

/ 摘要 /

数据标准化是选煤厂智能化建设的基础。针对选煤厂的生产和管理数据意义不明、通用性差和信息孤岛等问题,提出了选煤数据标准化及标准数据平台的建设方案。以建立和完善行业通用又兼容企业个性化的选煤标准数据字典为基础,以分类建构结构化和标准化的选煤生产和管理元数据为核心,基于数据特点分类采用时序数据库和关系数据库进行数据存储,以API接口模式建立基于权限管理基础上的安全、灵活、高效的数据输入、查询和供给服务。通过数据标准化和标准数据管理平台建设,实现选煤厂大数据的规范化、易识别、可携带和全面共享。

选煤厂数据标准化及标准数据平台建设方案

林喆 匡亚莉 王章国

选煤厂智能化是智能矿山重要组成部分,研究与实践是当前热点问题。数据是实现智能化的关键基础,提高各环节数据使用效率是高质量建设智能化选煤厂的重要工作。中国矿业大学匡亚莉课题组,长期从事选煤厂信息化和智能化的研究实践,表明数据便捷、高效利用的前提是实现数据标准化。数据标准化通过定义和明确相关业务含义,改进和再分配数据,提高业务规范性,便于通畅沟通交流,避免不必要分歧。用户和设备使用相同结构数据,提高数据可携带性,减少参数单位间转换,数据在工具软件、计算平台和数据库间高保真自由传输,实现数据互操作性,提高数据处理效率和数据平台管理效率。

近些年,已逐步开展煤炭企业层面数据标准化的研究和实践,准能集团公司在已有生产系统信息化基础上,设计了数字矿山一体化构架,建立了准能数字矿山生产实时数据库系统,并提出了数据标准化构架和数据标准化体系,实现了数据标准化、结构化、实时化的高效管理;国家能源集团神东煤

炭集团有限公司在数字矿山平台基础上,规划了数据标准化项目,解决数据分散、共享程度低、标准不统一等问题,并通过数据共享交换平台和大数据平台,逐步实现数据标准化。此类研究与实践是从煤炭企业顶层设计数据治理框架,为选煤数据标准化提供了参考依据。但选煤数据标准化易识别、可利用的成熟推广应用,还存在诸多问题。本文基于选煤标准数据平台建设,构建结构化选煤标准数据,探讨选煤数据标准化及管理平台建设的思路与方案。

选煤数据标准化定义及管理平台框架

数据标准化是指企业或组织对数据的定义、组织、监督和保护进行标准化的过程。标准化在经济、技术、科学和管理等社会实践中,对重复性事物和概念,通过制订、发布和实施标准达到统一,以获得最佳秩序和社会效益。

为实现选煤数据标准化,标准化平台应具备4个功能模块:标准字典、基础数据、元数据管理和安全管理,如图1所示。

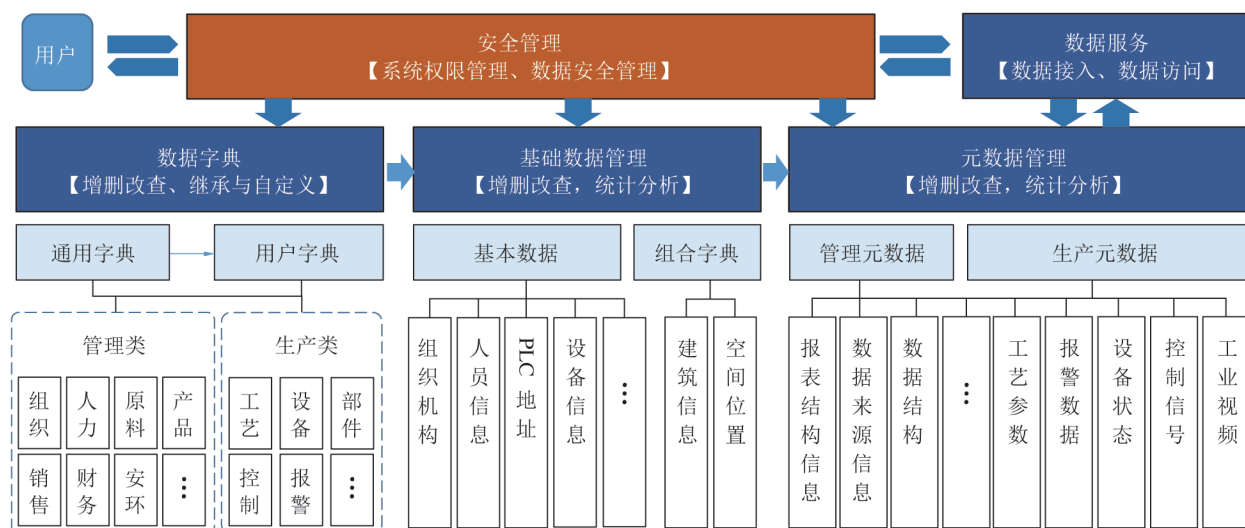


图1 选煤数据标准化平台架构

(1) 标准字典是数据标准化的最基本要素，是实现数据统一表达的基础。

(2) 基础数据是基于标准字典组合形成的具有特殊作用的数据，功能类似于字典，但又不具备字典的通用性，因此单独管理。

(3) 元数据是数据标准化描述的数据集，采用标准字典和基础数据的词条，结构化描述数据，使数据能被人识别，也能被计算机识别。

(4) 安全管理是管理平台数据使用权限和数据安全模块。

此外，更高级别的标准数据平台还应对外提供数据服务，包括数据接入、数据存储和标准化数据查询。

选煤标准数据字典定义与管理

数据字典是描述数据模型中数据对象或项目的集合，用于程序员和其他需要参考的人识别数据。选煤标准数据字典的定义是数据标准化的基础。在选煤生产实际中，不同单位、不同部门和不同专业对相同数据的描述存在不一致。在复杂的选煤厂智能化建设中，大量数据命名和描述存在同词异义或异词同义现象。例如“欠速”“失速”“打滑”都可用于描述带式输送机的驱动滚筒与输送带间的相对

滑动，导致输送带输送异常；“油泵”为液压油泵简称，也是润滑油泵简称。此类问题都为准确描述数据特点，特别是计算机识别数据点制造障碍。因此，将数据描述相关词条规范化为数据字典，并规定采用字典中的词条进行数据描述，可大幅提高数据描述的可读性和准确性。

选煤智能化的标准字典应满足通用性、准确性、全面性的要求。

(1) 字典通用性是数据流通的必然要求，数据字典要实现尽可能高层级通用，最好是通行国家标准或行业标准；在缺少相关标准的情况下，至少要实现集团公司级通用，数据定义才具有广泛意义上的可识别和可流通属性。

(2) 准确性是指字典的科学、准确，要符合科学术语和行业规范性。

(3) 全面性是指字典应能尽可能覆盖描述数据所需词条，各公司或选煤厂特殊的个性化词条，也应纳入字典管理，但可作为厂级私有字典，而不作为通用字典。

基于上述原则要求，字典纵向管理应分为“通用字典”和“用户自定义字典”2个以上层级。通用字典收集所有共性、通用词条，可在行业层面上定义；用户自定义字典可细分层级为集团公司级、公司级和厂级等。下级字典可继承上级字典，用户

可直接使用或对继承字典修改后使用；对于缺少字典，用户可自定义新增字典。下级字典修改或删除不影响上级字典。

横向字典分类管理，总体可分为管理类和生产类。管理类字典为公司管理过程使用数据项所涉及词条，包括公司组织架构信息、人力资源信息、计划统计信息（包括原料、产品和销售等）、财务、安全环保和其他行政管理业务使用的数据；生产类主要是与选煤生产实时相关的数据信息，包括工艺系统、设备仪表、辅助设备、零部件、控制信号、报警信息和工业视频等。

基础数据定义与管理

基础数据由2类数据构成：

(1) 选煤厂最基本的数据信息，如人员信息和设备信息等。

(2) 组合字典，由基础字典组合成选煤厂特有数据描述词条。具有数据字典的基本属性，是为描述数据服务的词条；该词条一般不具备通用性，是某个单位特有且由多个基本数据字典词条组成，具有较长字节数，包括厂房建筑及相关建筑空间定位点描述，描述数据点或数据监测对象位置等基础信息。

基础数据形成后等同于数据字典，元数据中相关描述必须引用基础数据表的值，形成标准化数据信息。此类数据介于字典和元数据间，由用户根据需要定义。基础数据可在公司级、厂级或部门级定义，与数据字典相同，下级可继承上级，下级不影响上级。

选煤厂智能化元数据定义与管理

元数据是描述数据的数据，又称为中介数据，是描述数据属性的信息。不同领域，元数据格式和定义不同，最基本的功能是实现数据资源的快速有效查找。对于选煤智能化系统，元数据定义是选煤数据标准化工作的核心。通过定义元数据，可有效

实现海量数据的格式化、规范化，并可快速实现数据来源、历史数据和实时数据的定位和读取等功能，使数据能清晰地被不同专业的用户识别，提高数据意义的清晰度和数据使用便捷性，奠定数据仓库共建共享基础。选煤元数据管理及其数据服务原理如图2所示。

根据智能化选煤系统的数据特点，元数据可分为2类：

(1) 控制系统通过传感器在线采集的各类数据，用于系统智能化分析决策或统计分析等，称为生产类元数据。

(2) 人工通过各管理系统录入或由生产实时数据和采制化数据，经过一定统计方法计算得到的数据，是管理和统计的非实时类数据，称为管理类元数据。

元数据的业务分类，不同于数据治理概念中元数据在技术、管理和业务上的分类标准。管理类元数据除明确自身定义外，重点描述数据来源和去向，主要字段包括数据分类、数据名称、数据内容、计量单位、采集频率、数据来源和数据用途等信息。管理类数据具有强相关性，应采用关系型数据库存储。

生产类元数据是工艺系统中主要设备运行状态和工艺参数的监控、控制和报警等信息，细分为6类数据。

(1) 工艺参数是直接反映生产过程运行状态的工艺类模拟量数据，包括物料流量、灰分、水分、浓度、阀门开度、液位、压力等。

(2) 控制指令是控制系统发出的控制指令，主要是开关量，包括设备启停、阀门开关、故障处理等控制命令。对工艺参数的设定值归入工艺参数，出于安全目的的设备动作和参数范围的限制设定归为控制指令。

(3) 运行状态主要指生产系统中设备的运行状态，包括设备开停、起车准备、带电、检修、故障、停运时间等状态信息。

(4) 安全监控是指设备和生产场所的安全状态信息，包括感温、感烟、振动、电流、设备温度等。

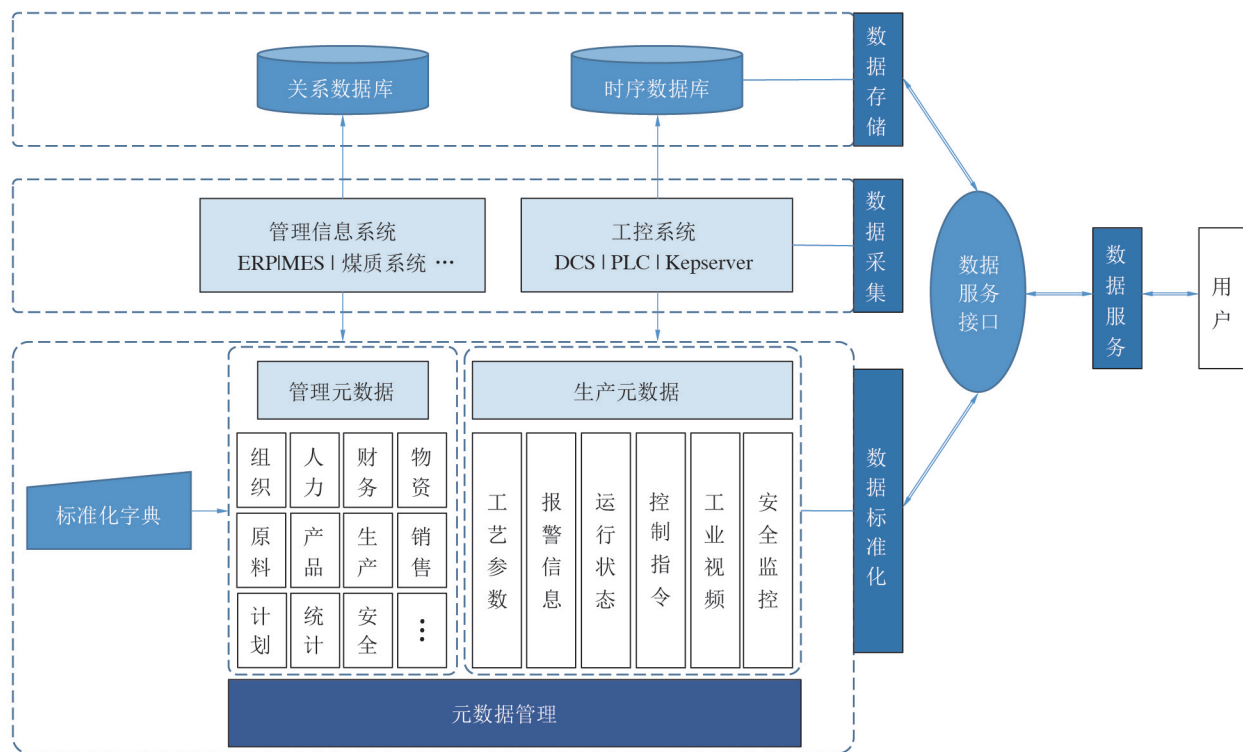


图2 选煤元数据管理及其数据服务原理

(5) 报警信息是指生产过程中发出的各类异常报警，包括设备故障、设备状态异常、工艺参数超限等信息。

(6) 视频监控是指各场所的实时监控视频。

生产类数据的字段主要包括工艺系统、子系统、设备编号、设备名称、数据监测点的定位信息（即被监测对象位置，根据实际情况细化到辅助设备、零部件、空间位置描述信息）、数据分类、数据阈值、计量单位、计数类型（瞬时值或是累计值等）、数据来源PLC的IP地址、数据点的地址标签、存储地址和存储标签等。根据工艺系统、设备编号和数据分类等信息，可方便地查找相关血缘关系的元数据家族；数据来源和数据存储信息为数据使用者提供便捷的数据支持。生产类元数据描述的数据项是与时间相关的一维数据，采用时序数据库存储。

元数据管理系统应具备数据的批次和单项添加，及数据项修改和删除等基本功能，提供基于数据结构的元数据分类查询与统计等功能，以支撑数据实体关联分析、实体差异分析、数据血缘分析和指标一致性分析等需求。

选煤标准数据平台的安全管理

除数据备份等通用性数据安全措施外，选煤数据标准化平台提供数据管理和数据服务，设计用户访问权限和数据接口访问权限，包括：用户信息管理、用户访问权限管理和接口访问权限管理。

选煤标准数据管理系统支持前后端用户管理与权限管理。前端管理包括访问用户信息管理、角色管理、角色权限配置管理，及角色与用户关联；后端管理包括后端访问token配置管理，及token对应数据权限管理。数据服务安全管理框架如图3所示。

后台数据处理服务通过模型-视图-控制器（Model-View-Controller, MVC）框架，实现数据库数据查询与聚合，通过HTTP服务（Web API数据接口）对外提供聚合结果，在前后端访问的过程中，通过访问验证和过滤器，实现访问权限控制。

权限控制包括2个层面。

(1) 通过前端和远程为用户提供唯一访问标识。

(2) 通过前端和远程用户访问标识，分配菜单及菜单下数据字段和按钮的访问权限。

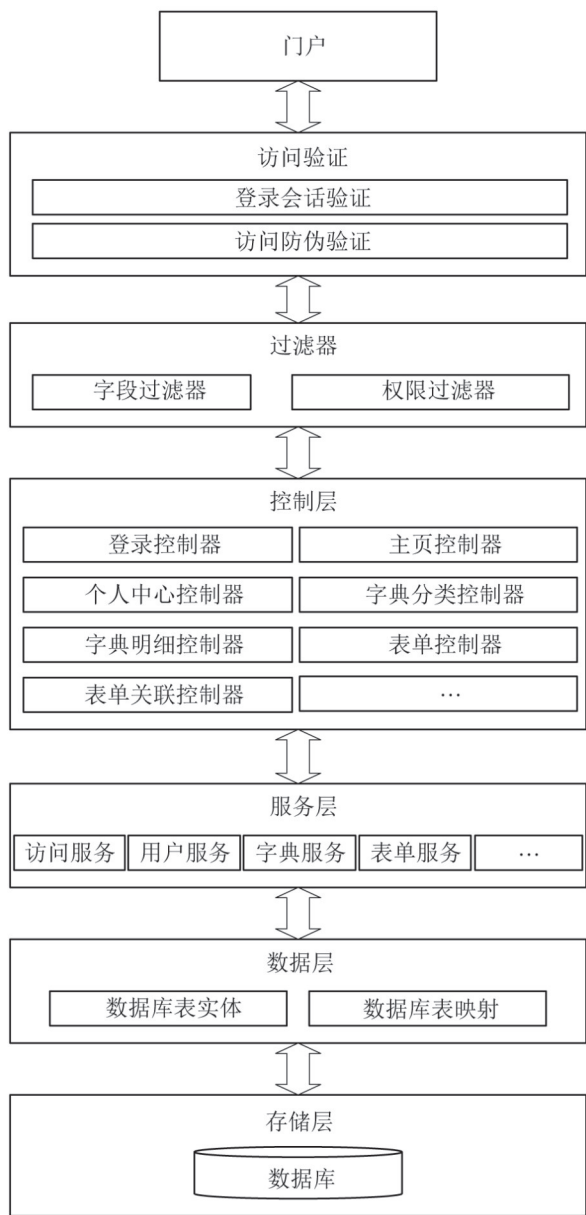


图3 数据服务安全管理框架

结 论

数据是选煤厂智能化的基础，数据标准化是提高数据利用效率的必要条件，选煤厂智能化实践已经着手开展数据标准化和数据治理工作。从数据管理和使用需求出发，从数据字典、基础数据、元数据、数据安全和数据共享服务等方面进行数据标准化定义和规范，并构建软件平台进行管理。针对选煤数据的特点，标准化方案应遵循以下原则：

(1) 选煤数据标准化应建立在行业通用标准字典的基础上，基于数据治理的理念，以建构结构化元数据为核心，形成易识别、可携带、能够在工具软件、计算平台和数据库间高保真自由传输的标准数据及其服务体系。

(2) 选煤元数据根据来源和用途分为管理类和生产类，管理类数据用关系数据库存储，生产类数据用时序数据库存储。

(3) 以MVC为框架、通过API接口对外数据服务是当前选煤标准数据管理平台的较优模式。

■ 策划:赵瑞 编辑:宫在芹

作者简介:

第一作者:林喆,副教授,博士,主要从事选煤厂智能化和细颗粒固液分离研究工作。

E-mail: z.lin@cumt.edu.cn

作者单位:中国矿业大学化工学院

热点问答

煤矿 5G 专网与 F5G、WiFi6 在智能矿山通信网络需求与场景包括哪些?

国家多部委联合印发《关于加快煤矿智能化发展的指导意见》。煤矿通信网络环境特殊,需按实际需求建设数据采集与传输系统。从应用场景看,对通信网络需求分三类:信息采集类(上行数据为主)、控制类(下行数据为主)、双向交互类。

智能采煤、掘进、运输、视频监控、智能巡检等场景都有通信需求。通信需求主要包括大带宽(如视频监控场景,上行传输需大带宽)、低时延/高可靠(远程控制等场景)、移动性(自动巡检等)、数据安全(封闭生产区域特点)、设备安全(井下设备需相关认证)。传统移动通信网络有短板,基于 5G 无线网络和 F5G 光纤有线网络融合的矿井通信网络是智慧矿山通信网络的最佳选择。

——来源:《中国煤矿智能化发展报告(2022 年)》