

团体标准

T/CESA XXXX—202X

标准信息服务平台 数据库建设要求

Standard information service platform — Database construction requirements

征求意见稿

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。

已授权的专利证明材料为专利证书复印件或扉页，已公开但尚未授权的专利申请证明材料为专利公开通知书复印件或扉页，未公开的专利申请的证明材料为专利申请号和申请日期。

202X-XX-XX 发布

202X-XX-XX 实施

中国电子工业标准化技术协会 发布



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

目 次

前 言.....	III
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 数据库构成.....	1
5 总体要求.....	2
5.1 系统要求.....	2
5.2 命名要求.....	2
6 标准管理数据库.....	2
6.1 构建过程.....	2
6.2 数据表设计.....	3
6.2.1 标准文件管理数据.....	3
6.2.2 制修订过程数据.....	3
7 标准内容数据库.....	3
7.1 构建过程.....	3
7.2 数据表设计.....	4
7.2.1 标准结构化内容数据.....	4
7.2.2 标准知识数据.....	4
8 产业链数据库.....	5
8.1 构建过程.....	5
8.2 数据表设计.....	5
9 标准评价数据库.....	5
9.1 构建过程.....	5
9.2 数据表设计.....	6
附 录 A（规范性） 标准信息服务平台数据库表结构设计.....	7

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。

本文件由之江实验室提出。

本文件由中国电子技术标准化研究院、中国电子工业标准化技术协会归口。

本文件起草单位： 。

本文件主要起草人： 。

标准信息服务平台 数据库建设要求

1 范围

本文件规定了标准信息服务平台数据库的构成、总体要求以及标准管理数据库、标准内容数据库、产业链数据库、标准评价数据库的具体构建要求。

本文件适用于标准信息服务平台数据库的设计和建设。

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。

GB/T 1.1 标准化工作导则 第1部分：标准化文件的结构和起草规则

GB/T 16733 国家标准制定程序的阶段划分及代码

GB/T 20273 信息安全技术 数据库管理系统安全技术要求

GB/T 22373 标准文献元数据

3 术语和定义

下列术语和定义适用于本文件。

3.1

标准信息服务平台 `standard information service platform`

运用高性能计算、大数据、区块链、知识计算技术，提供标准智能检索、标准对比分析、产业链用标情况分析等功能，满足标准推广与应用、标准实施效果分析与评价、产业链用标分析与决策等标准数字化应用的服务平台。

3.2

标准主题词 `standard subject term`

描述标准内容的关键词。

4 数据库构成

图1给出了标准信息服务平台的数据库构成，其主要组成部分为标准管理数据库、标准内容数据库、产业链数据库、标准评价数据库。

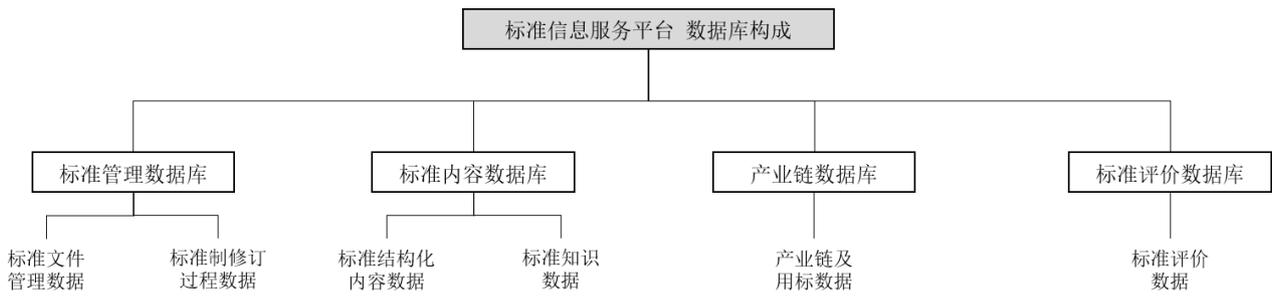


图1 标准信息服务平台数据库构成

- 标准管理数据库：包含标准文件管理数据和标准制修订过程数据，支撑标准信息服务平台标准文件的更新管理、上传下载和标准制修订过程信息及文件的查看、下载等基础功能。
- 标准内容数据库：包含标准结构化内容和标准知识数据，支撑标准信息服务平台对于标准全文内容的检索、智能问答、图谱探索等功能。
- 产业链数据库：包含产业链及用标数据，支撑标准信息服务平台的产业链数据查询、用标分析等功能。
- 标准评价数据库：包含标准评价相关数据，支撑标准信息服务平台的标准成熟度评价、标准实施效果评价等功能。

5 总体要求

5.1 系统要求

- 5.1.1 应面向业务主题建设独立于应用程序的数据库，根据需求支持不同语言及文字、符号、公式、图形、图像等多种形式数据的存储。
- 5.1.2 应具备安全可靠的基础设施与环境，包括服务器、存储硬件设备、软件环境、网络环境等。
- 5.1.3 应具备支持元数据管理、数据管理、安全管理、数据交换等功能的数据库管理系统。
- 5.1.4 数据库系统的建设应遵循 GB/T 20273 中对信息系统建设的安全技术规定。
- 5.1.5 数据库系统应具备易维护性、可扩展性，能够根据需求的变化维护、修改或扩充系统功能。

5.2 命名要求

- 5.2.1 数据库名称应体现数据库的数据内容，名称宜由大写英文字母、下划线组成；
- 5.2.2 数据表名称应体现表中的数据内容，命名简短、明确，名称宜由大写英文字母、下划线组成；
- 5.2.3 字段名应作为表中该字段的唯一标识并体现字段的含义，名称宜由大写英文字母、0-9 数字、下划线组成，字段名称不宜超过四个英文单词。

6 标准管理数据库

6.1 构建过程

标准管理数据库构建过程见图2，输入的标准相关原始文件上传后形成文件管理数据，标准制修订信息通过文件上传处理或外部接口接入数据，经制修订管理形成制修订过程数据。

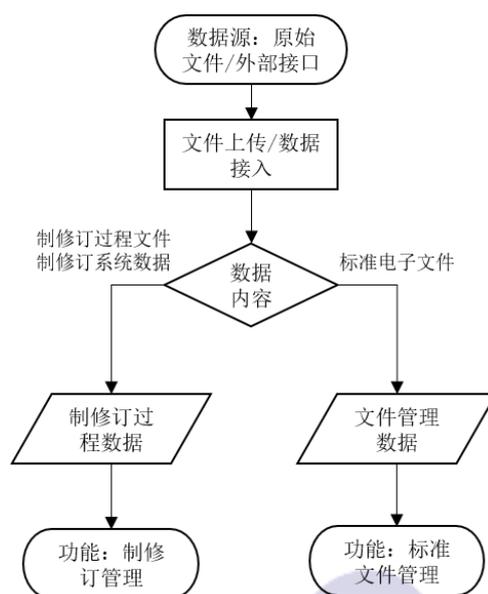


图2 标准管理数据库构建过程

6.2 数据表设计

6.2.1 标准文件管理数据

标准文件管理数据主要数据表应包括：

- a) 标准文件基本信息表，记录标准原始电子文件信息；
- b) 标准文件页表，记录标准文件的分页信息。

数据表具体字段设计参考附录A.1至A.2。

6.2.2 制修订过程数据

制修订过程数据主要数据表应包括：

- a) 制修订过程流转信息表：记录标准的制修订阶段状态、各阶段对应的时间段、专家等相关信息；
- b) 标准状态信息表：记录发布标准的当前状态，以及与其他标准之间的代替关系。
- c) 标准化技术组织信息表：记录组织基本信息、秘书处信息等技术组织基本信息；
- d) 标准化专家信息表：记录专家基本信息；
- e) 标准体系信息表：记录标准体系对应的基本信息、体系结构。

各数据表具体字段设计参考附录A.3至A.7。

7 标准内容数据库

7.1 构建过程

标准内容数据构建过程见图3，标准电子文件经过结构化处理，形成标准结构化的内容数据，基于结构化内容数据进一步做知识抽取、标注和图谱构建等，形成标准知识数据。

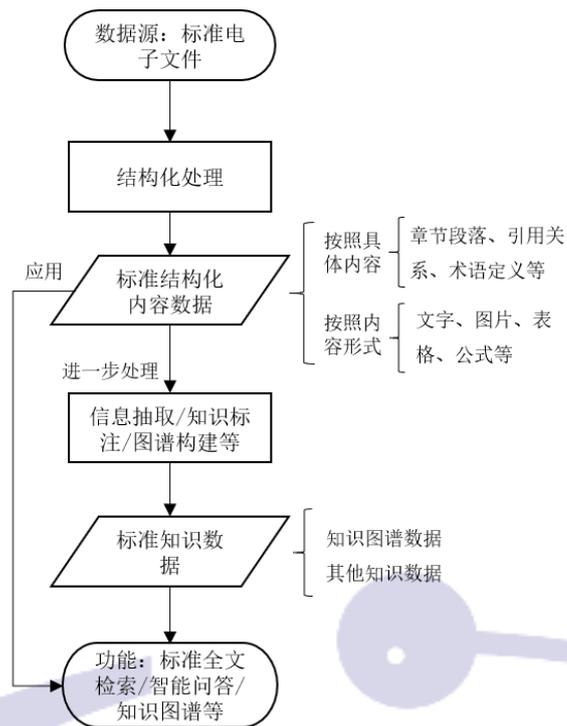


图3 标准内容数据库构建过程

7.2 数据表设计

7.2.1 标准结构化内容数据

标准结构化内容数据应包括：

- a) 标准结构化基本信息数据，其主要数据表应包括：
 - 1) 标准基本信息表：记录标准文献基本属性信息，应至少包括 GB/T 22373-2021 中定义的标准文献核心元数据，宜包括较常用的标准文献公共元数据；
 - 2) 图片基本信息表：记录标准中图片的基本信息、段落位置信息、存储信息等；
 - 3) 表格基本信息表：记录标准中表格的基本信息、段落位置信息、存储信息等；
 - 4) 公式基本信息表：记录标准中公式的基本信息、段落位置信息、存储信息等。
- b) 标准结构化加工后的内容数据，其主要数据表应包括：
 - 1) 标准全文内容表：记录标准按照标准编写要素以及章节结构划分的全文内容，数据表设计应考虑 GB/T 1.1-2020 6.2.2.2 中阐述的标准构成要素，包括必备要素和可选要素。
 - 2) 术语定义表：记录结构化加工后的术语及定义内容；
 - 3) 标准引用关系表：记录结构化加工后的标准引用其他标准的情况；
 - 4) 表格内容表：记录结构化加工后的表格的数据内容。

各数据表具体字段设计参考附录A.8至A.15。

7.2.2 标准知识数据

标准知识数据是由标准内容进一步处理形成的知识化数据，应包括知识图谱数据。图谱数据一般由图谱的基本信息、本体层数据和实例层数据组成，其中图谱基本信息和本体层数据宜采用关系型数据库存储，主要数据表应包括：

- a) 图谱基本信息表：记录标准图谱信息；
- b) 实体类型信息表：记录标准知识图谱本体模型中实体类型的基本信息；
- c) 实体类型属性表：记录标准知识图谱本体模型中实体类型的属性信息；
- d) 实体类型关系表：记录标准知识图谱本体模型中实体类型之间的关系信息。

各数据表具体字段设计参考附录A. 16至A. 19。

图谱实例层数据宜采用图数据库或多种数据库混合存储，满足图谱数据的高效存储和查询需求。若使用关系型数据库，其数据表应参照本体层，设计对应的实体表、实体关系表、实体属性表等。

面向标准信息服务平台的信息查询和知识分析需要，宜包括标准主题词信息表、标准问答系统(FAQ)信息表、标准文本主谓宾(SPO)信息表等标准知识数据表。

8 产业链数据库

8.1 构建过程

产业链数据库构建过程见图4，产业链、企业、标准基础信息经过数据收集、关系梳理后形成产业链及用标数据。

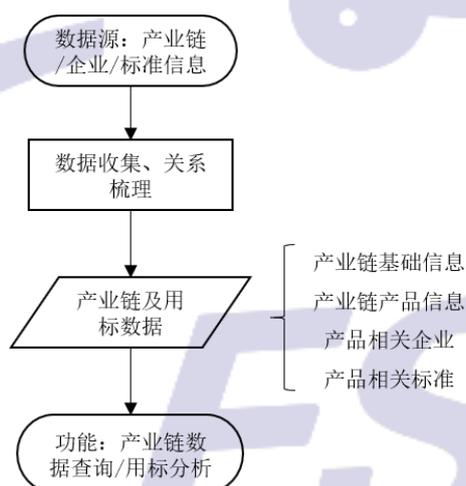


图4 产业链数据库构建过程

8.2 数据表设计

产业链数据主要数据表应包括：

- a) 产业链基本信息表：记录产业名称、描述、对应的产业链图谱编号等基础信息；
- b) 产业链产品信息表：记录产业链中各环节和产品的树形层级关系；
- c) 产品相关企业表：记录产品对应的企业清单；
- d) 产品相关标准表：记录产品对应的标准清单。

各数据表具体字段设计参考附录A. 20至A. 23。

宜包括企业基本信息表：记录产业链相关业的企业概况、经营情况等基本信息。

9 标准评价数据库

9.1 构建过程

标准评价数据库构建过程见图5, 基于标准基本信息和评价参考信息进行指标设计和评价数据采集, 形成标准评价指标和评价结果等标准评价数据。

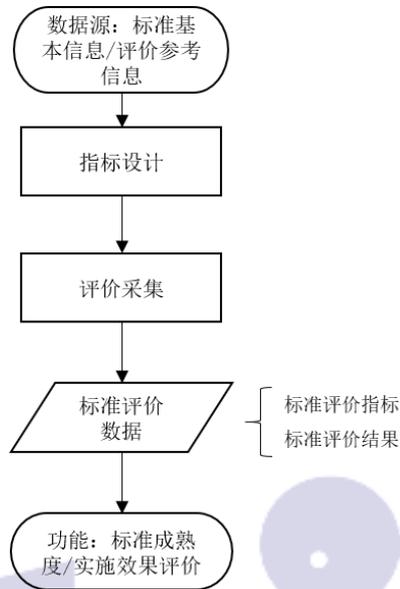


图5 标准评价数据库构建过程

9.2 数据表设计

标准评价数据应包括但不限于:

- a) 标准评价指标表: 记录每个标准与评价指标的对应关系;
- b) 评价指标体系表: 记录多级指标体系之间的从属关系;
- c) 标准评价结果表: 记录标准评价的结果数据。

各数据表具体字段设计参考附录A. 24至A. 26.

附 录 A
(规范性附录)

标准信息服务平台数据库表结构设计

标准文件基本信息表字段设计如表A.1所示，数据表设计时应考虑以下方面：

- a) 考虑标准文件大小和数量，有较高存储、性能要求时宜采用文件系统服务存储文件原件；
- b) 标准文件在磁盘或文件系统的存储采用标准文件唯一ID命名；
- c) 根据使用需求考虑是否采用加密方法；
- d) 当存在同时修改同一文件信息时，考虑使用文件锁。

表A.1 STANDARDS_FILE_INFO 标准文件基本信息表

字段名称	数据类型	约束/条件	字段说明
FILE_UUID	字符型	必选	标准文件唯一 ID
FILE_VERSION	字符型	必选	标准文件版本
ORIGINAL_NAME	字符型	必选	标准文件名
CONTENT_TYPE	字符型	必选	标准文件类型
STANDARD_NO	字符型	必选	标准编号
STORE_TYPE	数值型	必选	存储类型（0：本地磁盘，1：DFS 文件系统）
STORE_PATH	字符型	必选	标准文件存储位置
LENGTH	数值型	可选	标准文件实际大小（单位：字节）
STORE_IP	字符型	可选	标准文件存储节点 IP
LAST_MODIFIED_TIME	日期型	可选	标准文件最后修改时间
FROM_IP	字符型	可选	标准文件来源 IP

标准文件中页表字段设计如表A.2所示，其中页内容根据使用需求考虑是否采用加密方法。

表A.2 STANDARDS_MATERIAL_PAGE 标准文件页表

字段名称	数据类型	约束/条件	字段说明
FILE_UUID	字符型	必选	标准文件唯一 ID
FILE_VERSION	字符型	必选	标准文件版本
PAGE_UUID	字符型	必选	本页内容唯一标识
PAGE_NO	数值型	必选	页号
CONTENT_TEXT	字符型	可选	页内容
CONTENT_WORD_COUNT	数值型	可选	页字数
CONTENT_SENTENCE_COUNT	数值型	可选	页句子的数量

制修订过程流转信息表字段设计如表A.3所示：

表A.3 STANDARDS_REVERSION_PROCESS_INFO 制修订过程流转信息表

字段名称	数据类型	约束/条件	字段说明
FILE_UUID	字符型	必选	标准文件唯一 ID
FILE_VERSION	字符型	必选	标准文件版本
STANDARD_NO	字符型	必选	标准编号
PRESENT_REVERSION_NO	字符型	必选	当前所处制修订阶段编号
PRESENT_REVERSION_NAME	字符型	必选	当前所处制修订阶段名称
REVERSION_NO	字符型	必选	制修订阶段编号
REVERSION_NAME	字符型	必选	制修订阶段名称
START_DATE	日期型	可选	开始时间
END_DATE	日期型	可选	结束时间
EXPERT_LIST	字符型	可选	制修订过程对应专家列表

根据GB/T 16733-1997，国家标准的制修订阶段编号和名称对应数据字典如下：

表A.3.1 国家标准制修订阶段数据字典

制修订阶段编号	制修订阶段名称
00	预阶段
10	立项阶段
20	起草阶段
30	征求意见阶段
40	审查阶段
50	批准阶段
60	出版阶段
90	复审阶段
95	废止阶段

标准状态信息表字段设计如表A.4所示：

表A.4 STANDARDS_STATUS_INFO 标准状态信息表

字段名称	数据类型	约束/条件	字段说明
FILE_UUID	字符型	必选	标准文件唯一 ID
FILE_VERSION	字符型	必选	标准文件版本
STANDARD_NO	字符型	必选	标准编号
STANDARD_STATUS_NO	字符型	必选	标准状态编号
STANDARD_STATUS_NAME	字符型	必选	标准状态名称
IS_REPLACE	逻辑型	必选	是否代替其他标准
REPLACED_STANDARD_NO	字符型	可选	被代替标准的标准编号

标准技术组织信息表字段设计如表A.5所示：

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/068006011103007005>