

ICS 35.080

CCS L77

团 体 标 准

T/ZADT XXXX—XXXX

T/ZXCH XXXX—XXXX

供应链企业业务数据管理平台技术规范

Technical Specification for Business Data Management Platform for Supply
Chain enterprises

（征求意见稿）

XXXX-XX-XX 发布

XXXX-XX-XX 实施

浙江省国际数字贸易协会 发布
浙江省信息产业质量协会

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 基本要求 2

6 分类组织方式 3

7 数据获取 3

8 数据存储 4

9 数据服务 7

附录 A（资料性） 数据组织分类方式 9

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。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由物产中大数字科技有限公司提出，由浙江省国际数字贸易协会和浙江省信息产业质量协会联合发布、归口、实施。

本文件起草单位：。

本文件主要起草人：。

供应链企业业务数据管理平台技术规范

1 范围

本文件提供了供应链企业业务数据管理平台的基本要求、数据分类组织方式、数据获取要求、数据存储要求和数据服务要求。

本文件适用于供应链企业业务数据管理平台建设，为相关企业提供建设参考。

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 35295-2017 信息技术 大数据 术语

GB/T 36073-2018 数据管理能力成熟度评估模型

GB/T 42381.120-2023 数据质量 第 120 部分：主数据：特征数据交换：溯源性

GB/T 42381.130-2023 数据质量 第 130 部分：主数据：特征数据交换：准确性

GB/T 42381.140-2023 数据质量 第 140 部分：主数据：特征数据交换：完整性

3 术语和定义

下列术语和定义适用于本文件。

3.1

业务数据 business data

企业经营管理决策所依据的数据，通过对其深入分析和理解，更准确地把握市场脉搏，优化运营策略，实现更高效的资源配置和更快速的业务增长。

3.2

内部数据 internal data

企业运营过程中产生的数据，包括业务系统、运营流程以及管理活动产生的数据。

3.3

外部数据 external data

在组织或企业运营范围之外生成的数据，包括任何与组织外部环境相关的数据，如竞争对手经营情况、证券市场数据、行业数据等。

3.4

半结构化数据 semi-structured data

具有预定义的数据模型，但具有某种形式的层次或嵌套结构。

3.5

非结构化数据 non-structured data

数据结构不规则或不完整，没有预定义的数据模型，不方便用数据库二维逻辑表来表现的数据。包

括所有格式的办公文档、文本、图片、HTML、各类报表、图像和音频/视频信息等等。

3.6

数据对象 data object

具有复杂结构且属于同一主题的数据集合。

3.7

定性文本数据 qualitative text data

用于描述事物的文本。

3.8

定量文本数据 quantitative text data

从文本中实体抽取的指标名称、指标度量值、指标计量单位、指标计量货币构成的数组。

3.9

标准化数据 standardized data

将数据按照一定的规则进行归并、清洗、统一、去重等处理后的数据。

3.10

数据转换 data transformation

将非结构化数据解析为半结构化数据对象的过程。

3.11

数据融合 data fusion

将多个来源的异构数据按同一主题汇聚，形成同一主题的二维表结构的过程。

4 缩略语

下列缩略语适用于本文件。

ESG	Environmental, Social, and Governance	环境、社会和公司治理
PDF	Portable Document Format	可移植文档格式
HTML	Hyper Text Markup Language	超文本标记语言
Web	Website	网站
ERP	Enterprise Resource Planning	企业资源计划
CRM	Customer Relationship Management	客户关系管理系统
MES	Manufacturing Execution System	制造执行系统
WMS	Warehouse Management System	仓储管理系统
OA	Office Automatic	办公管理系统

5 基本要求

5.1 数据溯源

宜提供数据溯源分析和展示，支持采集数据、转换数据、融合数据和标准化数据之间流向间的相互溯源。

5.2 数据检索

应采用关系型数据库管理数据，具备将非结构化数据转换为二维表结构化数据的功能，实现采集的 PDF 文档、图片、HTML、XML、Excel、Word 等文档型的非结构化数据按表格、文本和图片的数据对象转换成二维表结构化数据的功能，支持表格对象数据按其主题融合到同一二维表、图片数据对象和文本数据对象标签类别管理，具备从文本数据对象中抽取实体及实体关系的能力，满足数据的快速检索和数据资源化要求。

5.3 数据规范

应满足代码统一、字段统一和格式统一要求，缺失数据统一用空值补全，并且标准化后的数值型数据要统一度量数据所属组织、数据名称、数据时间格式及其度量单位。

5.4 数据安全

采集外部数据至少要遵守采集网站等数据源的 Robots 协议、爬虫检测、加固 Web 站点等要求，采集的数据范围和频率等应遵守最小化的使用原则，不得妨碍被采集数据源系统的正常运行。

应将内外部数据隔离管理，满足内部数据安全要求。

应充分考虑系统的备份及出现各类故障时的恢复能力，以及恢复时间，满足数据服务及时性要求。

6 分类组织方式

6.1 采集数据

6.1.1 内部数据按管理域细分为战略管理、组织管理、运营管理、财务管理、科技创新、风险管理、人力资源、科技创新和信息化，详细见附录 A；按数据来源细分为企业资源管理系统（ERP）、客户关系管理系统（CRM）、生产管理系统（MES）、仓储管理系统（WMS）、财务管理系统、人力资源系统、办公自动化系统（OA）等。

6.1.2 外部数据按采集数据的主题细分为公开披露的报告信息、证券市交易市场数据、行业研究报告、商品价格与供需信息、新闻舆情、人力招聘等。

6.2 转换数据

按转换后的数据对象分为表格、图片、文本。其中文本可细分为定性文本数据和定量文本数据。

6.3 融合数据

6.3.1 表格数据对象按主题细分为资产负债表、利润表、现金流量表、主要会计数据、主要财务指标、分季度主要财务数据、非经常性损益项目和金额、采用公允价值计量的项目、主营业务分行业情况、产销量情况分析表、研发人员情况表、以套期保值为目的的衍生品投资、主要控股参股公司分析、员工情况等。

6.3.2 图片和文本数据对象按管理域细分为战略管理、组织管理、运营管理、财务管理、科技创新、风险管理、人力资源、科技创新和信息化，详细见附录 A。

6.3.3 定量文本数据按管理域分为经营数据、证券市场数据、行业数据、招聘数据等。

6.4 标准化数据

表格数据对象和定量文本数据的主题分别继承并融合了相应表格数据对象及定量文本数据的主题，并在库表级别的命名规范上采用标准化标识符以区别于融合数据库表。

7 数据获取

7.1 数据采集

7.1.1 采集方式

数据采集方式分为系统采集和人工采集。

a) 系统采集：支持关系型数据库、半结构化数据库、分布式文件数据、对象存储、时序数据库等多种类型数据源的配置化系统采集，具备接口对接和爬虫采集的配置管理功能。

b) 人工采集：支持录入、导入等方式采集数据。

7.1.2 采集对象

采集对象包括内外部数据。

a) 内部采集的数据应包括但不限于以下内容：

收集企业内部的战略、组织、运营、财务、风险、人力、科技和信息等领域数据，应覆盖采购、销售、库存、仓储、物流、资金等模块，包含结构化数据和非结构化数据，其中非结构化数据应包括但不限于各流程的附件文件、各系统中的日志文件和各系统中的图像文件。

b) 外部采集的数据应包括但不限于以下内容：

- 1) 上市公司公开披露的财务报告、ESG 报告、内部控制报告等 PDF 文档；
- 2) 证券交易市场的股票价格、市值、市盈率、市净率等指标构成的 HTML 文件；
- 3) 证券公司等投资咨询机构发布的行业研究报告等 PDF 文档；
- 4) 行业的商品价格、消费指数、开工率、库存量、进出口、物流价格、仓储成本等第三方接口 JSON 数据或 CSV 数据；
- 5) 公司官网、招聘网站的招聘职位、职责、薪资范围等组成的 HTML 文件；
- 6) 主流媒体报道的企业或竞争者密切相关的新闻报道、产业政策等新闻舆情。

7.2 数据整合

7.2.1 数据转换

应具备数据转换功能，支持从不同结构类型的数据里面抽取数据，转换为表格、图片和文本等数据对象，形成半结构化数据，并以结构化数据存储，满足数据资源化要求。

7.2.2 数据融合

应具备将多源异构转换抽取得到的表格对象数据和定量文本数据按同一主题进行融合汇聚；图片数据对象和定性文本数据按领域进行融合，形成不同主题或领域的知识库，满足数据的快速检索要求。

7.2.3 数据标准化

宜具备标准创建、标准修改、标准删除、标准查询、标准实施等功能。标准化包括但不限于统一以二维表结构融合入库多源表格数据对象和数值型文本中的数据指标名称、数据格式、计量单位、货币单位，满足多源异构数值型数据的可比性，易用用户开展数据的挖掘分析。

7.3 数据核验

数据核验宜提供监控数据采集、数据转换、数据融合、数据标准化等任务执行情况及其执行结果数据质量的核验功能，支持任务人工补调任务和数据原位核验、修改、补录，并记录相关的核验信息，满足数据的完整性、准确性、规范性和一致性要求。

8 数据存储

8.1 存储方式

宜采用关系型数据库对数据进行存储，记录数据间的流向关系，并以非结构化数据存储 URL 或路径与非结构化数据间建立联系。

8.2 存储策略

8.2.1 数据分层

应采取数据分层策略将数据分为采集数据层、转换数据层、数据融合层、数据标准化层、数据服务

层。

8.2.2 主题建库

将数据按第6节的分类组织方式，在不同的数据层划分成若干相对独立的主题域进行存储，形成满足特定场景需求的主题库。

8.2.3 按类汇聚

宜采取根据数据类别的完整性和独立性等分类原则汇聚多源异构数据，满足数据的快速检索要求，支持数据库的可读性和可维护性。

8.3 数据模型

8.3.1 采集数据

采集数据存储的数据模型包括其属性信息、数据源信息、主体信息、状态信息，数据项包括但不限于下表1所示。

表1 采集数据的数据模型

数据项	数据类型	必填项	备注
采集时间	日期型	是	格式为 yyyy-MM-dd hh:mm:ss
数据描述主体	字符型	是	按描述主体分为企业名称、商品名称等
数据来源	字符型	是	按来源分为接口名称、系统名称、人工导入等
数据类型	枚举值	是	按类型分为数据库、JSON、PDF、HTML 等
数据对象	字符型	是	采集的目标数据对象编码
对象存储路径	字符型	是	采集的目标数据对象存储的文件路径
数据核验状态	枚举值	是	分为未核验和已核验
数据转换状态	枚举值	是	分为未转换和已转换

8.3.2 转换数据

8.3.2.1 表格数据

转换得到的表格数据对象的数据模型包括其属性信息、位置信息、主题信息、主体信息和状态信息，数据项包括但不限于表2所示。

表2 转换表格数据的数据模型

数据项	数据类型	必填项	备注
数据对象名称	字符型	是	数据对象的名称，如 PDF 文档的名称
数据描述主体	字符型	是	数据描述的企业或商品等主体的名称
日期	日期型	是	数据描述主体的日期
表格所在页码	整数型	是	描述表格在 多页文档类型数据对象的页码
表格顺序	整数型	是	描述表格所在页面的顺序
表格起始行	整数型	是	描述表格在页面的起始位置行
表格终止行	整数型	是	描述表格在页面的终止行
表格区域信息	字符型	是	描述表格在页面区域的数组字符串 [x0, y0, x1, y1]
表格主题关键区域	字符型	是	描述表格主题信息的关键区域数据
表格标题	字符型	否	从表格的关键区域抽取到的表格标题
表格计量单位	字符型	否	从表格的关键区域抽取到的表格计量单位
表格货币单位	字符型	否	从表格的关键区域抽取到的表格货币单位
表格主体数据	字符型	是	表格的主体数据 ArrayList 字符串
数据核验状态	枚举值	是	分为未核验和已核验
数据融合状态	枚举值	是	分为未融合和已融合

8.3.2.2 图片数据

转换得到的图片数据对象的数据模型包括其属性信息、位置信息、主体信息和状态信息，数据项包括但不限于表 3 所示。

表 3 转换图片数据的数据模型

数据项	数据类型	必填项	备注
数据对象名称	字符型	是	数据对象的名称，如 PDF 文档的名称
数据描述主体	字符型	是	数据描述的企业或商品等主体的名称
日期	日期型	是	数据描述主体的日期
图片所在页码	整数型	是	描述图片在多层文档类型数据对象的页码
图片顺序	整数型	是	描述图片所在页面的顺序
图片区域信息	字符型	是	描述图片在页面区域的数组字符串[x0, y0, x1, y1]
图片上文	字符型	否	图片的上文文本
图片下文	字符型	否	图片的下文文本
图片主题关键区域	字符型	是	描述图片主题信息的关键区域数据
图片标题	字符型	否	描述图片的标题信息
图片对象编码	字符型	是	描述图片对象数据的编码
图片主题	枚举值	是	图片分类管理的主题标签
对象存储路径	字符型	否	转换图片数据对象存储的文件路径
数据核验状态	枚举值	是	分为未核验和已核验
数据融合状态	枚举值	是	分为未融合和已融合

8.3.2.3 定性文本数据

定性文本数据对象的数据模型包括其属性信息、位置信息和主体数据，数据项包括但不限于表 4 所示。

表 4 转换定性文本数据的数据模型

数据项	数据类型	必填项	备注
数据对象名称	字符型	是	数据对象的名称，如 PDF 文档的名称
数据描述主体	字符型	是	数据描述的企业或商品等主体的名称
日期	日期型	是	数据描述主体的日期
文本所在页码	整数型	是	描述文本在多层文档类型数据对象的页码
文本起始行	整数型	是	描述文本在页面的起始位置行
文本终止行	整数型	是	描述文本在页面的终止行
文本区域信息	字符型	是	描述文本在页面区域的数组字符串[x0, y0, x1, y1]
文本主体数据	字符型	是	文本的主体数据
文本数据编码	字符型	是	文本的主体数据编码
文本主题	枚举值	是	文本分类管理的主题标签
数据核验状态	枚举值	是	分为未核验和已核验
数据融合状态	枚举值	是	分为未融合和已融合

8.3.2.4 定量文本数据

定量文本数据对象的数据模型包括其属性信息、位置信息和主体数据，数据项包括但不限于表 5 所示。

表 5 转换定量文本数据的数据模型

数据项	数据类型	必填项	备注
数据对象名称	字符型	是	数据对象的名称，如 PDF 文档的名称
数据描述主体	字符型	是	数据描述的企业或商品等主体的名称
日期	日期型	是	数据描述主体的日期
文本数据编码	字符型	是	描述定量文本数据的来源文本数据的编码
文本主体数据	字符型	是	描述定量文本数据的来源文本数据
指标名称	字符型	是	从文本中抽取的指标名称
指标数值	数值型	是	指标的数值

指标计量单位	字符型	是	指标的计量单位
指标货币单位	字符型	否	指标的货币单位
数据核验状态	枚举值	是	分为未核验和已核验
数据融合状态	枚举值	是	分为未融合和已融合

8.3.3 融合数据

融合数据模型包括其属性信息、主体信息、来源信息和状态信息，数据项包括但不限于表 6 和表 7 所示。

8.3.3.1 表格和定量文本数据项包括但不限于 6。

表 6 表格和定量文本数据的数据模型

数据项	数据类型	必填项	备注
数据描述主体	字符型	是	数据描述的企业或商品等主体的名称
日期	日期型	是	数据描述主体的日期
指标名称	字符型	是	从文本中抽取的指标名称
指标数值	数值型	是	指标的数值
指标计量单位	字符型	是	指标的计量单位
指标货币单位	字符型	否	指标的货币单位
数据来源表	字符型	是	指标来源的数据库表名称
数据来源表行	整数型	是	指标来源的数据库表的行 ID
数据核验状态	枚举值	是	分为未核验和已核验
数据标准化状态	枚举值	是	分为未标准化和已标准化

8.3.3.2 图片和定性文本数据项包括但不限于表 7。

表 7 转换定量文本数据的数据模型

数据项	数据类型	必填项	备注
数据描述主体	字符型	是	数据描述的企业或商品等主体的名称
日期	日期型	是	数据描述主体的日期
数据编码	字符型	是	文本或图片的数据编码
数据主题	数值型	是	文本或图片的主题名称
数据来源表	字符型	是	文本或图片来源的数据库表名称
数据来源表行	整数型	是	文本或图片来源的数据库表的行 ID
数据核验状态	枚举值	是	分为未核验和已核验
数据标准化状态	枚举值	是	分为未标准化和已标准化

8.3.4 标准化数据

标准化数据模型包括其属性信息、主体信息和状态信息，数据项包括但不限于表 8 所示。

表 8 标准化数据的数据模型

数据项	数据类型	必填项	备注
数据描述主体	字符型	是	数据描述的企业或商品等主体的名称
日期	日期型	是	数据描述主体的日期
标准指标名称	字符型	是	指标标准化后的名称
指标名称	字符型	是	指标的原名称
指标数值	数值型	是	指标的数值
指标计量单位	字符型	是	指标的计量单位
指标货币单位	字符型	否	指标的货币单位
数据来源表	字符型	是	指标来源的表名称
数据来源表行	整数型	是	指标来源的表的行 ID
数据核验状态	枚举值	是	分为未核验和已核验

9 数据服务

9.1 数据资源目录

应具备数据资源目录编目功能，实现数据的目录式管理，支持数据按主题、数据描述主体、数据项、数据范围、数据对象等方式检索数据，实现数据资源的统一管理，为运营决策者等提供数据资源清单。

数据资源目录可具备应用系统注册、目录编制、目录审核与发布、资源挂载、目录查询、目录变更与停用等功能，数据资源目录至少应包括采集数据资源目录、转换数据资源目录、融合数据资源目录和标准化数据资源目录。

9.2 数据展示

供应链企业业务数据管理平台应提供以下数据展示功能：

- a) 标准化数据的数据集统计性描述；
- b) 标准化数据的在线浏览；
- c) 标准化数值型数据的折线图、柱状图、雷达图、饼图等可视化图表展现；
- d) 采集数据、转换数据、融合数据和标准化数据的溯源展示和分析。

9.3 数据调用

应支持数据调用申请、调用审核、数据准备、数据分发和数据反馈的闭环服务管理流程，可根据需求在数据资源目录选择采集数据、转换数据、融合数据及标准化数据。数据调用形式支持导出数据和接口调用。第三方系统通过认证 token 后数据访问接口调取数据，实行可追溯管理。宜建立调用数据反馈机制，数据使用者及时反馈数据质量等问题。

9.4 数据需求

应具备数据需求收集功能，支持数据采集、转换、抽取、清洗、加工、开发和应用等需求类别的填报。填报内容包括数据来源、需求数据目标、需求业务原因、设计内容、预计完成时间、数据服务方式、需求申请者、联系方式等。

附 录 A
(资料性)
数据组织分类方式

一级主题	二级主题
战略管理	战略规划、投资管理、主业资源、国际化
组织管理	总部组织、授权放权、组织流程、法人治理、组织文化
运营管理	精益管理、现场管理、供应链、营销与服务
财务管理	财务管控、资本结构、资本管理、市值管理
科技创新	科技规划、技术创新、协同创新、创新激励
风险管理	风险管控、内部控制、企业法务、合规管理、责任追究
人力资源	人才规划、市场化选人用人、薪酬激励、人才培养
信息化	信息规划、信息平台、数字化转型、网络安全