

中国计算机用户协会团体标准

《版本典藏网站资源长期保存 核心元数据》

（征求意见稿）编制说明

一、标准编制的背景

典藏是文化传承和保护的重要形式，版本典藏网站资源是具有极高的历史、文化、科研和艺术价值的互联网内容资源，将这些具有收藏价值的网站资源保存起来，不仅可以为后代留下重要的历史遗产，使人们更好地理解和研究文化发展脉络；同时也为文化传承提供了宝贵资源，为现代文化的多元化和繁荣奠定基础；更为科学研究提供重要的依据和材料，有助于推动学术研究，促进科技进步。

当前，在互联网资源保存研究与实施中还存在的重要性认识不足、标准规范不健全、技术能力较弱等情况，难以适应互联网异构数据快速增长、移动互联网创新应用不断涌现等问题。为满足面向 50 年以上互联网资源存储、检索和还原展示的需求，需要研制数据格式、操作流程、系统接口等相关标准，用以规范技术开发和系统运行，实现互联网资源长期安全存和科学管理以及与国际互联网保存标准互联互通的目标。

二、任务来源

根据中国计算机用户协会下达的 2023 年下半年标准制修订计划，中国国家版本馆作为主要牵头单位筹建了标准起草组，承担《版本典藏网站资源长期保存 核心元数据》标准的研制任务。该标准的立项计划号为 T/CCUA LX009-2023，技术归口单位为中国计算机用户协会。

三、编制过程

在下达计划号前，2023 年 8~9 月期间多次召开起草组大纲讨论会。

2023 年 11 月 1 日召开标准立项评审会通过项目立项。

2023 年 3 月完成标准工作组稿。

2024 年 3 月，标准起草组完成了标准草案的初稿，于 2023 年 4 月初开始

在起草组评审委员会内进行评审。共经历了 1 次组内评审，期间收到了 10 条意见。

2024 年 4 月征求意见稿初稿提交标准委员会审核。

2024 年 5 月 20 日收到标准专家冯惠（电子四院[退休]）意见 10 条，对不符合标准规范性的内容进行了修改。

2024 年 6 月 12 日收到标委会意见 4 条，对相应内容进行了修改。

四、编制原则

标准的用语、格式按照 GB/T1.1-2020 给出的规则起草。

标准内容的编制坚持以下原则：

1. 实用性

本标准从网站资源长期保存应用的实际需要出发，采用归纳、抽象和总结的方法形成具有通用性的实体和语义单元。以实际需求为基础，结合长期保存活动的特点整合而成，使标准具有较强的实用性。

2. 科学性

本标准科学地定义核心元数据的实体、和语义单元，同时也描述了实体间的关联关系以及实体间的继承关系。

3. 通用性

本标准充分考虑到资源对象、保存活动、保存技术和环境等方面的需求，确保在应用过程中充分满足不同目的、不同技术和不同管理的多重需求。

4. 可扩展性

网站资源长期保存活动是需要持续维护和应用的，在标准的编制过程中充分考虑了各方面扩展的需求，提高标准的适应性和使用周期。

五、标准主要内容

本标准规定了长期保存核心元数据模型的构成，元数据模型由两级 11 个实体组成，同时还描述了实体间的关系。本标准还描述了不同实体的语义单元属性，包括名称、组分、定义、说明、数据约束、范例、重复性和必备性。

本标准适用于各类图书馆、团体、企事业单位、社会组织从事互联网资源长期保存活动时参考。

本标准不包含元数据著录方法和规范。

六、有关技术的说明

有关本标准起草过程中的一些技术问题说明如下：

1、本标准描述的元数据是长期保存活动中不同实体的元数据组成的元数据体系，用于描述长期保存过程中的各种活动、各种资源对象、各种责任主体以及各种软硬件环境等。

2、本标准中的资源对象分为知识对象、表现和文件。其中知识对象是抽象的概念，例如一个网站；表现则是知识对象的一种表现形式或者版本，例如一个网站的备份；文件则是组成表现的各类文件，例如某一个 WARC 文件。

参考的主要标准：

1. GB_T 18787.3-2015 信息技术 电子书 第 3 部分：元数据
2. WH-Z1-2012 图书馆数字资源 长期保存元数据规范
3. T_BAS 004-2023 民间文物核心元数据规范

七、关于标准的性质

鉴于本标准作为团体标准发布，属于推荐性标准。由本团体成员约定采用或者按照本团体的规定供社会自愿采用。

八、有关专利的说明

本标准不涉及专利问题。

标准起草组

2024 年 4 月 14 日