

# 数字保存的风险检测<sup>\*</sup>

臧国全 李哲

**摘要** 选择 PREMIS《保存型元数据词典》作为证据基础对数字保存的风险进行检测。首先设计检测方法;其次实施保存型元数据与保存风险之间的映射,并对映射过程中的映射相关性、核心职责的覆盖程度、保存系统的独立性进行分析;再次设计检测实施方案,并对实施方案中的风险检测级别、编码、保存政策、保存事件进行分析;最后进行全风险检测和单风险检测实验,并对实验结果进行分析。研究表明,PREMIS《保存型元数据词典》可以用于保存风险的检测,但也存在一些不足,有必要设计数字保存的风险型元数据。

**关键词** 数字保存 保存风险 风险检测

**分类号** G250

**DOI** 10.16603/j.issn1002-1027.2017.06.008

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,检测风险应是数字保存的一项常规工作,也是规避和降低风险的基础。

## 1 检测方法的设计

### 1.1 设计思路

数字保存需要周期性(或实时)的风险检测,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,数字保存的风险检测可以通过检查数字对象与风险相关的属性来实现。因此,实现数字保存的风险检测需要两个基础元素:一是检测对象,是对“数字对象的与风险相关的属性”的描述和记录;二是检测项目,是对数字对象进行检测所需的风险识别列表。

在保存系统中,数字对象的属性描述都是通过元数据实现的。不同类型元数据的作用不同,设置的元数据元素也就不一样。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,采用的基本方法是识别实现每项核心职责的保存风险,进而采取相应措施规避和降低识别出的风险。目前,业已存在的保存型元数据方案有多个,其中 PREMIS 的《保存型元数据词典》(以下简称“《PREMIS 元数据词典》”)是应用最广的一个,且已经成为国际上的“事实标准”<sup>[1]</sup>。由于设计一个全新的元数据方案工作量巨大,因此本研究直接选用该词典。

目前业已存在一些数字保存的风险识别模型,但在对风险识别模型的基本要求(包括:风险概念界定的清晰性、风险专指度的合适性与一致性、风险的全面性、风险识别模型的易用性等)满足方面都存在不同程度的缺憾。为此,作者设计了一个《数字保存的风险识别模型》(以下简称“《风险识别模型》”),并直接选用该风险识别模型。

实施数字保存风险检测的思路为:(1)识别《PREMIS 元数据词典》中与数字对象保存风险相关的元数据(PREMIS 使用的是“语义单元”,为了便于理解,文中沿用传统称谓“元数据”);(2)根据《风险识别模型》中每种保存风险的内涵,建立保存风险与 PREMIS 元数据之间的映射关系;(3)基于映射关系,设计保存风险检测逻辑流程图;(4)基于逻辑流

<sup>\*</sup> 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)的研究成果之一。

通讯作者:臧国全,ORCID:0000-0002-9606-6455,zangguoquan@zzu.edu.cn。

程图,编制实现保存风险自动检测的程序;(5)选择一个保存系统,采集实验样本,运行检测程序,输出检测结果,并对结果进行分析。

### 1.2 选择《PREMIS 元数据词典》的原因

《PREMIS 元数据词典》是由 OCLC 和 RLG 于 2005 年在对全球主要数字保存的研究机构和保存项目进行调查和深度采访、全面考察保存型元数据的基础上并参考开放档案信息系统(Open Archival Information System, OAIS)制定的,且于 2008 年和 2015 年进行了两次修订。最新版本(V3.0)包括 4 个实体(对象、保存事件、行为主体、权利)、48 个容器型元数据、199 个元数据元素。

《PREMIS 元数据词典》将其设计的保存型元数据界定为“为了支撑数字保存过程,保存系统所使用的信息,尤其是为实现其核心职责所需的信息”。因此,保存和维护保存型元数据的目的是提高保存系统实现其核心职责的能力,实现方式是通过记录数字对象的主要特征、出处、保存过程中涉及的保存事件、行为主体、权利等相关信息,使每件数字对象都拥有一组保存型元数据对其进行描述,从而使数字对象具有“自我记录性(Self-documenting)”,这种“自我记录性”特征能够维持数字对象本身的真实性、完整性、可理解性和可呈现性的实现。另外,保存型元数据记录的信息常被视为“应急信息”,对其保存和维护以备未来之需,因为长期保存过程充满不确定性,这种“应急信息”可以帮助识别和规避由不确定性导致的潜在风险。因此,《PREMIS 元数据词典》可以作为数字对象的保存风险识别的证据基础,与本研究思路吻合。

### 1.3 选择《风险识别模型》的原因

《风险识别模型》首先识别数字保存的核心职责,然后针对每项核心职责识别保存系统中影响该核心职责实现的各种风险。核心职责和包含的风险数量为:可用性(4 种风险)、可识别性(3 种风险)、持续完整性(5 种风险)、可呈现性(5 种风险)、可理解性(4 种风险)和真实性(3 种风险)。

《风险识别模型》中的数字保存核心职责是基于对数字保存领域的理论与实践全面考察的结果,在业界已有很强的共识。另外,该模型和《PREMIS 元数据词典》都是基于保存系统生命周期设计的,包括的阶段有:数字对象的获取、保存、维护和传播使用,两者的兼容性较好。最后,该模型的所有风险都是

仅针对保存的数字对象,而不针对保存系统本身(如保存系统经济方面的风险和法律方面的风险),这与《PREMIS 元数据词典》也一致。

## 2 保存型元数据与保存风险之间的映射

### 2.1 映射方案

针对《PREMIS 元数据词典》的对象实体,映射元数据包括两部分:一是针对容器型元数据,将其包含的所有下级元素归并一起,作为一个单元整体参加映射;二是针对非容器型元数据元素,直接参加映射。上述第一种情况有一个例外,即对象特征(Object Characteristics)容器型元数据,因为该容器型元数据包含的各个元素都是用来描述数字对象的不同特征,因此将该容器型元数据中的每个元素都作为一个独立的单元参加映射。

针对《PREMIS 元数据词典》中其他实体(保存事件、行为主体、权利),各自仅作为一个单元整体参加映射。针对《风险识别模型》,映射也仅在核心职责的层面上实施。

采用上述映射方案的原因有二:一是减少参加映射的元数据和保存风险的数量,降低风险检测实验的复杂性。二是如果一个元数据元素与一种风险相关,那么该元素所属的容器型元数据也一定与该风险相关;反过来,如果一个风险与一个容器型元数据相关,那么该风险所属的核心职责也一定与容器型元数据相关,这样可以保证风险检测实验结果的真实性,但没有精确到具体的风险和具体的元数据元素。

### 2.2 映射结果及分析

基于上述设计的映射方案进行映射操作的结果如表 1 所示。

表 1 《风险识别模型》中核心职责与《PREMIS 元数据词典》的映射结果

《风险识别模型》 核心职责	《PREMIS 元数据词典》 映射元数据数量
可用性	18
可识别性	21
持续完整性	10
可呈现性	16
可理解性	17
真实性	13

### 2.2.1 映射的相关性

保存风险与元数据之间相关性的判断没有一个严格清晰的标准,有些是直接相关,而有些是间接相关,但针对后者无法度量其相关的程度。在映射实践中,对间接相关没有考虑相关程度,都进行了映射处理。可以根据风险检测的结果,分析这些间接相关是否保留。

基于对相关性的上述处理方式,再加上元数据和保存风险的本身特征,导致映射结果并不是一一对应关系。一个元数据可能映射到多个风险,而这多个风险可能隶属于不同的核心职责,因此一个元数据可能映射到多个核心职责。反过来,一个风险可能对应于多个元数据元素,而这多个元素可能隶属于不同的容器型元数据,并且一个核心职责又包括多个风险,所以一个核心职责可能映射到多个元数据。

### 2.2.2 核心职责的覆盖程度

每个核心职责在《PREMIS 元数据词典》中的覆盖程度取决于两个因素:一是相关元数据的数量;二是元数据与核心职责的风险之间的针对性。因此,相关元数据的数量多并不意味着核心职责的覆盖程度高,比如,核心职责“持续完整性”映射的元数据数量最少(10个),但覆盖程度相对较高,因为这些元数据所记录的信息都是针对持续完整性的风险。相反,虽然核心职责“可理解性”对应的元数据数量比“持续完整性”多,但是“可理解性”的覆盖程度较低,因为映射到“可理解性”的元数据虽然相关,但针对性差,大多数元数据对识别“可理解性”相关风险的作用小,因为《PREMIS 元数据词典》没有定义一组元数据提供帮助理解或解释所存储数字对象内容的信息,虽然这些信息可以存储在其他数字对象中,并且保存系统可以将该数字对象与包含该类信息的其他数字对象进行链接,但这种方式毕竟不是直接的。

### 2.2.3 保存系统的独立性

《PREMIS 元数据词典》的设计基于一个假设:保存系统是一个独立系统,且控制所有的保存过程。但是,现行实践中不乏将一部分保存活动外包给第三方的实例。在这种情况下,整个保存过程将分布在多个保存机构中,可能会导致整个保存系统面临的风险不同于《风险识别模型》中界定的风险,并由此导致对作为风险检测证据基础的保存型元数据需求不同。

比如,对“可识别性”风险检测造成的影响。假如一个保存系统是由多个子系统集合而成(如数据库子系统、文件子系统、Web 服务子系统等),基于一个子系统(如数据库)自动生成的数字对象唯一标识符,使用范围仅局限于该子系统,也仅支持该系统内的保存活动(如检索、复制、删除等),但无法支持跨越两个及其以上子系统的保存活动,也不适用于针对保存系统外包的第三方保存服务。因此,用于确保数字对象长久可识别性的唯一标识符不应该是保存系统中某一个子系统的属性,应该是整个保存系统的一个属性,而上述保存系统的分布式结构所产生的仅适用于某一子系统的数字对象唯一标识符,给“可识别性”风险的检测带来困难。

再比如,对“持续完整性”风险带来的影响。“持续完整性”的风险之一是比特保存,这是保存系统对数字对象存储的一项功能,但该项功能经常被外包给保存机构内部的一个 IT 部门或者外部的第三方。在这种情况下,保存系统就无法控制与比特保存相关的保存型元数据的信息采集和维护,导致“持续完整性”的风险检测困难。

为了充分反映数字保存的分布式结构,需要对《PREMIS 元数据词典》和《风险识别模型》做一些改造,以满足分布式结构保存系统的风险检测需要。

## 3 保存风险的检测实施

### 3.1 逻辑流程图

鉴于篇幅,文章仅以“持续完整性”的风险检测为例,因为该类风险在《PREMIS 元数据词典》中的覆盖程度相对较高。

“持续完整性”指构成数字对象的比特流持续存在且没有被破坏,并处于可使用、可被操作的状态,这些比特流可以从所保存的存储介质中被检索出来并进行相应的操作。《风险识别模型》中“持续完整性”的风险有:(1)数字对象的不合理或过失的操作或存储(如不适宜的保存条件);(2)存储介质超出有效期;(3)用于访问存储介质所需的设备已无法获取;(4)存储介质或比特流序列的蓄意破坏(如存储介质的有意破坏或被盗,计算机病毒);(5)由于软硬件或者操作人员的失误导致对存储介质和比特流序列的无意破坏。总的来说,对数字对象“持续完整性”的主要风险来自存储介质的管理、刷新方针、迁移政策和数据安全方针。

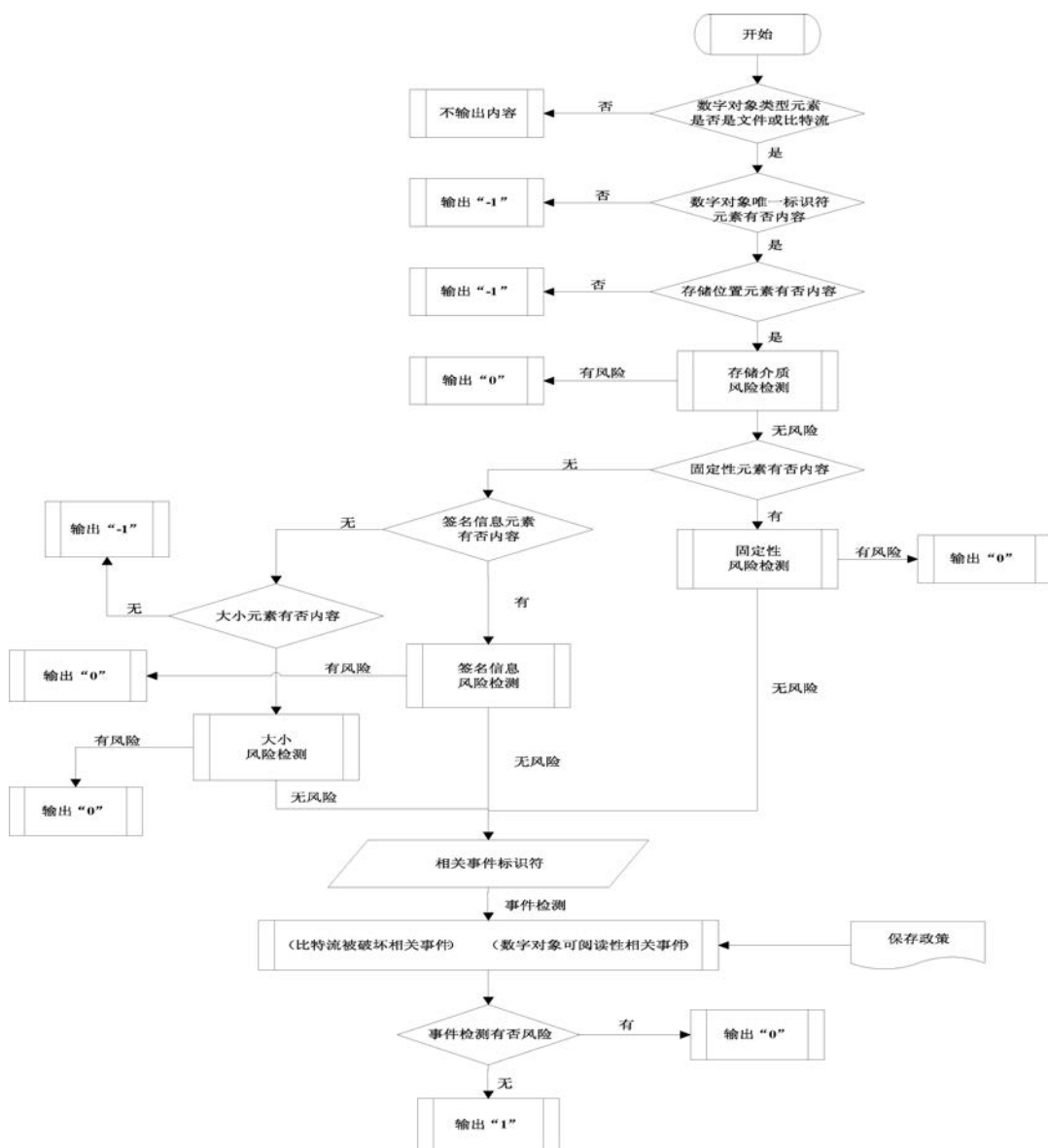


图1 “持续完整性”风险检测的逻辑流程图框架

“持续完整性”风险检测的逻辑流程图框架如图1所示。针对每件数字对象,流程图输出一个用数字表示的检测结果,其中:“-1”表示对应元数据内容缺失;“0”表示检测出风险;“1”表示未检测出风险。

在数字保存实践中,数字对象有4类:文件对象、比特流对象、知识实体和表现。其中,知识实体是一件独立的知识作品,形式上包含一组文件,比如一本书(是一个知识实体)的每一页都是单个独立的PDF文件。表现是知识实体的展现形式,比如一篇论文(是一个知识实体)共12页,每页为一个TIFF

图像,加上一个结构化元数据的XML文件显示页面顺序,这13个文件构成了一个表现。因此,知识实体和表现都为多个文件对象的逻辑组合,但持续完整性风险产生的数据单元是单个文件或比特流,所以图1不适用于知识实体和表现的检查,形式上,这两种数字对象的检查结果将不输出内容。

“持续完整性”的风险检测,除了数字对象范畴(Object Category)和数字对象标识符(Object Identifier)之外,还需两类元数据:数字对象存储(Object Storage)和数字对象的特征。

数字对象存储包含存储位置(Content Loca-



tion)和存储介质(Storage Medium)两个元数据。需要检查数字对象有否存储介质,但无须检查具体的存储位置,因为“持续完整性”与具体存储位置无关,但与存储介质直接相关。不同存储介质的使用寿命差别较大,保存系统需根据介质使用寿命对存储的数字对象进行迁移,以避免存储介质的过期导致数字对象二进制数据流的破坏和无法完整读取呈现,因此“持续完整性”的风险检测需知晓数字对象的存储介质并由此获取其使用寿命。使用寿命可由相关元数据记录,也可从其他来源中获取,比如,从存储子系统记录的存储介质生产企业、出厂日期等信息中推算,因为任何一种存储介质都有一个大概的使用周期。

数字对象的特征包含多个元数据,其中只有数字对象的固定性(Fixity)和“持续完整性”风险检测密切相关。然而,如果保存系统没有记录 Fixity 元数据,可选择数字对象的签名信息(Signature Information)元数据检测数字对象的比特流是否被破坏,这是替代固定性检查的一种较好方式;如果连签名信息也没有记录,还可使用数字对象的大小(Size)元数据粗略判断数字对象是否被改变,但准确度不高。

接下来,通过数字对象的保存事件标识符(Linking Event Identifier),识别并检查相关的保存事件,在长期保存过程中保存系统执行这些事件的目的是防止或检查“持续完整性”的风险。事件有两类:一是与比特流顺序被破坏的相关保存事件;二是与数字对象可读性相关的保存事件。前者是针对数字对象比特流的被破坏进行检测和规避的活动(如固定性检查、病毒检测、备份和存储活动等),后者是针对保存介质的风险进行检测和规避的活动(如存储介质的刷新、更换等)。这两类事件的界限不太清晰,可能还存在重复(如都有备份事件),但不影响检测结果。

### 3.2 检测实施的相关问题

#### 3.2.1 风险检测级别

检测实施前,需考虑在哪个层面上实施:在单件数字对象层面?一类数字对象层面?还是在整个保存系统层面?检测级别的选择取决于以下两个因素:

(1)风险的性质。“真实性”风险检测需在单件数字对象层面进行,因为同一类型或整个保存系统

中不同数字对象的真实性可能不一样;“可呈现性”风险检测除了可在单件数字对象层面上实施外,还可在一类数字对象层面上进行,因为一类数字对象使用的软硬件环境(是“可呈现性”风险的主要来源)是相同的;如果数字对象的唯一标识符是由保存系统产生,或由外部注册中心提供且可以共享,那么“可识别性”风险检测(唯一标识符是主要的风险来源)还可在保存系统层面上进行。

(2)保存系统的风险检测需求。“可呈现性”风险可在一类数字对象层面上进行,也可以在单件数字对象层面上进行,不同保存系统可根据不同的风险检测需求来选择。因为风险检测需求有多个层面,所以用于不同层面的风险检测逻辑图也有所区别,比如,检测对象是某种格式的一类数字资源,那么在逻辑图中就无需判断数字资源的格式。

#### 3.2.2 编码

元数据的赋值方式给风险的自动检测带来两个挑战:一是 PREMIS 中的一些元数据采用控制词表赋值,但仅推荐可采用的控制词表,没做统一的规范性要求,这样,不同保存系统可能采用不同控制词表,导致含义相同的赋值内容所使用的词汇或表达方式不一样;二是有些元数据采用自由文本式赋值,也会导致同样的问题。解决方法是对赋值内容进行编码,对含义相同的内容给一个编码。编码方式应具有通用性,如果存在相应国际标准应予以采用。但目前 PREMIS 没有涉足该项工作,增加了风险检测程序编制的工作量,且检测程序在不同保存系统之间也难以通用。

#### 3.2.3 保存政策

映射结果表明,《PREMIS 元数据词典》支持风险检测的一个缺陷是缺乏用于描述保存政策的元数据,而这类元数据对于风险检测是至关重要的。比如,数字对象存储介质的刷新频率是一项保存政策,记录该保存政策对“持续完整性”核心职责的“保存介质不可用”风险的检测起重要作用。实际上,任何一个保存系统在数字对象长期保存过程中都需要一系列保存政策,但目前《PREMIS 元数据词典》没有提供元数据给予记录(除保存级别元数据之外)。如果一些政策仅适用于某些类型的数字对象,而不适用于保存的所有数字对象,这样不同类型的数字对象就要采用不同的保存政策,这些政策记录的缺失使得相应的风险检测愈加困难。

## 3.2.4 保存事件

(1)保存事件的定义与归类。《PREMIS 元数据词典》对保存事件的定义是开放的,没有预先定义事件的类型,保存系统可根据需求具体定义。这对保存风险检测的实施是一个挑战,因为不仅要对各种事件进行定义,而且需将性质相同或相近但名称不同的事件搜集归类(如备份事件和复制事件),以避免保存系统已经执行了该种事件但由于事件名称不同而导致风险检测结果的误判。实现这种归类的一种有效方法是参考和分析 IT 领域中数据存储实践的保存事件以及与数据完整性和可使用性相关的保存事件,因为这些类型的事件与持续完整性的风险关联度较大。

(2)保存事件的来源。一部分保存事件是由保存政策决定的,比如,保存政策要求存储介质的刷新周期为半年,那么就需要一个实施存储介质刷新的保存事件。因此,保存系统应该制定各种保存政策,但 PREMIS 没有设计元数据记录保存政策。为此,本研究的风险检测需从外部获取保存政策信息。

(3)安全性事件不予考虑。安全性事件也与持续完整性相关,密码泄露、黑客入侵和过失操作等都会对数字对象的完整性、机密性和可用性造成风险,执行相应的安全性事件可以规避和降低这些风险,通过对安全性事件实施情况的检查可以实现对这类风险的检测。但安全性事件风险是针对整个保存系统,文章的“持续完整性”风险检测是针对具体的数字对象,所以没有考虑安全性事件的风险。

## 4 实验

## 4.1 数据采集

数字对象样本来源选取 CNKI 的《中国期刊全文数据库》,样本集为 1 万件数字对象。

数字对象样本采集方法采用分层随机抽样法,使各类数字对象以接近的概率被抽样,而后通过套录形成有代表性的数字对象样本集。针对 CNKI 的《中国期刊全文数据库》,首先,依据学科、年代区间等分类标准,将数字对象划分为若干个层次;然后,根据各层次单元的数字对象数量与《中国期刊全文数据库》数字对象总数量的比例,依据设定的抽取样本总量,确定从各层次单元中抽取样本的数量;最后,基于无重复抽样的简单随机抽取法从各层次单元中抽取样本。如,依据上述方法计算出层次单元

为“2015 年图书馆学的期刊论文”的样本抽取量为 10 篇,选择《中国知网》的文献类型为“期刊”,在高级检索界面中输入“中图法分类号=G250”且“年代=2015”,获得检索结果为 3610 篇,而后在 1 到 3610 之间随机生成 10 个不重复的数,套录该 10 个数对应的检索结果数字对象,获得该层次单元的样本。

保存型元数据内容的样本采集主要针对上述采集的数字对象,纯粹为了研究的目的,由 CNKI 帮助提供相应保存型元数据和描述型元数据内容。将保存型元数据转换为《PREMIS 元数据词典》的格式,然后针对转换后为空的保存型元数据元素,若描述型元数据对应元素有内容,则直接套录。上述过程由代码实现,但需人工进行干预,比如需对名称与《PREMIS 元数据词典》不同但含义相同的元素赋值给 PREMIS 对应的元数据。

## 4.2 实验结果

基于风险检测的逻辑流程图(图 1 仅为整个流程图的很小一部分),编制“数字保存风险检测”程序,针对上述采集的实验数据样本集运行该检测程序,输出检测结果。实验的相关界定为:检测级别为单件数字对象层面,实验用的元数据内容进行规范编码,定义必要的保存政策(如磁性存储介质刷新频率为 30 天),对保存事件名称进行规范化处理。另外,为了简化结果,将图 1 中输出为“-1”(表示元数据内容缺失)和“0”(表示检测出风险)聚类,同视为检测出风险。

## 4.2.1 全风险检测

全风险检测结果如图 2 所示,横轴代表风险类型,纵轴代表存在风险的数字对象数量(检测的数字对象数量为 1 万件)。

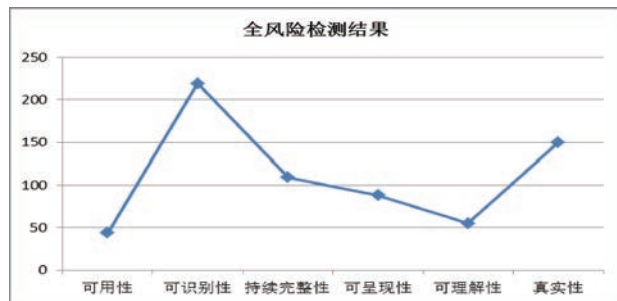


图 2 全风险检测结果

全风险检测结果显示,最大风险来自数字对象的可识别性,可能原因是元数据没有被充足抓取或良好维护,或虽被抓取但元数据与其所描述的数字

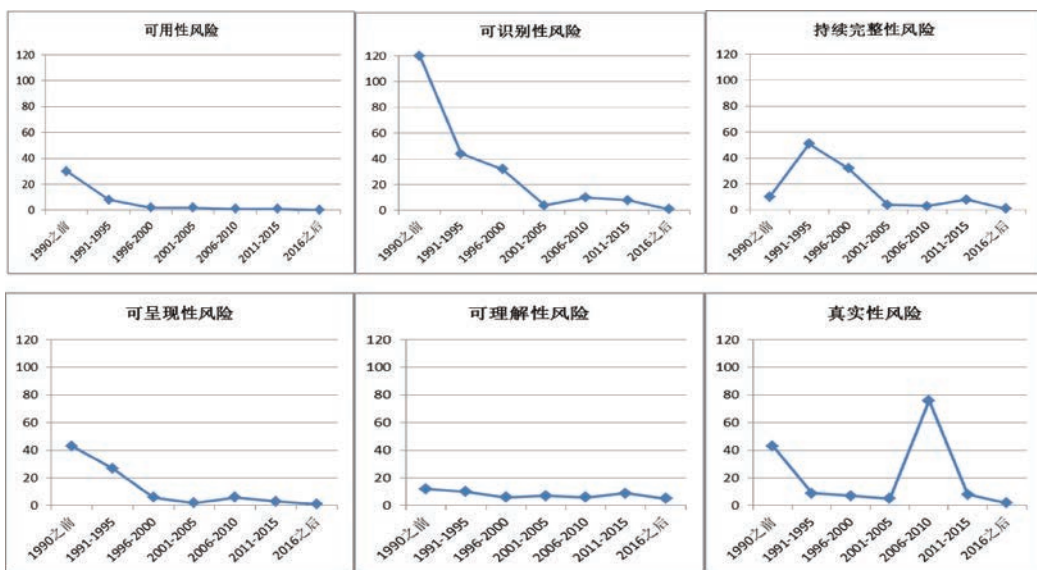


图3 单风险检测结果

对象之间的链接没有被抓取或维护。其次是数字对象的真实性风险,可能原因在于长期保存过程中,数字对象发生了改变(可能是合理改变、恶意修改或者错误改变),但合理改变部分没有被完整记录。再次是持续完整性风险,可能原因是没有执行存储介质的刷新操作、存储介质超出有效期、由于病毒或操作失误等因素导致比特流序列破坏等。其他类型的风险较小。

#### 4.2.2 单风险检测

针对每类风险分年代区间进行检测,结果见图3,横轴代表年代区间,纵轴代表存在对应风险的数字对象数量(检测的数字对象数量为1万件)。

可用性风险主要分布在1995年之前,这部分数字对象完全由数字转化生成,可能原因是转换质量不理想。

可识别性风险总体最差,主要分布在2000年之前,这部分数字对象大多由数字转化生成,元数据需要人工赋值,可能原因是这部分的元数据描述充分性较差。

持续完整性风险主要分布在1991—2000年,可能原因是这部分数字对象的存储介质维护不到位。

可呈现性风险主要分布也在1995年之前,同样由于这部分数字对象完全来自数字转化,在转化过程中数字对象的主要特征抓取效果不理想,一些体现数字对象重要特征的细节没有被有效抓取。

可理解性风险整体情况最好,但存在些许风险,可能原因是保存系统没有获取和保存用户群体所需的充足的附加信息。

真实性风险主要分布在1995年之前和2006—2010年,可能原因是该两个时间段的数字对象在保存过程中改变较多,但缺乏详细记录。

风险检测的目的是为保存系统的维护提供依据。根据上述实验,保存风险主要发生在2000年之前的数字对象,保存系统需要加强对这部分数字对象的质量、元数据描述及其与数字对象之间的链接、保存事件导致数字对象变化的记录等方面的维护。

## 5 结语

实验表明,《PREMIS元数据词典》可以作为证据基础支持数字保存的风险检测。但是,存在下述问题:(1)缺乏保存政策方面的元数据,而这类元数据不仅关系到保存事件的定义,而且还直接关系到风险的判断;(2)虽然有保存事件元数据,但缺乏具体保存事件的清晰定义,导致不同保存系统可能定义不同的保存事件,难以建立通用的风险检测算法;(3)元数据赋值的欠规范化,不仅体现在采用控制词表进行赋值的元数据上,而且也体现在采用自由文本进行赋值的元数据上,导致风险检测软件运行的困难;(4)缺乏对分布式保存系统的支持,这与现行保存系统的实践不完全一致。

因此,有必要建立针对数字保存风险检测的风险型元数据,以弥补上述保存型元数据的不足。风险型元数据的建立可以采用两个步骤:一是根据全面风险管理理论,从数字对象、保存事件、行为主体、



保存政策、知识产权、保存系统等6个角度分析数字保存风险的产生因素,分别建立6种类型的风险型元数据;二是参考已有的保存型元数据、管理型元数据、描述型元数据等,并在征询专家(尤其是数字保存的实践专家)的意见和面向公众讨论的基础上,完善风险型元数据。

## 参考文献

1 PREMIS Editorial Committee. PREMIS Data Dictionary for

Preservation Metadata (version 3.0) [EB/OL]. [2016-04-24]. <http://www.loc.gov/standards/premis/>.

2 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]. [2016-03-02]. [http://www.oclc.org/research/activities/past/orprojects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/activities/past/orprojects/pmwg/pm_framework.pdf).

作者单位:郑州大学信息管理学院,郑州,450001

郑州大学公共管理学院,郑州,450001

收稿日期:2017年1月12日

# Risk Check for Digital Preservation

Zang Guoquan Li Zhe

**Abstract:** In this paper, the PREMIS Data Dictionary is chosen as the evidence base to support risk check. The following four steps are used to check risk: (1) designing the check method; (2) implementing the mapping between the preservation metadata and preservation risks, and analyzing the related issues including the correlation between the both sides of the mapping, the coverage of core responsibility and the independence of the preservation system; (3) designing and testing the implementation, and analyzing the related issues including the level of risk check, encoding, preservation policies and preservation events; (4) testing and analyzing the results, including full risk check and single risk check. The research result shows that although the PREMIS Data Dictionary can be used to check the risks of the digital preservation, it does have some shortcomings. It is necessary to design the risk metadata for digital preservation.

**Keywords:** Digital Preservation; Preservation Risks; Risk Check

## 《中华人民共和国公共图书馆法》出台,将于2018年1月1日起施行

十二届全国人大常委会第三十次会议2017年11月4日表决通过了《中华人民共和国公共图书馆法》(以下简称“公共图书馆法”)。这是党的十九大之后出台的第一部文化方面的法律,也是公共文化领域继公共文化服务保障法之后的又一部重要法律,对于进一步健全我国文化法律制度、促进公共图书馆事业发展、保障人民群众基本文化权益具有重要意义。

公共图书馆法将于2018年1月1日起施行,共6章55条,对公共图书馆的设立、运行、服务以及相关法律责任等分别作了详细规定。

出台一部立得住、行得通、切实管用的公共图书馆法,是社会各界的夙愿。这部法律从酝酿到出台,历时十几年时间,得到了党中央、全国人大常委会、国务院的高度重视,经过了多个部门反复论证、深入调研、广泛征求各方意见,回应了社会关切,凝聚了社会各界的智慧和共识。

公共图书馆是传承人类文明、传播先进文化、开展社会教育的重要场所,是公共文化服务体系的重要组成部分,对满足公民精神文化需求、提高公民科学文化素质和社会文明程度有重要作用。公共图书馆法规定,本法所称公共图书馆,是指向社会公众免费开放,收集、整理、保存文献信息并提供查询、借阅及相关服务,开展社会教育的公共文化设施。明确公共图书馆应当将推动、引导、服务全民阅读作为重要任务。

法律明确了公共图书馆事业的发展方向,提出发挥公共图书馆功能,保障公民基本文化权益,坚定文化自信,必须坚持社会主义先进文化前进方向,坚持以人民为中心,坚持以社会主义核心价值观为引领,传承发展中华优秀传统文化,继承革命文化,发展社会主义先进文化。

法律规定县级以上人民政府应当将公共图书馆事业纳入本级国民经济和社会发展规划,将公共图书馆建设纳入城乡规划和土地利用总体规划,加大对政府设立的公共图书馆的投入,将所需经费列入本级政府预算,并及时、足额拨付;国家扶持革命老区、民族地区、边疆地区和贫困地区公共图书馆事业的发展。

(来源于《人民日报》2017年11月5日05版,记者郑海鸥报道)