

●张正强

电子文件管理元数据中时间元素的语义结构研究*

摘 要 电子文件管理元数据中,时间元素存在4种语义结构:时间元素的前端模式语义结构、后端模式语义结构、连续体模式的语义结构、分面语义结构。建立电子文件管理元数据标准时,最好以连续体模型为依据来建构,同时采用分面语义结构。表6。图3。参考文献5。

关键词 电子文件管理 元数据 时间元素 语义结构

分类号 G250.76

ABSTRACT In the metadata for the management of electronic documents, date elements have four semantic structures. The author thinks that when we establish a metadata standard for the management of electronic documents, we had better use the continuum-model and the facet semantic structure. 6 tabs. 3 figs. 5 refs.

KEY WORDS Electronic document management. Metadata. Date element. Semantic structure.

CLASS NUMBER G250.76

时间元素是元数据集中一个核心的元素,也是任何元数据标准中都不缺少的元素之一。在电子文件管理元数据中,时间元素也具有如此重要的地位。但是电子文件管理元数据中的时间元素在其语义结构上却要比通用元数据如“都柏林元数据集”中的时间元素的结构更复杂,而这种复杂性又是由电子文件管理所需历经其整个生命周期的性质所决定的,因此研究、理解电子文件管理元数据中的时间元素,必须研究、理解电子文件的生命周期,并从中得出规律性的认识,对于更科学地建立我国电子文件管理元数据

标准将是十分有益的。

1 电子文件的生命周期与通用元数据集中时间元素的语义结构

在著名的《都柏林核心元数据集》中和以此为基础制定的国际标准《ISO 15836:2003(E)信息和文献——都柏林核心元数据集》中^[1],时间元素被定义为“与资源生命周期中某一个事件相关的时间”。由于这两个元数据集是国际上跨学科、跨专业领域的通用标准,因此在这个定义中的“资源”也就具有高度

- A Conceptual Model of Service Quality and Its Implications for Future Research. Journal of Marketing, 1985, 49(4)
- 2 Nitecki. Changing the Concept and Measure of Service Quality in Academic Libraries. Journal of Academic Librarianship, 1996, 22
- 3 Thompson, B., & Cook, C. Stability of the reliability of LibQUALTM scores: A “Reliability Generalization” meta-analysis study. Educational and Psychological Measurement, 2002, 62
- 4 Thompson, B., Cook, C., & Thompson, R. L. Reliability and structure of LibQUALTM scores. portal: Libraries and the Academy, 2002, 2
- 5 <http://www.libqual.org/>
- 6 Brinley Franklin and Terry Plum. Library usage patterns in the electronic information environment. Information Research, 2004, 9(4)
- 7 New Ways of Listening to Library Users: Tools for Measuring Service quality. <http://www.arl.org/libqual/events/listen/index2.html>
- 8 钱佳平. 基于差距分析的大学图书馆服务质量评价体系建构: 研究与实践. 北京大学硕士学位论文, 2004
- 9 朱德全. 现代教育统计与测评技术. 重庆: 西南师范大学出版社, 1998
- 钱佳平 浙江大学图书馆副研究馆员, 管理学硕士。通信地址: 杭州市。邮编 310027。
- 刘兹恒 北京大学信息管理系教授。通信地址: 北京。邮编 100871。
- (来稿时间: 2005-04-25)

* 本文为国家社科基金重点资助项目“基于XML电子文件管理元数据标准研究”系列研究成果(项目批准编号04ATQ002)。

的概括性,而这个“资源”在电子文件管理领域,它就被具体化为“电子文件”,在电子文件管理元数据中的时间元素实际上就是“与电子文件生命周期中某一个事件相关的时间”。

对电子文件生命周期有许多描述类型,在此我们先以传统经典的模型来描述。经典的模型认为:电子文件生命周期分为3个阶段:第一阶段是电子文件的创建阶段,第二阶段是电子文件的现行阶段,第三阶段是电子文件的非现行阶段即归档文件保管利用阶段,其具体的模型见图1。

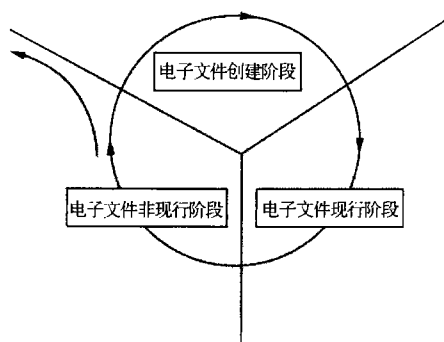


图1 电子文件的生命周期

从图1可以看出,在电子文件生命周期的3个阶段中,都可能发生与某一个事件相关的时间。如在电子文件的创建阶段,就有电子文件创建的事件,与此相应的就有文件创建的时间;在电子文件的现行阶段,就有电子文件有效期被截止的事件,与此相应的就有截止时间;而在电子文件的非现行阶段,就有归档文件被处置的事件,与此相应的就有归档文件被处置的时间,等等。

在通用元数据《都柏林核心元数据集》和《ISO 15836:2003(E)信息和文献——都柏林核心元数据集》中,时间元素都只设了1个元素,而没有任何下位子元素,其语义结构非常简洁明了。

但是尽管在《都柏林核心元数据集》时间元素没有下位子元素,但是在《都柏林元素集》(即 DCMI Metadata terms)中对时间元素还是列出了它的8个下位子元素,它们分别是:创建时间、修改时间、发布时间、可获取时间、生效时间、版权时间、接收时间和提交时间。将这8个下位子元素与《都柏林核心元数据集》中的时间元素结合起来就构成了如表1的语义结构。

表1 《都柏林元素集》时间元素语义结构

上位元素	下位子元素——元素限定词
时间	创建时间
	修改时间
	发布时间
	可获取时间
	生效时间
	版权时间
	接收时间
	提交时间

如果我们结合电子文件生命周期来看,作为时间元素的这8个下位子元素所表示的时间正好都是发生在电子文件生命周期的文件的创建阶段,见图2所示。

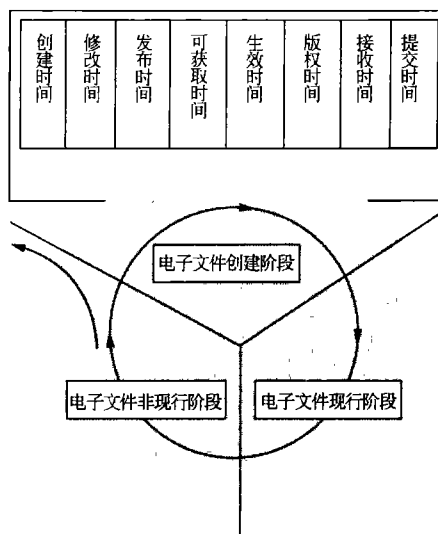


图2 《都柏林元素集》时间元素的下位子元素与电子文件生命周期的对应

由于都柏林元数据中的时间元素及其下位子元素与电子文件生命周期中的电子文件创建阶段对应,所以它也适用于电子文件管理中对相关事件的时间的描述,具有很强的通用性与适用性。也正是因为这种通用性与适用性,所以在电子文件管理元数据标准的建立中,其时间元素必须与都柏林元数据中的时间元素相兼容,实际上,这种兼容已经在电子文件管理元数据中时间元素的前端模式语义结构中得到了应用。

2 电子文件管理元数据中时间元素的前端模式语义结构

在国际电子文件管理领域,一般把电子文件生命周期中的电子文件创建阶段称为“前端”,而把电子文件生命周期中的电子文件非现行阶段称为“后端”。相应于这两端就产生了电子文件管理元数据时间元素的前端模式语义结构与后端模式语义结构这两种结构。

时间元素的前端模式语义结构,其特点是:“时间”作为一个上位元素,而它的下位子元素主要是由与电子文件生命周期中电子文件创建阶段和电子文件现行阶段中事件相关的时间构成,这种语义结构主要体现在英国颁布的《电子政务元数据标准》中^[2]。在该标准中,时间元素的语义结构是由 14 个作为元素限定词的下位子元素构成的,比《都柏林元素集》中的时间元素的下位子元素多了 9 个,但是这 14 个下位子元素都是与电子文件生命周期的“前端”——电子文件创建阶段和电子文件现行阶段中的事件相关的时间。其具体语义结构及与电子文件生命周期相对应的结构示意图如表 2。

表 2 英国《电子政务元数据标准》时间元素的语义结构与电子文件生命周期的对应

上位元素	下位子元素——元素限定词	文件生命周期
时间	1. 创建时间	电子文件创建阶段
	2. 修改时间	
	3. 发布时间	
	4. 可获取时间	
	5. 生效时间	
	6. 接收时间	
	7. 提交时间	
	8. 版权时间	
	9. 获得时间	
	10. 宣布时间	
	11. 更新时间	电子文件现行阶段
	12. 下一版本时间	
	13. 截止时间	
	14. 取消时间	
	无	电子文件非现行阶段

从表 2 时间元素的语义结构可以看出,其中 8 个下位子元素(1~8)都是《都柏林元素集》中的时间元素。这说明《都柏林元素集》中的时间元素在该结构中都得到了很好继承与兼容。而且,在以上时间元素的语义结构中,时间元素的 14 个下位子元素中有 10 个是集中在电子文件生命周期的电子文件创建阶段,由此可以看出,该时间元素的语义的重心是在电子文件生命周期的“前端”,至于文件生命周期的“后端”,在时间元素中没有设置任何下位子元素来反映它。

电子文件管理元数据时间元素的这种前端模式的语义结构,其优势主要有两点:

(1)通过时间元素下位子元素的向下扩展,可以详尽地反映电子文件生命周期中电子文件创建阶段,即“前端”中与事件有关的各种时间。

(2)通过时间元素下位子元素的向上兼容,可以实现时间元素纵向上的语义互操作。

但也有局限性,主要表现为:

(1)该模式的时间元素的下位子元素主要反映的是电子文件生命周期的“前端”,因此,在时间元素下没有下位子元素来反映电子文件生命周期中电子文件非现行阶段的相关事件的时间,从而也就造成这一阶段的下位子元素的“空白”。

(2)当电子文件处于生命周期的“后端”时,时间元素难以向后扩展对处于该阶段的电子文件的进行时间描述。

3 电子文件管理元数据中时间元素的后端模式语义结构

电子文件管理元数据中时间元素的后端模式语义结构,其特点是:“时间”作为一个上位元素,下位子元素却主要是由与电子文件生命周期“后端”中事件相关的时间构成,这种结构主要体现在南非颁布的《电子文件核心元数据》标准中^[13]。在该标准中,时间元素的语义结构由 9 个下位子元素构成,而且这 9 个下位子元素中有 7 个是集中在电子文件生命周期的“后端”,其具体的语义结构如表 3。

从以上时间元素的语义结构可以看出,在时间元素的 9 个下位子元素中只有 2 个是关于电子文件创建阶段的时间的,如“文件创建时间”和“文件最后编辑时间”。对于电子文件现行阶段,没有时间元素的任何下位子元素来反映。而时间元素中的其他 7 个

下位子元素(表3中3~9)集中反映的是电子文件在非现行阶段的时间的,也就是说,这个元数据标准中,时间元素其语义的重心是在电子文件生命周期的“后端”。时间元素的这种后端模式的语义结构,其优势主要有两点:

表3 南非《电子文件核心元数据》标准时间元素的语义结构与电子文件生命周期的对应

上位元素	下位子元素——元素限定词	文件生命周期
时间	1. 文件创建时间	电子文件创建阶段
	2. 文件最后编辑时间	
	无	电子文件现行阶段
	3. 文件登录时间	电子文件非现行阶段
	4. 文件归档时间	
	5. 归档文件集创建/结束时间	
	6. 归档文件组创建/结束时间	
	7. 归档文件移交时间	
	8. 归档文件版本创建时间	
	9. 归档文件阅读时间	

(1)结构简单,通过时间元素下位子元素的向下扩展,可以反映电子文件生命周期“后端”中与事件有关的时间。

(2)向上兼容,通过时间元素下位子元素的向上兼容,可以实现时间元素纵向上的语义互操作。

但它也有局限性,主要表现在:

(1)该模式时间元素的下位子元素主要反映的是电子文件生命周期的“后端”,因此,在时间元素下没有设任何下位子元素来反映电子文件现行阶段的相关事件的时间,从而也就造成这一阶段的下位子元素的“空白”。

(2)当电子文件处于生命周期的现行阶段时,时间元素难以向前扩展对该阶段的电子文件的进行时间描述。

4 电子文件管理元数据中时间元素的连续体模式的语义结构

由于电子文件管理元数据中时间元素的前端模式语义结构和后端模式语义结构,都难以全面反映电子文件在其整个生命周期中各个阶段与事件相关的时间,所以,在国际电子文件管理元数据领域,又出现

了一种时间元素的连续体模式的语义结构,这种结构能够连续地反映电子文件生命周期的各个阶段,也就是说它既能够反映电子文件生命周期的“前端”和电子文件的现行阶段,又能够反映电子文件生命周期的“后端”,即完整地反映电子文件的整个生命周期。但是这种时间元素的语义结构,它所依据的电子文件生命周期的模型,同我们前面所引用的传统经典的电子文件生命周期模型不一样,它依据的是现代电子文件生命周期的“电子文件连续体”的模型,该模型示意如图3。

从电子文件管理轴上看,整个过程是:首先创建的是“电子文件文稿”,完成文稿创建后,电子文件便进入现行阶段,成为“现行电子文件”,当现行电子文件其现行职能结束后,便进入归档阶段,形成“归档电子文件”,然后再经过移交,电子文件进入文件保管部门形成了“电子档案文件”。根据文件连续体模型的这一原理,澳大利亚摩纳大学研制的《澳大利亚电子文件管理元数据》标准对时间元素建立了如表4的语义结构^[4]。

表4 《澳大利亚电子文件管理元数据》时间元素的语义结构与电子文件连续体的对应

上位元素	下位子元素——元素限定词	所对应的电子文件连续体模型中的文件状态
时间	1. 文稿创建时间	电子文件文稿
	2. 文稿登录时间	
	3. 文件接收时间	
	4. 文件发送时间	现行电子文件
	5. 文件传递时间	
	6. 文件移交时间	归档电子文件
	7. 文件处置时间	电子档案文件

从上面时间元素的语义结构中可以看出,其下位子元素的分布是从文稿创建开始,一直连续到了电子档案文件的处置,也就是相当于传统经典的电子文件生命周期的全部过程。我们将该时间元素的下位子元素与传统经典的电子文件生命周期的各个阶段对应起来就可以看得更清楚,其对应如表5。

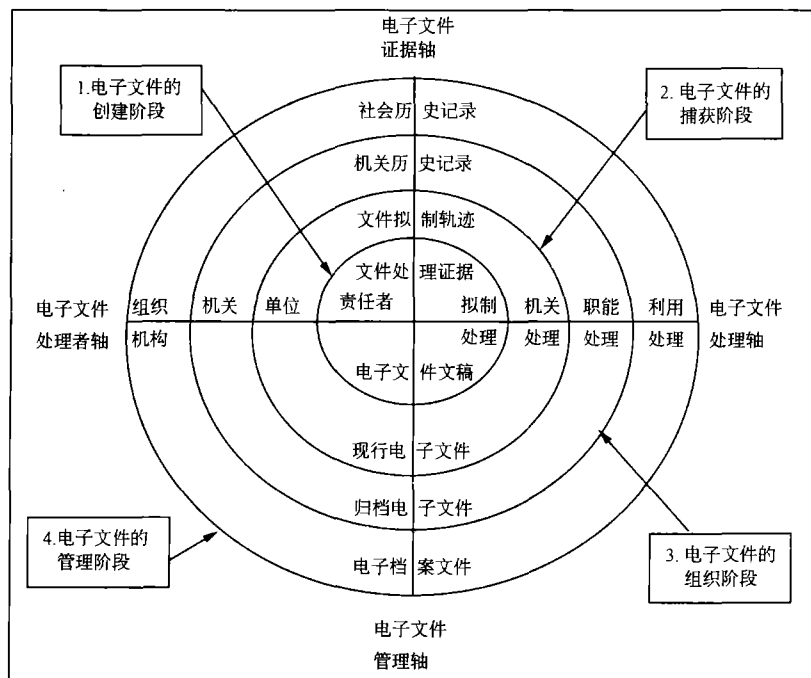


图3 电子文件连续体模型

表5 《澳大利亚电子文件管理元数据》时间元素的语义结构与电子文件生命周期的对应

上位元素	下位子元素——元素限定词	所对应的传统电子文件生命周期
时间	1. 文稿创建时间	电子文件创建阶段
	2. 文稿登录时间	
	3. 文件接收时间	
	4. 文件发送时间	电子文件现行阶段
	5. 文件传递时间	
	6. 文件移交时间	电子文件非现行阶段
	7. 文件处置时间	

通过上面的对应可以看出,在时间元素的连续体模式的语义结构中,其下位子元素在整个文件生命周期过程中的分布是连续的。这种连续体模式的语义结构具有如下优势:

(1)在纵向语义上既具有很好的向上兼容性,又具有很好的向下扩展性。

(2)由于时间元素的下位子元素是连续分布的,所以在整个电子文件生命周期中具有很好的向前与向后的动态兼容性。

(3)由于时间元素的下位子元素是连续分布的,所以十分便于对时间元素的集成管理和一体化管理。

当然这种结构也有局限:反映与电子文件生命周期中各个阶段相关事件时间的详尽程度还不够,或者说反映电子文件生命周期中各个阶段时间元素的下位子元素的数量还不足够的多。

5 电子文件管理元数据中时间元素的分面语义结构

在整个电子生命周期中,实际上有很多事件的发生与电子文件管理元数据中的时间元素都是紧密相关的,尤其是当电子文件进入到“后端”,它所涉及的事件比其他阶段都要多得多,因此,所需描述相关事件所要求的时间元素的下位子元素也要比其他阶段多得多,而前面所述的时间元素的前端模式语义结构、后端模式语义结构和连续体模式的语义结构,在这种情况下,都难以满足这种要求。为了解决这个问题,2003年,美国明尼苏达州颁布的《电子文件管理元数据标准》中^[5],就对时间元素的语义结构进行了分面化处理。该标准中除了时间元素的本体的语义结构,还设立了3个具有时间元素分面的语义结构,其结构具体如表6。

表6 美国明尼苏达州《电子文件管理元数据标准》
时间元素的分面语义结构

时 间		
管理类时间 (分面结构一)	利用类时间 (分面结构二)	保存类时间 (分面结构三)
1. 利用审查时间	1. 列表时间	(当前保存类时间)
2. 鉴定时间	2. 利用元数据时间	1. 备份时间
3. 审记时间	3. 内容利用时间	2. 转存图像时间
4. 条件变更时间	4. 非授权利用时间	3. 载体更新时间
5. 密级变更时间	5. 预定时间	4. 缩微时间
6. 代码变更时间	6. 拷贝时间	5. 平台迁移时间
7. 移交时间	7. 下载时间	6. 软件迁移时间
8. 解密时间	8. 屏幕转储时间	7. 脱机备份时间
9. 销毁时间	9. 审读时间	(下次保存类时间)
10. 分发时间	10. 安全攻破时间	1. 下次审查时间
11. 案卷结卷时间		2. 下次载体审查时间
12. 保管地变更时间		3. 下次备份时间
13. 出版时间		4. 下次转存图像时间
14. 相关关系时间		5. 下次载体更新时间
15. 发布时间		6. 下次缩微时间
16. 发布号变更时间		7. 下次平台迁移时间
17. 保管期限设定时间		8. 下次软件迁移时间
18. 保管期限变更时间		9. 下次脱机备份时间
19. 档号变更时间		10. 其他时间
20. 数字签名时间		

从以上时间分面的语义结构可以看出,时间元素首先被分出了3个面,如分面结构一“管理类时间”、分面结构二“利用类时间”和分面结构三“保存类时间”,然后在每个分面下再详尽地列出各自分面的下位子元素,3个分面加起来共有48个子元素。如果将如此多的子元素直接列在时间元素下作为它的下位子元素,时间元素的语义结构会变得臃肿不堪很难处理。所以,该标准采用了一个比较好的办法,就是将各个分面提升,作为电子文件管理元数据标准中独立的元素来处理,如将分面“管理类时间”提升为电子文件管理元数据中的“管理史”元素,将分面“利用类时间”提升为“利用史”元素,而将分面“保存类时间”提升为“保存史”元素,于是就形成了时间元素的分面语义结构。

这种结构具有如下优势:

- (1) 结构简洁,分面清晰,便于元数据著录。
- (2) 既便于时间元素下位子元素的向下详尽扩展,又避免了时间元素语义结构的臃肿不堪。
- (3) 可以很好地满足处理电子文件生命周期中各个阶段事件与时间元素相关的详细需求。

但是这种分面结构也有不足。由于这种结构是把时间分面提升为独立的元素结构,所以,相对于被提升上来的分面而言,其上位元素“时间”则成为它们隐性的上位元素,而不是显性的,不便于向上兼容。

6 结论

电子文件管理元数据中时间元素的语义结构是一个复杂问题,但是从上面分析可知,相比较而言,时间元素的语义结构最好是以电子文件的连续体模型为依据来建构,并同时采用时间元素的分面语义结构。这样既可以满足反映电子文件在它整个生命周期中各个阶段与事件相关的各个时间,又可以具有最大的兼容性;既可以使时间元素的语义结构简洁清晰,又可以具有最大的扩展性。目前,电子文件管理元数据中时间元素的连续体模型的语义结构和分面语义结构都已经引起了国际标准化组织的重视。如何建立适合我国国情的电子文件管理元数据标准,值得我们认真思考和深入研究。

参考文献

- 1 International Standard Organization (ISO). Information and Documentation—The Dublin Core Metadata Element Set. Feb 26, 2003
- 2 Office of the e-Envoy of UK. E-Government Metadata standard (e-Gms). Version 2.0, December 2003
- 3 National Archives and Records Service of South Africa. Minimum Mandatory Metadata. Version 1.1, April 2004
- 4 Monash University SPIRT Project of Australia. Australian Recordkeeping Metadata Schema. Version 1.0, May 2000
- 5 Minnesota State Archives of US. Minnesota Recordkeeping Metadata Standard. Version 1.2, April 2003

张正强 教授,国家社科基金重点资助项目“基于XML电子文件管理元数据标准化研究”项目首席负责人。通信地址:上海四平路2575号南京政治学院上海分院信息管理系统。邮编200433。(来稿时间:2005-03-31)