

基于元数据智能治理的电力数据要素价值挖掘

黄祖源
云南电网有限责任公司信息中心



目录 CONTENT

01 成果背景

03 成果价值

02 成果内容

04 典型应用案例



中国南方电网
CHINA SOUTHERN POWER GRID



01

成果背景

自十八大以来，从国家发展到南网数字化转型都对数据治理提出了明确要求，数据治理已经成为释放数据要素价值、占据国际数字竞争制高点的迫切需要。





➤ 国家发展方向

党的十八大以来，以习近平同志为核心的党中央高度重视数字化发展，明确提出建设数字中国。2022年6月22日，习近平总书记在主持召开中央全面深化改革委员会第二十六次会议时强调，促进数据高效流通使用、赋能实体经济，统筹推进数据产权、流通交易、收益分配、安全治理，加快构建数据基础制度体系。会议还审议通过了《关于构建数据基础制度更好发挥数据要素作用的意见》。数据要素国家治理体系是指统筹数据要素流通与应用的一系列相互关联、相互协调的制度安排，体现在国家、行业、区域、企业、个人等不同层面，涉及科技、产业、社会等不同应用领域以及数据采集、存储、加工、分析等不同管理阶段。构建数据要素国家治理体系不仅是我国推进国家治理现代化的迫切需要，而且是我国释放数据要素价值、占据国际数字竞争制高点的迫切需要。

➤ 南方电网发展方向

南方电网公司明确提出了建设“数字电网”和数字化转型的目标，进一步强调了强化数据资产管理，挖掘数据资产价值，落实国家大数据发展战略的战略举措，这些都要求**夯实数据治理基础**，通过**加强元数据管理、数据标准管理、数据质量管理等数据治理工作**，为开展数据资产管理创造良好的战略环境和制度基础，也为促进数据价值释放打好数据基础。



数据治理面临的问题-痛点分析

PROBLEM STATUS

在大数据时代，**传统的数据治理手段已无法应付治理过程产生的问题**，例如多源异构数据的获取和定义、数据治理效率、标准和规则的时效性、治理知识的利用率、多维治理知识的融合等问题。

传统数据治理过程

数据
定义
管理

元数据管理：电力企业主要针对结构化数据的技术元数据实现了自动采集，而通过管理手段完善管理元数据和业务元数据。此外在非结构化数据的元数据管控上处于起步阶段。

数据
质量
治理

数据标准管理：内外部不同来源的业务数据，其数据标准存在不一致的情况；其次是随着业务变化而发生的数据标变化，无法及时更新对应的数据治理规则。

数据质量管理：数据质量规则清单主要依赖于人工经验梳理和维护，目前情况下只能覆盖核心业务和部分常规业务，对于新兴业务和外部业务覆盖率仍有待加强。

知识
沉淀

质量问题分析：在对数据问题进行分析时，无法完全摆脱对业务专家的依赖，在支持基层业务人员开展治理时，服务质量有待提高。

数据治理知识管理：元数据、数据标准、数据质量规则、问题数据分析结果、整改指引、整改经验等方面的知识独立积累，数据治理知识未能全面有效地形成体系。

传统数据治理问题

多源异构数据的数据定义与标准不统一，数据覆盖不全

数据治理过程缺乏自动化和智能化手段，效率较低

数据标准和数据质量规则的时效性无法保障

治理专家知识未能有效利用，问题发现效率不高

数据治理知识融合程度低，无法多维度的查询和展示关联信息



中国南方电网
CHINA SOUTHERN POWER GRID



数据治理-智能化是发展趋势

DEVELOPMENT TREND

在面临上述问题时，如果还是依靠传统“人海战术”的数据治理模式，将无法适应业务数字化转型的要求。因此，云南电网在实践中尝试通过自动化和智能化的数据治理手段实现基于元数据的智能治理，深挖电力数据要素价值，“以智促质、以质增值”。

“在Gartner 2020年发布的“数据与分析领域的十大技术趋势”中给出了这样的预测”

增强型数据管理利用ML（机器学习）和AI（人工智能）技术优化并改进运营。它还促进了元数据角色的转变，从协助数据审计、沿袭和汇报转为支持动态系统。

增强型数据管理产品能够审查大量的运营数据样本，包括实际查询、性能数据和方案。利用现有的使用情况和 workload 数据，增强型引擎能够对运营进行调整，并优化配置、安全性和性能。

智能化技术的发展驱使数据治理变革！

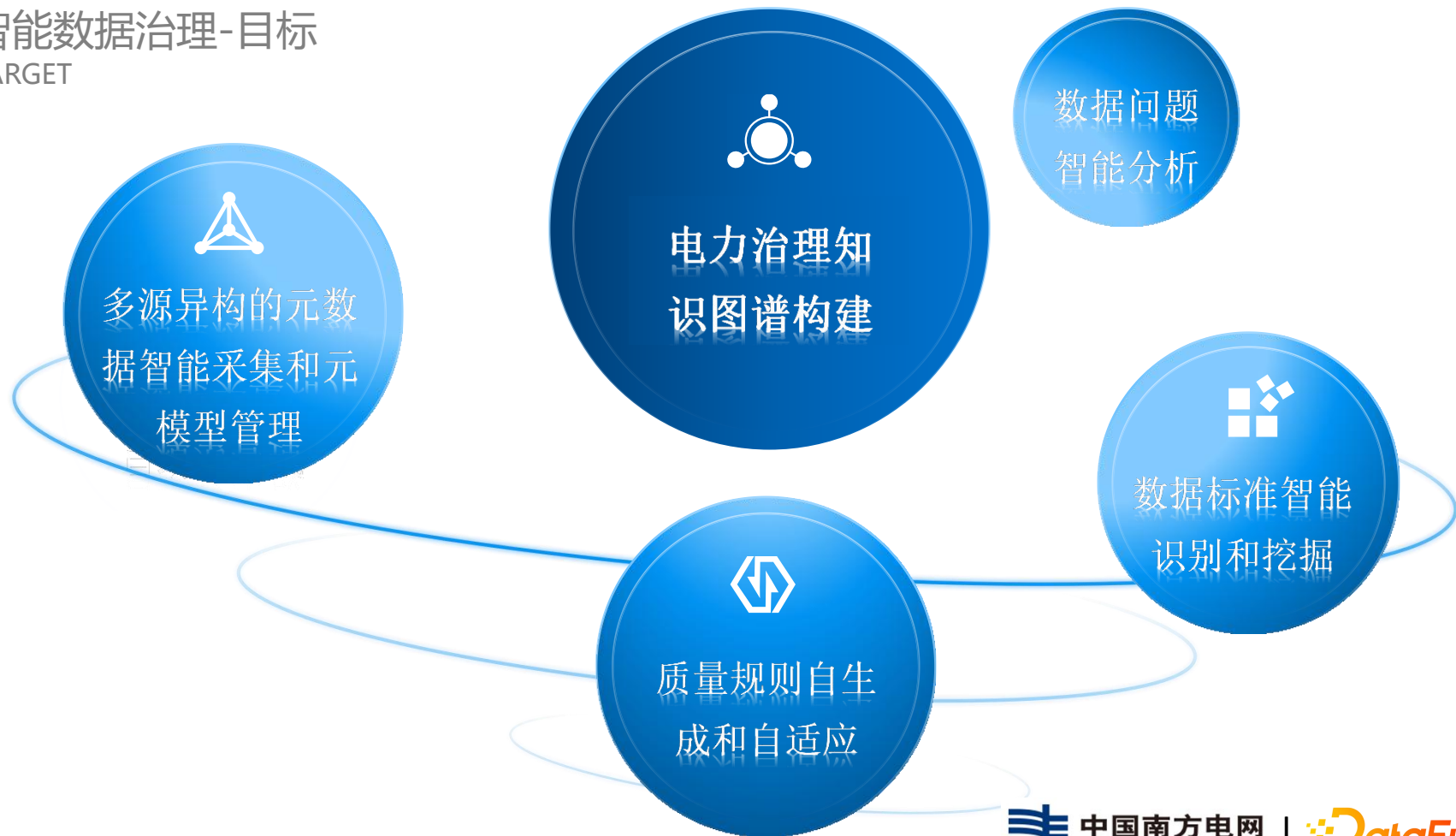


中国南方电网
CHINA SOUTHERN POWER GRID



智能数据治理-目标

TARGET



中国南方电网
CHINA SOUTHERN POWER GRID

| DataFun.

02 成果内容

云南电网基于元数据的智能治理经历理论研究、平台建设、场景应用三个阶段，从“治理知识图谱化”、“元数据自动化管理”等五个关键方向进行突破。



实现路径

TECHNOLOGY ROADMAP

为实现“以智促质、以质增值”的目标，云南电网公司通过对治理理论体系的研究提出了“治理知识图谱化”等五个改进点，打造数据治理智能化技术的平台建设，以应用场景实现价值挖掘，初步探索了一条智能化数据治理的前进方向。

01

统一大数据治理体系研究

- 治理知识的图谱化构建研究
- 元数据自动化管理技术研究
- 数据标准自动生成技术研究
- 数据规则自动生成技术研究
- 数据质量问题智能分析技术研究

02 平台建设

基于数据治理智能化技术的平台建设

- 知识图谱构建
- 数据规则生成
- 元数据管理
- 数据问题智能分析
- 数据标准生成

03 场景应用

数据治理智能化技术应用场景

- 应用场景方案



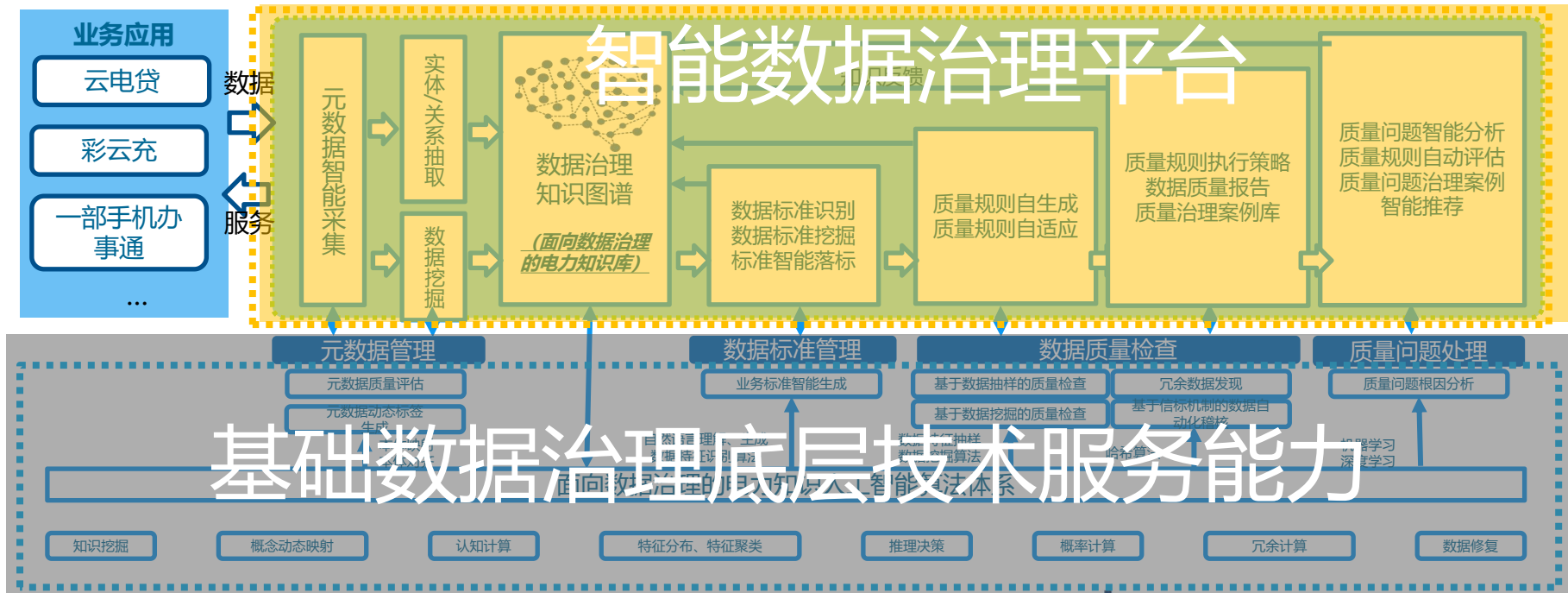
中国南方电网
CHINA SOUTHERN POWER GRID



基于元数据的智能治理体系

INTELLIGENT GOVERNANCE SYSTEM

基于元数据的智能治理体系主要分为两层，一是基础数据治理底层技术服务能力，基于机器学习、自然语言处理、智能算法等技术能力，面向电力数据治理问题所构建的基础技术服务能力。二是智能数据治理框架，在应用数据治理技术服务能力的基础上，根据治理应用需求搭建的智能数据治理平台。



中国南方电网
CHINA SOUTHERN POWER GRID



技术实现思路-多源异构数据的元数据智能识别

TECHNICAL REALIZATION IDEAS

多源异构数据不可知



数据资产不清晰

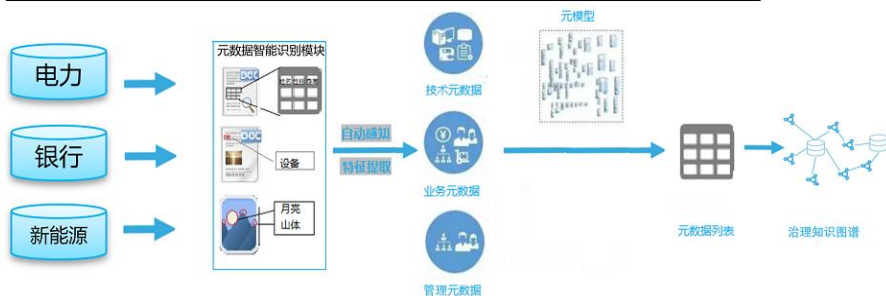
构建统一的元数据模型，支持多源异构数据的元数据智能识别

通过构建多源异构的元数据模型，将企业内外部数据的元数据基于业务对象进行重新组织；利用元数据智能识别，在技术元数据基础上提取完善业务元数据和管理元数据，构建结构统一、维度多样的数据资产目录。

传统思路



智能化思路

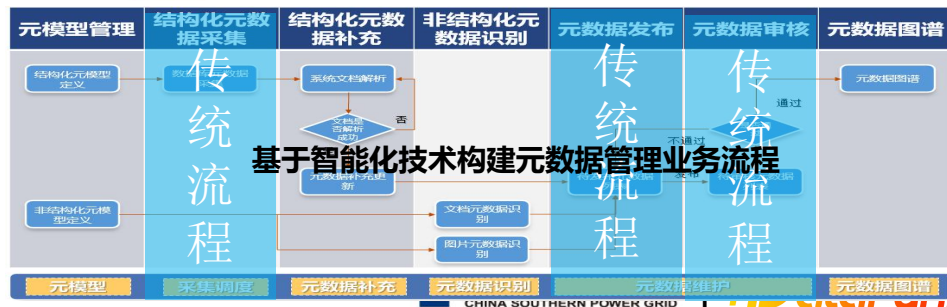


应用价值

提供元数据自动化和智能化技术，为跨域数据的数据构建统一的元数据模型，统一数据的含义描述，为数据治理打好基础。

技术亮点

- ✓ 元数据智能识别
- ✓ 统一元数据模型
- ✓ 元数据自动补充

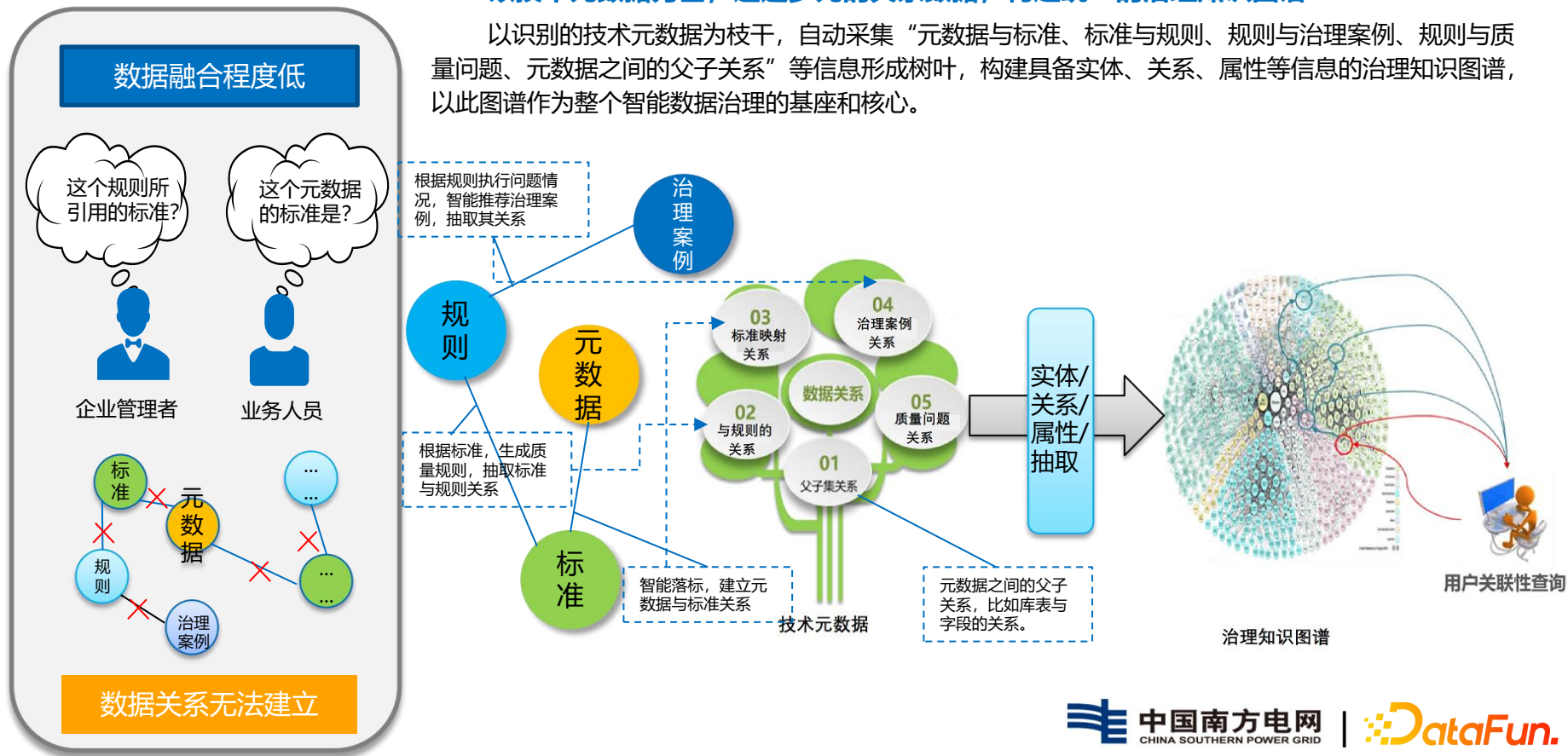


技术实现思路-构建数据治理知识图谱，实现多元数据知识的融合

IMPLEMENTATION SCHEME

以技术元数据为基，通过多元的关系数据，构建统一的治理知识图谱

以识别的技术元数据为枝干，自动采集“元数据与标准、标准与规则、规则与治理案例、规则与质量问题、元数据之间的父子关系”等信息形成树叶，构建具备实体、关系、属性等信息的治理知识图谱，以此图谱作为整个智能数据治理的基座和核心。



中国南方电网
CHINA SOUTHERN POWER GRID

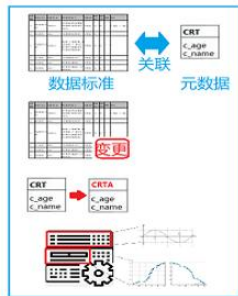
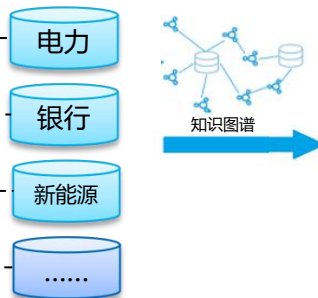
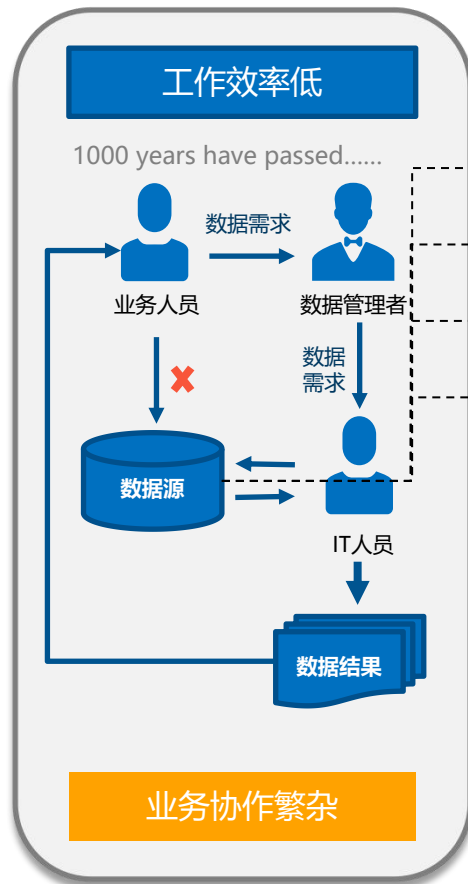


技术实现思路-通过智能化技术，实现提标准和规则的自生成

IMPLEMENTATION SCHEME

基于元数据实现数据标准和质量规则自动生成，并支持规则的动态更新

基于治理知识图谱，定位元数据对应的数据信息，通过对数据特征进行智能挖掘和识别生成数据标准；根据元数据与数据标准落标关系，遵循规则生成规范自动生成数据质量规则，将数据质量指标作为元数据的管理元数据反哺、完善治理知识图谱。

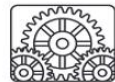


基于元数据落标的数据标准

基于数据标准变更

基于元数据变更

基于数据变化趋势



SQL脚本生成引擎



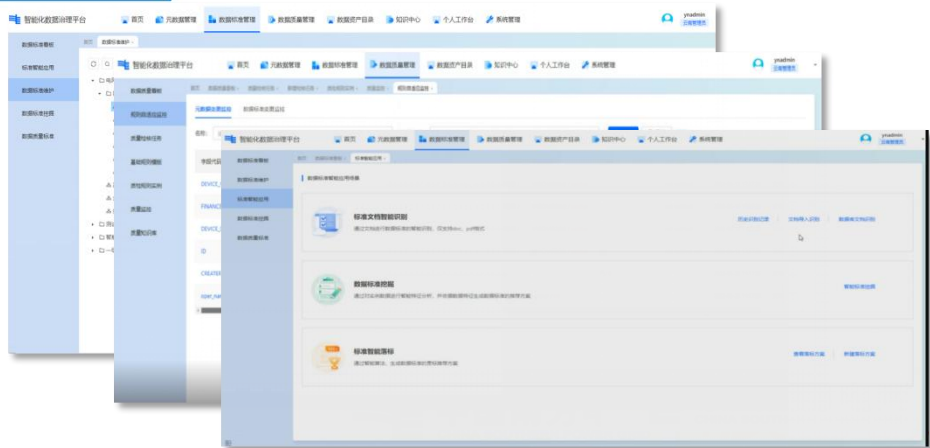
数据质量规则

应用价值

减少人为的参与，提高数据治理工作效率。

技术亮点

- ✓ 数据标准智能挖掘
- ✓ 数据标准智能落标
- ✓ 质量规则自动生成
- ✓ 质量规则自动适应



技术实现思路-利用治理案例知识沉淀，实现质量问题的智能分析

IMPLEMENTATION SCHEME

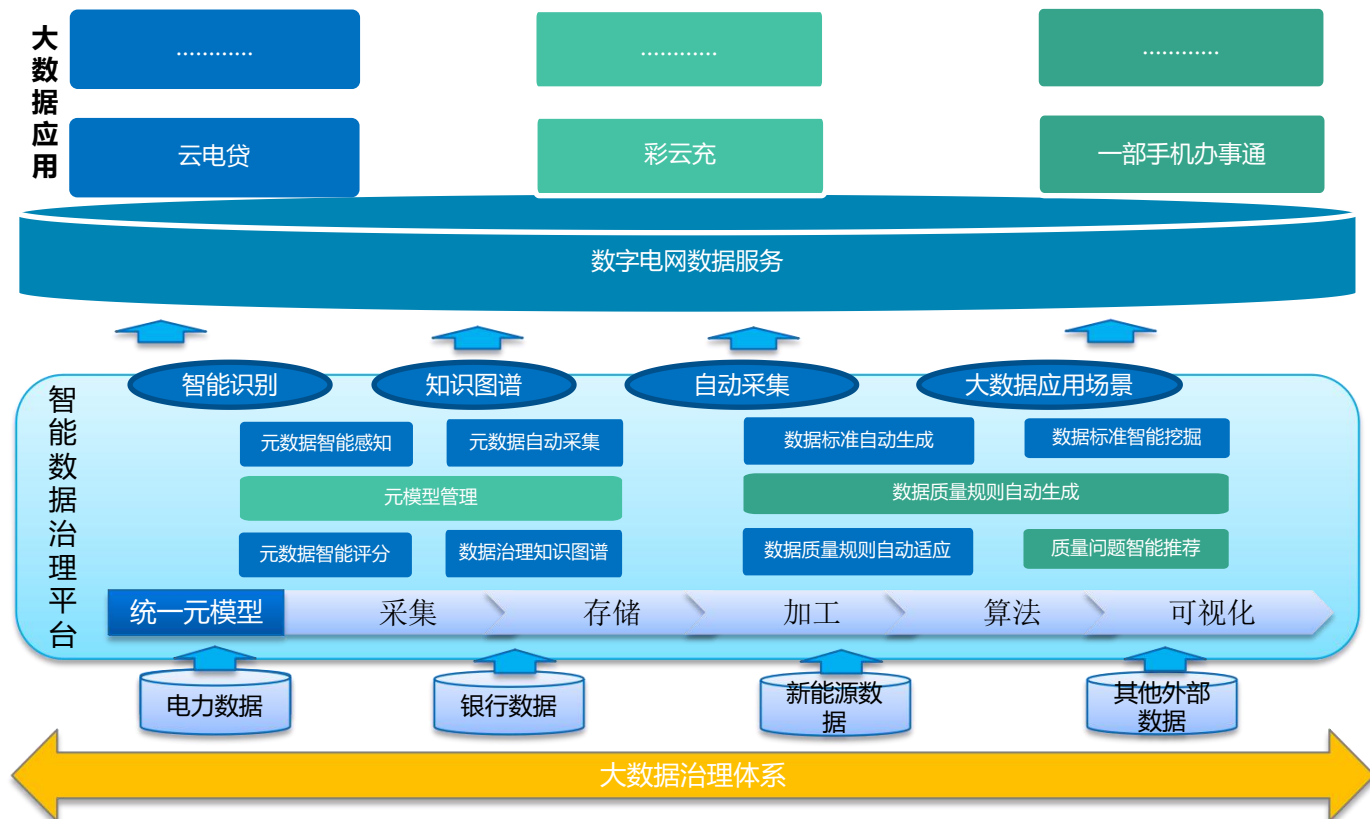
根据学习治理案例问题特征，利用算法实现质量问题的快速定位

结合云南电网数据治理所积累的知识库，利用机器学习技术，训练质量问题分析模型，根据数据质量规则的执行结果中质量问题的特征情况，通过算法技术计算特征的相似程度，自动的推荐问题分析原因，以及治理案例，使用户通过治理知识图谱能够一站式查询所有相关治理知识。



智能数据治理平台

OVERALL STRUTURE DESIGN



智能治理平台

通过多源异构数据，利用元数据的智能识别技术，实现元数据自动管理，并基于元数据构成，实现数据标准和质量规则的自动生成，以及智能化的问题分析处理和治理知识图谱的自动构建。

平台通过智能和自动化手段实现多源异构数据的治理，有效的支撑数据交互服务和大数据应用。



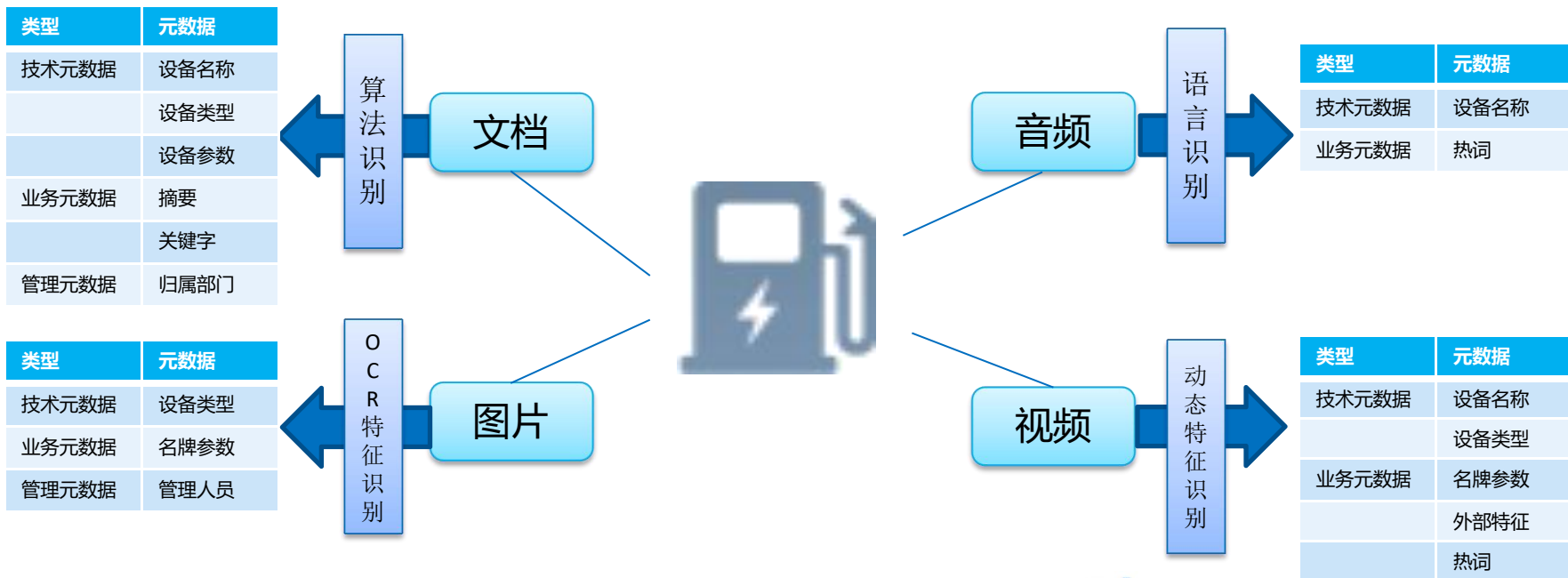
中国南方电网
CHINA SOUTHERN POWER GRID



关键技术-非结构化数据的元数据智能感知技术

TECHNOLOGICAL INNOVATION

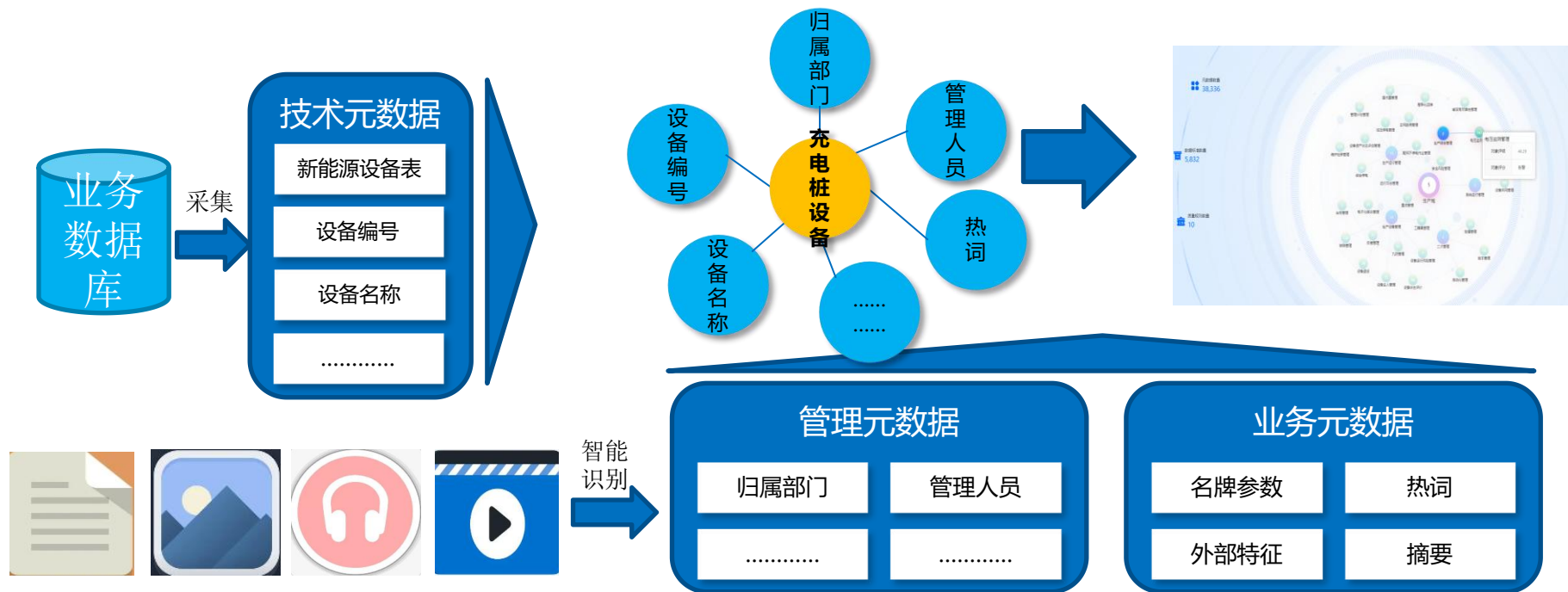
通过对业务对象非结构化数据的识别，利用自然语言处理、图像特征识别、语言识别、动态特征识别等智能技术，可快速对图像、文档、音频、视频等非结构化数据的元数据信息的分析和识别。



关键技术-构建电力元数据知识图谱

TECHNOLOGICAL INNOVATION

首先对业务系统数据库采集技术元数据，构建知识图谱原型；其次根据从技术文档、规范制度、数据特征中智能识别的业务元数据和管理元数据，提取实体、关系、属性等信息，不断循环完善知识图谱，形成体系化、一站式的图谱。



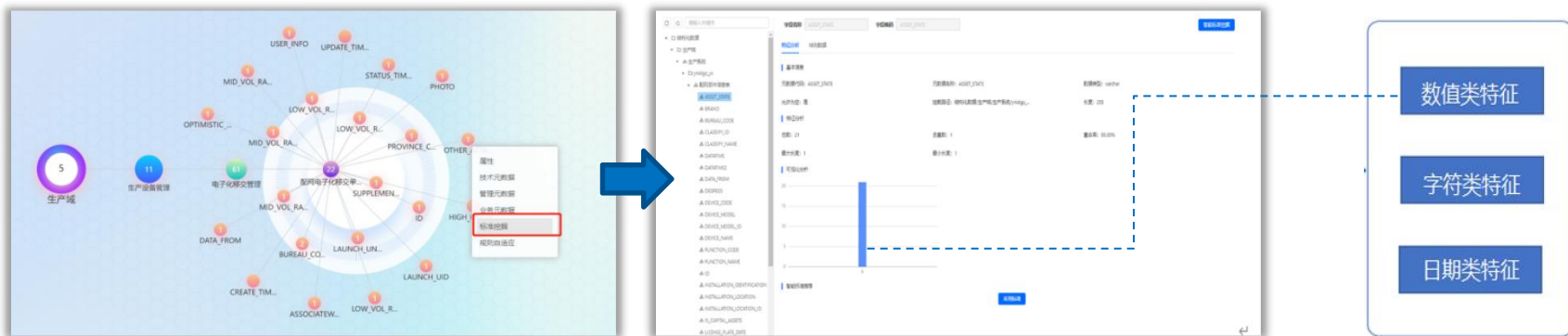
中国南方电网
CHINA SOUTHERN POWER GRID



关键技术-数据标准智能挖掘技术

TECHNOLOGICAL INNOVATION

通过在数据治理知识图谱中选择业务对象的相关字段，可以在数据挖掘页面可通过数据特征算法计算出该字段的数据类型特征分布情况，以此作为在业务发生变动时，对数据标准的变化感知和完善补充，重点强化对外部数据的元数据管控能力和数据治理能力。



步骤一：通过知识图谱，选择需要进行标准挖掘的字段元数据

步骤二：自动定位到该字段，通过智能算法进行数据特征分析，可分析该字段长度、枚举值、值域等数据特征，最后人工确定后形成标准。



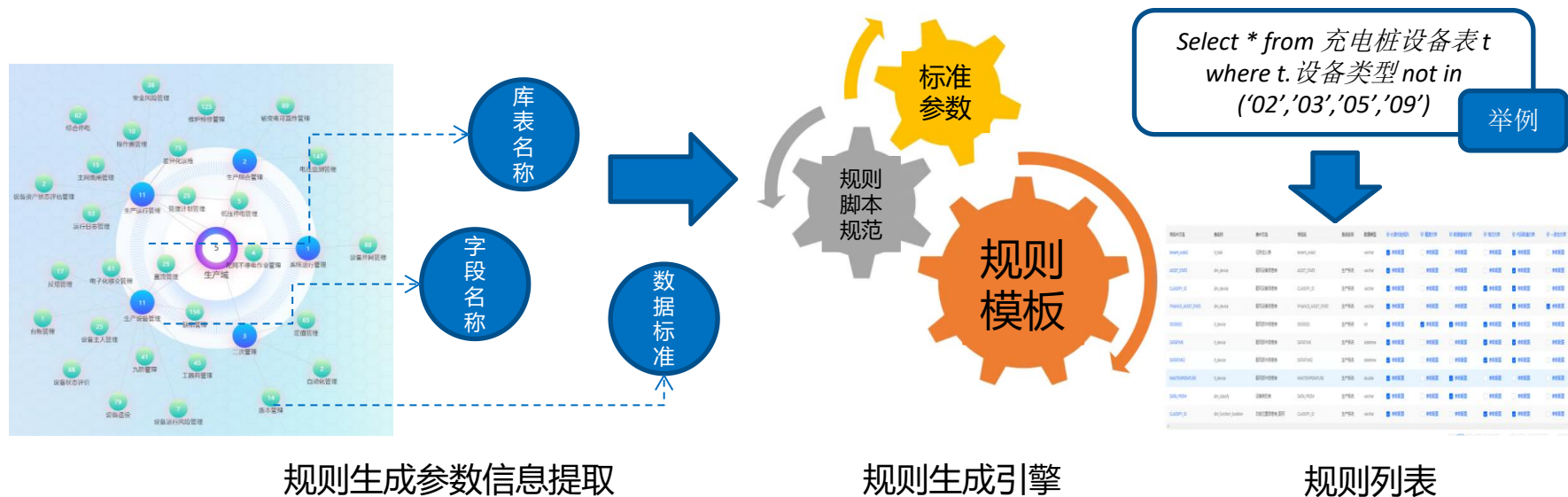
中国南方电网
CHINA SOUTHERN POWER GRID



关键技术-数据质量规则自生成自适应技术

TECHNOLOGICAL INNOVATION

数据质量规则通过元数据的库表名称、字段名称、数据标准的关系信息，利用规则生成引擎，对参数和标准信息按照规则脚本规范进行组装，形成可执行的SQL脚本。并支持通过感知元数据、数据标准和数据特征变化情况，进行自适应更新。通过这项技术实现普适性数据质量规则的生成，实现数据治理的全域覆盖。



关键技术-质量问题自动化归因分析技术

TECHNOLOGICAL INNOVATION

数据质量问题自动化归因分析基于云南电网历年积累的数据质量问题分析报告、整改指引等治理案例信息，采用基于特征相似度计算，分析匹配，形成最佳的综合治理建议推荐。以治理知识图谱为基础，实现通过元数据对数据治理相关所有信息的统一查询。



中国南方电网
CHINA SOUTHERN POWER GRID



03 成果价值

基于元数据的智能治理经过一年的实践，已覆盖银行、新能源、客户服务等应用场景，有效促进了电力数据要素价值挖掘与释放。



实用价值

PRACTICAL VALUE

成果应用于云南电网公司智能数据治理平台，为公司提供了全域数据质量辅助：通过**1年**的应用，在扩大元数据覆盖、解决多源异构问题、实现智能化管控等方面均取得明显成效，治理后的数据交换服务**已覆盖银行、新能源、客户服务等应用场景**，有效支撑大数据应用落地，促进了电力数据要素价值挖掘与释放。



5000+

已管理的元数据



3000+

自动生成规则



1000+

处理数据问题



100+

支持数据服务



10+

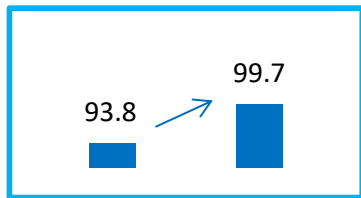
支撑大数据应用



20+

支撑应用场景

数据质量指数

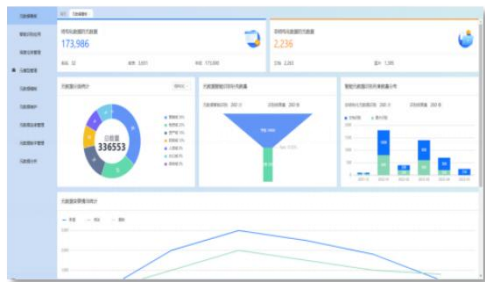


60%

减少数据
问题处理
时间

50%

释放全职
人力工时



中国南方电网
CHINA SOUTHERN POWER GRID



04 典型应用案例

通过基于元数据的智能治理，云南电网将治理数据与对外应用深度结合，实现了“云电贷”、“彩云充”等对外数据价值释放的典型案列。



SUPPORTING APPLICATION CASES

通过对用户电力数据的元数据定义，以及元模型构建，给“云电贷”产品在数据库设计和数据模型构建提供了极大的便利性，**将应用落地推广的时间提前了半年**，并率先在昆明、曲靖、玉溪三地同时落地推广，并成功由用电客户自主申请达成首单。

通过智能化数据治理，云电贷产品所管理的数据得到大量治理，数据服务访问的成功率从81.3%提升至95.7%，有效的支撑了“云电贷”数据增值服务的访问，目前产品已有2118次访问量。

[illegible]

电力能源视角刻画客户信用画像

典型应用案例-彩云充

SUPPORTING APPLICATION CASES



接入电动汽车**25000**余辆，充电站（点）**800**余个，充电桩（枪）**9000**多个，累计完成充电电量**3000**万千瓦时以上。

1、数据服务支撑

通过对充电数据的分析和治理，将数据问题从8%下降到2%，支撑了新能源汽车及充电设施的实时数据统计服务，以及新能源汽车及充电设施一体化的大数据应用分析服务。

2、业务支撑

通过对充电数据和电网布局数据进行协同治理，为充电桩精准定位和布局夯实基础，解决新能源汽车用户找桩难、充电难的业务困境。

3、第三方支撑

通过电网数据接口与其他平台数据接口标准化的元数据统一定义，优化数据交互的兼容性。



中国南方电网
CHINA SOUTHERN POWER GRID



非常感谢您的观看



中国南方电网
CHINA SOUTHERN POWER GRID

