

T/CCUA

中国计算机用户协会团体标准

T/CCUA XXXX—XXXX

版本典藏网站资源长期保存 服务接口要求

Long-term preservation website resources Service interface requirement

（征求意见稿）

（本草案完成时间：2024-6-18）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

中国计算机用户协会 发布

目 次

前 言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 2

5 技术要求 2

 5.1 通讯协议要求 2

 5.2 接口技术类型 3

 5.3 性能要求 3

 5.4 对象实体 3

6 接口要求 3

 6.1 访问身份验证接口 3

 6.1.1 接口要求 3

 6.1.2 输入与输出 3

 6.2 资源对象检索接口 3

 6.2.1 接口要求 3

 6.2.2 输入与输出 4

 6.3 资源对象详情获取接口 4

 6.3.1 接口要求 4

 6.3.2 输入与输出 4

 6.4 资源对象提取接口 4

 6.4.1 接口要求 4

 6.4.2 输入与输出 4

 6.5 保存实体检索接口 4

 6.5.1 接口要求 4

 6.5.2 输入与输出 5

 6.6 保存实体详情获取接口 5

 6.6.1 接口要求 5

 6.6.2 输入与输出 5

 6.7 网站回放接口 5

 6.7.1 接口要求 5

 6.7.2 输入与输出 5

参 考 文 献 6

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由中国计算机用户协会提出并归口。

本文件起草单位：中国计算机用户协会创新技术应用分会、中国国家版本馆、江苏嘉图网络科技有限公司

本文件主要起草人：刘成勇、王志庚、张琦、吕健泳、张颖，马驰、马腾飞、王文清、周长岭

本文件于 XXXX 年首次发布。

版本典藏网站资源长期保存 服务接口要求

1 范围

本文件规定了版本典藏网站资源长期保存服务接口的基本要求、技术要求和接口要求，界定了有关的术语和定义。

本文件适用于各类图书馆、团体、企事业单位、社会组织从事网站资源长期保存服务接口的开发、管理等活动。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

T/CCUA AAAA-XXXXX 版本典藏网站资源长期保存 核心元数据

RFC 2616 超文本传输协议—HTTP/1.1 (Hypertext Transfer Protocol—HTTP/1.1)

3 术语和定义

下列术语和定义适用于本文件。

3.1 长期 long-term

虑及技术变化(包括支持新载体、新数据格式)和用户群体变化对典藏信息影响的足够长的时间段。这个时间段延伸到无限未来。

[ISO 14721:2012 定义1.7.2]

3.2 长期保存 long-term preservation

一种可靠的长期维护版本典藏网站资源真实、完整、有效的行为。

注：改写DA/T 47—2009 3.3

3.3

资源对象实体 resource object entity

一个数字形式的信息单元，包含三个子类：知识对象实体、表现实体和文件实体。

[T/CCUA XXX—2004 术语和定义 3.3]

3.4

知识对象实体 intellectual entity

一个聚合在一起的内容单元（或是信息单元），例如，图书、地图、照片、数据库或软件。一个知识对象实体可以包含多个二级的知识对象实体。一个知识对象实体可以有一个或多个数字的表现形式。

[T/CCUA XXX—2004 术语和定义 3.4]

3.5

表现实体 representation entity

用于对知识对象实体进行完整而合理再现的一组文件的整体表示。

[T/CCUA XXX—2004 术语和定义 3.5]

3.6

文件实体 file entity

用于描述资源文件主要特征的实体。

[T/CCUA XXX—2004 术语和定义 3.6]

3.7

活动实体 activity entity

涉及或影响至少一个资源对象或责任者的行为。

[T/CCUA XXX—2004 术语和定义 3.7]

3.8

权利实体 rights entity

属于资源对象与/或责任者的一种或多种权利或许可的声明主张。

[T/CCUA XXX—2004 术语和定义 3.9]

3.9

责任者实体 agent entity

长期保存活动中所涉及的个人或组织。包括两个子类：机构实体和个人实体。

[T/CCUA XXX—2004 术语和定义 3.10]

3.10

软件实体 software entity

长期保存活动中所涉及的各种类型的应用软件、系统软件、中间件等软件系统。

[T/CCUA XXX—2004 术语和定义 3.13]

3.11

环境实体 environment entity

长期保存活动中所涉及的各种应用软件、系统软件、中间件等软件系统、硬件和网络环境。

[T/CCUA XXX—2004 术语和定义 3.14]

3.12

服务接口提供方 service interface provider

拥有版本典藏网站资源长期保存系统并通过服务接口方式提供网站资源数据服务的组织或个人。

3.13

服务接口使用方 service interface consumer

通过服务接口使用版本典藏网站资源长期保存系统的组织或个人。

4 缩略语

下列缩略语适用于本文件。

HTTP: 超文本传输协议 (Hyper Text Transfer Protocol)

HTTPS: 安全超文本传输协议 (HyperText Transfer Protocol Secure)

REST: 表述性状态转移 (Representational State Transfer)

XML: 可扩展置标语言 (Extensible Markup Language)

JSON: Javascript对象标记法 (JavaScript Object Notation)

IP: 网际互连协议 (Internet Protocol)

SOAP: 简单对象访问协议 (Simple Object

5 技术要求

5.1 通讯协议要求

通讯传输协议应采用 HTTP 标准协议或HTTPS标准协议：

- a) 服务接口采用 HTTP 标准协议时应采用 HTTP1.0 或 HTTP1.1 标准协议；
- b) 服务接口采用 HTTPS 协议时，应采用在中国境内注册的权威证书认证机构颁发的 SSL 证书。

5.2 接口技术类型

版本典藏网站资源长期保存服务接口技术类型应采用REST接口或Webservice接口：

- a) 服务接口采用 REST 接口时应使用 GET 或 POST 方法，返回报文数据格式为 JSON 格式；
- b) 服务接口采用 Webservice 接口时，应采用 SOAP1.1 或 SOAP1.2 访问协议，传输数据格式为 XML 格式。

XML格式为保证接口数据有效传输，应对接口相关的消息报文中的特殊字符进行转义处理，确保各服务接口使用方能够成功处理有效的XML报文数据，具体见表1所示。

表1 特殊字符转义对应表

序号	转义前	转义后
1	&	&
2	<	<
3	>	>
4	“	"
5	,	'

5.3 性能要求

服务接口的响应时间应小于或等于5s。

输入、输出相关数据在传输过程中都应加密。使用HTTPS协议进行通信，确保数据的机密性和完整性。

5.4 对象实体

本文件涉及的实体均出自T/CCUA XXX-2024。

6 接口要求

6.1 访问身份验证接口

6.1.1 接口要求

访问身份验证接口用于验证服务接口使用方身份的合法性，只有通过验证才能使用其他访问接口。接口要求包括：

- a) 服务接口使用方必须提供有效的用户名和密码进行身份验证，为了提高安全性，接口可以要求使用多因素认证，如手机验证码等；
- b) 当服务接口使用方通过身份验证后，系统应生成一个访问令牌，服务接口使用方在后续的请求中必须携带这个令牌以证明其身份，令牌需要定期更新或刷新，以防止令牌被滥用；
- c) 如果验证失败，应向服务接口使用方返回明确的出错消息和代码；
- d) 接口将记录所有身份验证相关的活动，包括成功的登录、失败的尝试、令牌的生成和销毁等。

6.1.2 输入与输出

- a) 输入包括：用户名和密码以及其他认证因素，如手机验证码等；
- b) 输出包括：验证结果，验证成功输出令牌，验证失败输出出错消息和代码。

6.2 资源对象检索接口

6.2.1 接口要求

资源对象检索接口用于检索版本典藏网站资源长期保存系统中符合条件的资源对象。资源对象包括三种：知识对象实体、表现实体、文件实体。

接口要求包括：

- a) 应支持对三种资源对象公开的元数据的检索，检索包括：单词检索、多词检索、通配符检索、范围检索等；
- b) 检索结果中应包含命中记录的主要元数据，如对象类别、标识符、名称等；
- c) 应支持对检索结果的特定元数据排序和过滤；
- d) 对于大量结果的检索，接口应支持分页显示，并允许设定返回结果的数量限制；
- e) 使用本接口前应获得系统认证和授权，只有经过身份验证和授权才能访问和使用接口。

6.2.2 输入与输出

- a) 输入包括：资源对象、检索词、检索字段、分页要求；
- b) 输出包括：结果数量、命中资源对象的主要元数据。

6.3 资源对象详情获取接口

6.3.1 接口要求

资源对象详情获取接口用于获取版本典藏网站资源长期保存系统中某个特定资源对象的详细信息。资源对象包括三种：知识对象实体、表现实体、文件实体。

接口要求包括：

- a) 输入参数中应含有能唯一确定某个资源对象的标识符；
- b) 接口返回的详细信息应以结构化的数据格式提供；
- c) 如果输入对象的标识符无法获取到详情，应向服务接口使用方返回明确的出错消息和代码；
- d) 使用本接口前应获得系统认证和授权，只有经过身份验证和授权才能访问和使用接口。

6.3.2 输入与输出

- a) 输入包括：资源对象、标识符；
- b) 输出包括：如命中则输出资源对象详细元数据，否则输出出错消息和代码。

6.4 资源对象提取接口

6.4.1 接口要求

资源对象提取接口用于从版本典藏网站资源长期保存系统中提取特定资源对象的数据。资源对象包括两种：知识对象实体、表现实体、文件实体。

接口要求包括：

- a) 输入参数中应含有能唯一确定某个资源对象的标识符；
- b) 应能够稳定、高效地连接到数据源，如数据库、文件系统、API；
- c) 应清晰地定义如何从数据源中提取资源对象；
- d) 应能够妥善处理各种错误情况，如数据源连接失败、数据格式错误等；
- e) 应提供适当的容错机制，如重试、回退、断点续传等；
- f) 使用本接口前应获得系统认证和授权，只有经过身份验证和授权才能访问和使用接口。

6.4.2 输入与输出

- a) 输入包括：资源对象、标识符；
- b) 输出包括：资源对象或下载资源对象的地址。

6.5 保存实体检索接口

6.5.1 接口要求

保存实体检索接口用于检索版本典藏网站资源长期保存系统中符合条件的实体对象。实体对象包括：资源对象实体、活动实体、权利实体、责任者实体、软件实体和环境实体。

接口要求包括：

- a) 应支持对各类保存实体对象的公开元数据的检索，检索包括：单词检索、多词检索、通配符检索、范围检索等；
- b) 检索结果中应包含命中记录的主要元数据，如对象类别、标识符、名称等；
- c) 应支持对检索结果的特定元数据排序和过滤；
- d) 对于大量结果的检索，接口应支持分页显示，并允许设定返回结果的数量限制；
- e) 使用本接口前应获得系统认证和授权，只有经过身份验证和授权才能访问和使用接口。

6.5.2 输入与输出

- a) 输入包括：实体类型、检索词、检索字段、分页要求；
- b) 输出包括：结果数量、命中实体的主要元数据。

6.6 保存实体详情获取接口

6.6.1 接口要求

保存实体检索接口用于获取版本典藏网站资源长期保存系统中某个特定实体对象的详细信息。实体对象包括：资源对象实体、活动实体、权利实体、责任者实体、软件实体和环境实体。

接口要求包括：

- a) 输入参数中应含有能唯一确定某个资源对象的标识符；
- b) 接口返回的详细信息及实体间的关系应以结构化的数据格式提供；
- c) 如果输入对象的标识符无法获取到详情，应向服务接口使用方返回明确的出错消息和代码；
- d) 使用本接口前应获得系统认证和授权，只有经过身份验证和授权才能访问和使用接口。

6.6.2 输入与输出

- a) 输入包括：实体类型、标识符；
- b) 输出包括：如命中则输出实体详细元数据，否则输出出错消息和代码。

6.7 网站回放接口

6.7.1 接口要求

网站回放接口是版本典藏网站资源长期保存系统存储的网站对象实体整体发布服务，可以回溯发布过去某个时间点的网站，重现网站的状态和内容，查看当时的页面内容、布局、交互等。

接口要求包括：

- a) 输入参数中应含有能唯一确定某个网站对象的标识符和时间节点；
- b) 提供高效的数据检索机制，使用户能够方便地找到并回放特定时间点的网站版本，回放功能应能够准确地呈现当时网站的外观和交互体验；
- c) 设计健壮的错误处理机制，能够在网站回放过程中遇到问题时进行妥善处理，并提供相应的错误提示和解决方案；
- d) 使用本接口前应获得系统认证和授权，只有经过身份验证和授权才能访问和使用接口。

6.7.2 输入与输出

- a) 输入包括：知识对象实体或表现实体的标识符；
- b) 输出包括：回放地址。

参 考 文 献

- [1] GB/T 29829-2022 信息安全技术 可信计算密码支撑平台功能与接口规范
 - [2] DB52T 1541.4-2021 政务数据平台 第4部分：接口规范
 - [3] T/CCUA 017-2021 政务数据共享平台 服务接口要求
-