappert^⑥, 2009)。《可信任数字保存审查表》(OCLC/CRL^⑦, 2007)及其认证程序(CRL^⑧, 2009)也间接地展示了数字保存的威胁。由于这类模型的应用环境和目的不同, 导致它们在威胁的种类、数量及模型展现形式等方面存在差异。

数字保存威胁的实证研究。主要有: 对美国国会图书馆数字保存的威胁检查(Littman^⑨, 2007); 对大英图书馆数字保存介质的威胁评估(McLeod^⑩, 2008)。这类研究是对已有模型的实证分析, 有助于数字保存项目选择合适的威胁评估模型。

②国内研究现状。对数字保存威胁的专门研究未见报道。为数不多的国外情况介绍出现在论述其他主题的文献中(臧国全^⑪, 孔繁清^⑫, 王军^⑬等)。

(3)数字保存威胁与风险型元数据的对应方面。目前为止, 国内外均未见到完全切题的报道。仅发现的一项相关研究是Kool等^⑭(2014)对SPOT模型中6项保存威胁与PREMIS的《保存型元数据词典》的语义单元之间所做的一个对应实验。

(4)数字保存的风险监控方面。国内外在该方面的研究为:

实践上, 风险监控已经应用到了众多行业, 比如工程项目、金融证券、IT、自然环境、企业运营、产品质量、医疗卫生、信用体系、生产安全、行政管理, 等等。但是, 针

① Clifton G. Risk and the preservation management of digital collections [EB/OL]. [2018-05-20]. <http://archive.iffa.org/V1/4/news/ipnn36.pdf>.

② Rosenthal D S H. Requirements for Digital Preservation Systems[J]. D-Lib Magazine, 2005, 11(11).

③ Thomaz K P. Critical Factors for Digital Records Preservation[J]. Journal of Information, Information Technology, and Organizations, 2006 (1): 21-39.

④ PARSE. Insight consortium. Science Data Infrastructure Roadmap[EB/OL]. [2018-04-02]. http://www.parse-insight.eu/downloads/PARSE-Insight_D2-2_Roadmap.pdf.

⑤ DCC. Digital Repository Audit Method Based on Risk Assessment[EB/OL]. [2017-12-02]. <http://www.repositoryaudit.eu/>.

⑥ Dappert A. Report on the Conceptual Aspects of Preservation[EB/OL]. [2017-12-21]. http://www.planets-project.eu/docs/reports/Planets_PP2_D3_ReportOnPolicyAndStrategyModelsM36-Ext.pdf.

⑦ OCLC, CRL. Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist[EB/OL]. [2018-04-22]. <http://www.crl.edu/PDF/trac.pdf>.

⑧ CRL. Audit and Certification of Digital Archives[EB/OL]. [2018-02-29]. <http://www.crl.edu/content.asp?l1=13&l2=58&l3=142>.

⑨ Littman J. Actualized Preservation Threats[J]. D-Lib Magazine, 2007, 13(7/8).

⑩ McLeod R. Risk Assessment: Using a Risk-Based Approach to Prioritise Handheld Digital Information[EB/OL]. [2018-05-22]. http://www.bl.uk/ipres2008/presentations_day1/20_McLeod.pdf.

⑪ 臧国全, 焦克非. 论数字保存[J]. 现代情报, 2007(8): 50-52.

⑫ 孔繁清. 数字保存经济问题探讨[D]. 郑州: 郑州大学硕士学位论文, 2011.

⑬ 王军. 图书馆数字保存的现实思考[J]. 现代情报, 2006(2): 23-37.

⑭ Kool W, et al. Preservation Health Check: Monitoring Threats to Digital Repository Content[EB/OL]. [2018-08-01]. <http://www.oclc.org/research/publications/library/2014/oclcresearch-preservation-health-check-2014.html>.