

标准数字化转型中标准标签集发展与应用分析

赵青青 蔡焱

(上海市质量和标准化研究院)

摘 要: 本文聚焦标准数字化转型中标准标签集这一关键技术,系统回顾了标准标签集从ISO STS演变到NISO STS的发展历程,并分析了两者的区别。随后围绕NISO STS的组成,明确了标准标签集中元素和属性的特征及用法,并分析了<standard>和<adoption>两种根元素结构。最后结合标准标签集在国内外的应用情况,分别介绍了相关典型案例,并指出NISO STS需要进一步引入语义更丰富的元素,从而增强标准的机器可读能力。

关键词: 标准数字化转型, 标准标签集, 机器可读标准

DOI编码: 10.3969/j.issn.1674-5698.2024.01.014

Development and Application Analysis of Standards Tag Suite in Standards Digital Transformation

ZHAO Qing-qing CAI Yan

(Shanghai Institute of Quality and Standardization)

Abstract: This paper focuses on standards tag suite, which is the key technology in the digital transformation of standards, systematically reviews the development course of standards tag suite from ISO STS to NISO STS, and analyzes the differences between the two ones. Then, based on the composition of NISO STS, the characteristics and usage of elements and attributes in tag suite are clarified, and the two root element structures of <standard> and <adoption> are analyzed. Finally, the typical application cases at home and abroad are listed, and this paper points out that NISO STS needs to introduce more semantic elements to enhance the machine readable ability of the standard.

Keywords: standards digital transformation, standards tag suite, machine readable standards

0 引言

当前标准数字化转型正在如火如荼地进行中。标准标签集(Standards Tag Suite, STS)作为标准数字化转型的关键技术之一,是推动标准数字化转型的基础。标准标签集通过定义一套

XML元素和属性,对标准的规范性和非规范性内容、标准采用和类标准文档进行XML编码,使标准的知识内容独立于最初交付内容的形式。因此,标准标签集又被称为“标准的标准”。

1 标准标签集发展历程

基金项目: 本文受上海市市场监督管理局2023年度科技项目“标准数字化转型中标准元数据标签集构建研究”(项目编号:2023-59)资助。

作者简介: 赵青青,硕士,标准化工程师,研究方向为标准馆藏建设与标准信息服务。

蔡焱,硕士,标准化工程师,上海标准文献馆主任,研究方向为标准信息管理和标准化服务。

1.1 ISO STS

2011年底, 国际标准化组织ISO对其出版系统进行改进, 并与Mulberry公司合作, 共同开发了用于ISO标准出版的ISO标准标签集(ISO STS)。自2011年第一版发布以来, ISO STS已经经过5个版本的修订, 最新版本ISO STS 1.1于2013年发布, 以美国国家信息标准组织NISO发布的ANSI/NISO Z39.96《期刊文章标签集》(Journal Article Tag Suite, JATS)的0.4版草案为基础。

自ISO STS发布以来, 一些ISO成员如英国标准协会BSI、澳大利亚标准协会SA和部分分销商已经采用该STS^[1]。但由于ISO STS并非官方标准, 无法与JATS保持耦合, 因此美国的一些标准开发组织和发行商并未采用。此后, 不同标准开发组织使用了不同的XML模型来标记标准和类标准文档, 众多的非标准化XML模型使得标准组织之间的互操作变得困难, 增加了开发成本。推动ISO STS走向标准化并与JATS建立正式关系势在必行。随后, ISO与NISO讨论创建基于ISO STS 1.1的“标准的标准”, 2017年NISO发布NISO STS标准, 并正式命名为ANSI/NISO Z39.102-2017(NISO STS 1.0)^[2]。2020年春季, ISO在其在线标准开发平台(Online Standards Development platform, OSD)中正式采用NISO STS。

1.2 NISO STS

1.2.1 NISO STS 1.0

ISO STS 1.1和NISO STS 1.0均基于JATS进行开发, 前者基于JATS的0.4版, 后者基于2015年发布的JATS 1.1版。同时NISO STS 1.0还以ISO STS 1.1为基础, 完全向后兼容ISO STS 1.1。

NISO STS 1.0包括两个标签集: 交换标签集和扩展标签集, 这两个标签集由定义的元素和属性构成, 旨在为标准发布和互操作提供模型, 不同之处在于交换标签集中唯一的表模型是基于XHTML, 而扩展标签集还提供OASIS/CALS表模型, 为每个包含MathML2或MathML3的标签集提

供了语法(DTD、XSD和RNG形式)。

截至2018年3月底, ISO、IEC、CEN、BSI、DIN、AS等国际和国家标准机构已采用NISO STS 1.0, ASTM、ASME、IEEE、API和SAE等机构或采用NISO STS 1.0, 或采用与其结构基本相同的模型^[3]。因为JATS是期刊出版中的XML标准, 而NISO STS又基于JATS, 因此JATS和NISO STS共享一组通用模块, 这对IEEE、ASME等同时出版期刊和标准的组织非常有利。

1.2.2 NISO STS 1.2

2022年11月14日, NISO宣布发布更新版本STS, 即NISO STS 1.2^①, 该版本已被美国国家标准协会ANSI批准为正式标准ANSI/NISO Z39.102-2022^[4]。

NISO STS 1.2扩展了元素和属性描述, 以及交换标签集和扩展标签集, 它还包含了对JATS 1.3(ANSI/NISO Z39.96-2021)所做的适用修改。新版本变化包括: 描述XML文件本身的规定、增加/修改了元素和属性、扩展NISO STS TBX简介和术语显示结构等^[5]。此外, NISO STS 1.2是完全向后兼容的, 任何对版本1.0有效的文档也对版本1.2有效。

1.3 NISO STS与ISO STS的关系

此处主要以ISO STS 1.1和NISO STS 1.0为例, 进行具体分析。NISO STS 1.0中的一些关键改进使其比ISO STS 1.1更加灵活^[6]。

(1) NISO STS 1.0扩展了元数据功能, 通过新增<std-meta>, 旨在满足任何标准组织的需求, 而不是局限于ISO及国家标准机构;

(2) NISO STS 1.0简化了术语显示模型<term-display>, 并添加了额外的语义标签, 从而允许进行更注重显示的建模;

(3) NISO STS 1.0为标准采用提供了一种新的递归模型;

(4) NISO STS 1.0同时容纳XHTML和CALS表, 以及MathML2和MathML3。相比之下, ISO STS 1.1仅支持XHTML和MathML2。

注: ① 鉴于ISO STS 1.1, 故命名为NISO STS 1.2, 以作区别。

2 NISO STS的组成

2.1 元素

元素是名词,例如standard、paragraph和ICS,它们是标准本身、标准的组成部分以及元数据。在NISO STS中,每个元素都有两个名称,一个是标签名称,一个是元素名称。标签名称是在标记文档、DTD片段和模式以及软件中使用的较短的机器可读名称,元素名称是较长的描述性名称,例如<p>是元素paragraph的标签名称。NISO STS 1.2定义了353种元素。

在NISO STS中,描述元素以元素的标签名称开始,后面紧跟元素名称,许多元素还带有用法或备注,以便将该元素与其他类似元素区分开。NISO STS还提供元素允许使用的属性类型、元素的父元素类型以及以何种组合使用的描述等,同时大多数元素都包含如何使用的标记示例。

2.2 属性

属性是与元素相关联的名称-值对,用于修改元素的某些特性。属性保存有关元素的事实,例如:在使用元素<sec>时,可使用属性@sec-type确定相应的章节类型(例如scope、foreword、norm-refs)。属性也有两个名称,一个是较短的机器可读名称(标签名称),另一个是较长的人类可读的描述性名称,例如@id是属性Document Internal Identifier的标签名称。NISO STS 1.2定义了190种属性。

属性的描述方式与元素非常相似,但因属性不能有子属性,所以只描述该属性可以应用于哪些元素、该属性的性质,以及属性允许的值和默认值。

3 NISO STS的主要内容

3.1 根元素

NISO STS主要通过文档层次图展示层次结构。每个层次都有一个根元素,可以说明其他几个附属元素的结构。针对标准本身、采用标准这两种文本的特点,NISO STS 1.2定义了两个根元素<standard>、<adoption>,任何一个都可用作包含

标准文本的文档元素。<standard>用于标准和诸如指南或手册等其他类标准文档;<adoption>包含关于采用和原始标准的信息,采用可能包括采用标准的组织信息,并包括采用组织提供的前页(如前言)和附属信息。<adoption>可以嵌套采用,或采用一个或多个<standard>。

3.2 根元素<standard>

<standard>可以分为如下结构部分(如图1所示)。

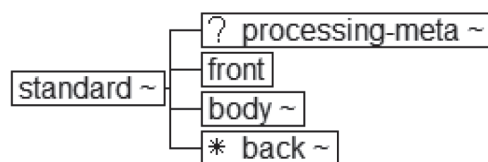


图1 标准文本<standard>的结构

(1) 前页内容<front>,描述标准的封面及前言(如图2所示)。

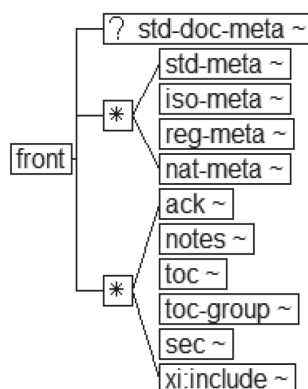


图2 标准前页<front>结构

(2) 主体<body>,描述标准的正文(如图3所示)。

(3) 附属信息<back>,描述标准的附录和参考文献(如图4所示)。

(4) 处理元数据<processing-meta>,描述有关XML文件本身的处理信息(不是由XML文件编码的标准)。

3.3 根元素<adoption>

<adoption>可以分为如下结构部分(如图5所示)。

(1) <adoption-front>描述采用标准自己的前页内容,包括文档级元数据及采用组织的元数据等;

(2) <back>描述采用组织提供的可选附属信

息。`<adoption>`模型允许在被采用标准之前或之后使用`<back>`，但不能同时使用；

(3) `<adoption>`描述被采用标准的内容，可能包括嵌套使用一个或多个`<standard>`、`<adoption>`；

(4) `<processing-meta>`描述有关XML文件本身的处理信息（不是由XML文件编码的采用）。

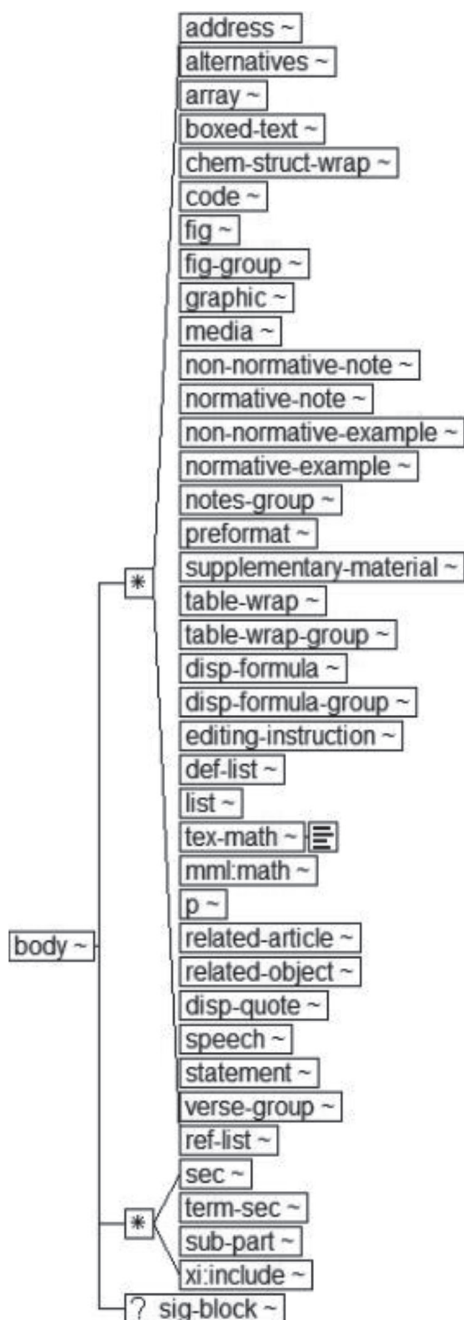


图3 标准主体`<body>`结构

4 国内外应用情况分析

4.1 国外应用情况

4.1.1 ISO/IEC在线标准开发平台OSD

在线标准开发平台OSD由ISO和IEC联合开发，为标准开发人员提供了一个全新的数字化工具，以简化起草和编辑国际标准的过程^[7]，提高了从标准准备到最终发布的整个过程的效率和协作。

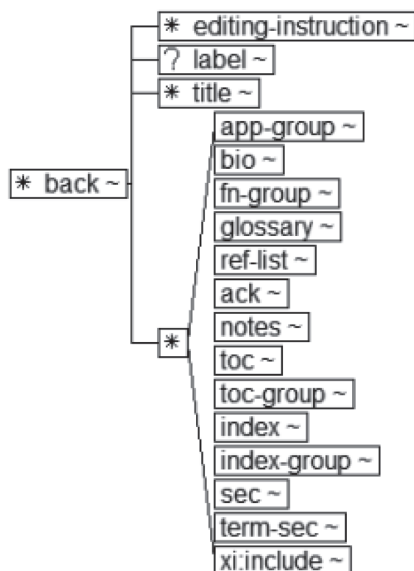


图4 标准附属信息`<back>`结构

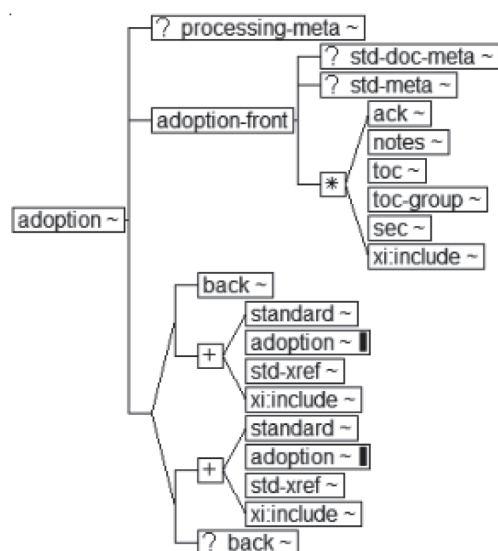


图5 采用标准文本`<adoption>`结构

OSD平台旨在用用户友好的XML编辑器(Fonto XML编辑器)取代传统的基于Word的内容创建过程,帮助标准开发人员从源头创建复杂的结构化内容。虽然标准内容基于NISO STS编码,但是标准开发人员却无需了解XML或结构化内容编写。XML优先流程不仅便于编辑处理,而且由于编辑指令规则嵌入到OSD中,因此在标准开发的早期阶段还可以提高内容质量。高效的工具允许标准开发人员专注于内容而不是格式,从而生成语义丰富和结构化的标准。CEN和CENELEC也引进了该平台。

4.1.2 可互操作规范和标准的语义网SWISS

SWISS(The Semantic Web for Interoperable Specifications and Standards)由美国XSB公司开发,是用于互操作规范和标准的语义网平台^[8]。SWISS通过建立在NISO STS XML之上的关联数据模型^[9],将PDF文档转换为上下文相关、可操作的数字数据对象,通过工程知识图谱技术,建立和管理文档和概念之间的联系,同时可通过API自动将结果数据传送到企业的内部系统^[10]。SWISS通过在内部内容和外部标准之间建立自动和实时连接,提高合规性并降低风险。

4.2 国内应用情况

4.2.1 中国标准出版社“标准在线起草系统”

中国标准出版社将传统的标准制修订业务与NISO STS相结合,从源头对标准文件进行XML数据化处理,实现对标准的全生命周期管理。标准在线起草系统可提供起草阶段的在线协同编制、可视化修改、在线沟通,出版阶段的智能审校、排

版精调,发布阶段的微信推送和动态提醒等功能,对于缩短标准制修订周期、提高工作效率、提升标准编写质量等方面发挥重要作用。

4.2.2 同方知网“数字标准智能应用平台”

数字标准智能应用平台基于机器可读和可理解的数字标准模型,面向标准化工作上下游场景及企业内部标准化业务,实现标准全过程数字化管理。该平台包含3个子系统,其中底层标准数字化加工系统按照ISO STS的要求,实现了标准智能标引、指标抽取、知识图谱、碎片化阅读等功能,具备ISO/IEC机器可读标准模型中二级水平。

5 结语

实施标准标签集是标准数字化转型的重要一环。我国对标准标签集的研究最早可见于2019年发布的国家标准GB/T 37967-2019《基于XML的国家标准结构化置标框架》^[11],但GB/T 37967-2019并未基于JATS,因此与ISO STS、NISO STS在结构上存在较大差异,故未能和国际接轨。随着ISO/IEC机器可读标准分级模型的提出,标准标签集成为描述基于XML编码的机器可读文件的关键技术,但是ISO STS、NISO STS主要面向标准出版,因此包含许多格式元素。当对标准中表格、段落中关键技术指标进行标记时,一般用HTML呈现,因此缺乏相应的语义和特定含义,故不能进行逻辑推理判断。要想实现更高水平的机器可读能力,还需进一步引入语义更丰富的元素,从而真正实现标准的机器可用、可读、可解析。

(下转第125页)

参考文献

- [1] 中国互联网络信息中心. 第51次中国互联网络发展状况统计报告[R/OL]. (2023-03-02) [2023-07-20]. <https://cnmic.cn/NMediaFile/2023/0322/MAIN16794576367190GBA2HA1KQ.pdf>.
- [2] 费威. 共享经济模式及其监管制度供给[J]. 经济学家, 2018(11): 75-82.
- [3] 李乃文, 荣帅, 赵宏霞. 双边市场环境下网购平台的质量诚信监控行为研究[J]. 软科学, 2017, 31(08): 129-133+138.
- [4] 颜海娜, 于静. 网络订餐食品安全“运动式”治理困境探究——一个新制度主义的分析框架[J]. 北京行政学院学报, 2018(03): 81-90.
- [5] 吕永卫, 霍丽娜. 网络餐饮业食品安全社会共治的演化博弈分析[J]. 系统科学学报, 2018, 26(01): 78-81.
- [6] 朱立龙, 孙淑慧. 消费者反馈机制下食品质量安全监管三方演化博弈及仿真分析[J]. 重庆大学学报(社会科学版), 2019, 25(03): 94-107.
- [7] 吴元元. 食品安全治理中的声誉异化及其法律规制[J]. 法律科学(西北政法大学学报), 2016, 34(02): 127-136.
- [8] Friedman D. On economic applications of evolutionary game theory[J]. Journal of Evolutionary Economics, 1998, 8(1): 15-43.

(上接第100页)

参考文献

- [1] International Organization for Standardization. ISO Standards Tag Set (ISOSTS) version 1.1: A customization of NISO JATS version 0.4[EB/OL]. [2023-09-28]. <https://www.iso.org/schema/isosts/v1.1/doc/index.html>.
- [2] ANSI/NISO Z39.102-2017[S/OL]. [2023-09-28]. <https://www.niso-sts.org/standard-html/v1-0/index.html>.
- [3] INERA. Why NISO STS is a better choice than DITA for your standards publishing[EB/OL]. [2023-09-28]. <https://www.inera.com/blog/why-niso-sts-is-a-better-choice-than-dita-for-your-standards-publishing/>.
- [4] ANSI/NISO Z39.102-2022[S/OL]. [2023-09-28]. <https://groups.niso.org/higherlogic/ws/public/download/27991/Z39.102-2022.pdf>.
- [5] National Information Standards Organization. NISO STS (Standards Tag Suite) Changes Between NISO STS 1.2 and NISO STS 1.0[EB/OL]. [2023-09-28]. <https://www.niso-sts.org/downloadables/NISO-STS-1-2-compare-1-0.pdf>.
- [6] National Information Standards Organization. NISO STS (Standards Tag Suite) Differences Between ISO STS 1.1 and NISO STS 1.0[EB/OL]. [2023-09-28]. <https://www.niso-sts.org/downloadables/NISO-STS-1-0-compare-ISO-STS-1-1.pdf>.
- [7] 姜冠男, 施琴. 标准组织数字化转型国际趋势研究[J]. 质量与标准化, 2022(03): 38-41.
- [8] 本刊编辑部, 赵子军. 能否分享若干个标准数字化范畴的典型范例或“标准数字化故事”? [J]. 中国标准化, 2022(05): 28-31.
- [9] National Library of Medicine. SWISS: The Semantic Web for Interoperable Specifications and Standards[EB/OL]. [2023-09-28]. <https://www.ncbi.nlm.nih.gov/books/NBK430699/>.
- [10] 李松丽, 曹平, 姜盼. 国际标准化组织的标准标签集研究分析[J]. 航空标准化与质量, 2018(2): 52-56.
- [11] 全国语言与术语内容资源标准化技术委员会. GB/T 37967-2019, 基于XML的国家标准结构化置标框架[S]. 北京: 中国标准出版社, 2019: 1.