

数据母体产品白皮书



国脉集团

二〇二〇年十二月

目录

前言	错误! 未定义书签。
一、背景与需求	1
二、产品概述	2
(一) 产品定义	2
(二) 主要模块	3
(三) 产品价值	错误! 未定义书签。
三、产品数据体系与技术体系说明	5
四、产品应用场景	7
五、成功案例	10
六、公司介绍	11
七、联系我们	14

前言

当前，数据及其技术的融合应用在政府经济调节、市场监管、社会管理、公共服务、生态环境保护等各项工作中强劲助攻、潜力无限。但由于数据是新型生产要素，具有体量大、反应快、虚拟性、多维复杂、不断裂变等新特点，传统的资源管理模式无法完全适配。党的十九届四中全会《决定》在“优化政府职责体系”中，明确提出要“建立健全运用互联网、大数据、人工智能等技术手段进行行政管理的制度规则”。而互联网、大数据、人工智能等技术手段应用的基础是数据。由此，我们看出围绕数据生产要素，构建科学的数据治理规则体系，是政府治理体系和治理能力现代化的重要组成部分，也是优化政府职责体系、深化供给侧结构性改革过程中必须要解答好的问题。

一、背景与需求

（一）数据治理作为政府治理的重要内容具有历史必然性

数据成为新型生产要素，是具有历史必然性的，是社会生产力不断向前发展的结果。如同农业社会的土地、工业社会的资本和能源、信息社会初期的计算机、集成电路、互联网等一样，随着社会生产力不断积蓄、发展，数据爆炸式增长推动数据处理技术日益成熟，数据价值如冰山一角逐渐显露。数据由过去记录信息的载体逐渐演变为驱动社会变革的力量，成为智能时代的主角。

站在历史唯物主义视角，数据及其技术作为新时期的生产要素和生产力，其发展是必然的、不以人们的意志为转移的。在这一阶段，人们必然要不断调整、优化自身所处的生产关系，以期与物质生产力实现对立统一、相互促进、螺旋上升。于是，建立与数据生产要素、数据技术发展相适应的数据治理体系，是社会进入智能时代的必然选择，属于上层建筑范畴。数据治理能力的好坏，直接反映出政府乃至社会拥抱新时代、适应新特征、应对国内外复杂环境的能力和韧劲。

（二）数据治理正成为国家治理能力现代化重要内涵

1. 数据治理和国家治理成为国家时代性主题。工业社会向信息社会转型对国家治理提出严峻的挑战，总书记指出没有网络安全就没有国家安全，没有信息化就没有现代化。数据治理能力正在成为组织竞争新优势。

2. 数据治理是国家治理在信息社会的重要组成部分，没有数据治理现代化就没有国家治理能力现代化。当前，国家治理与数据治理相互促进，数据治理的理念、方法、工具和流程将为国家治理提供极为重要参考的依据。

3. 大数据成为信息化主战场。我们已进入数据时代，这对数据治理提出更高要求。大数据成为信息化主战场，数据治理是当前信息化建设应用中的重点、要点和难点，是数据时代的关键任务和制度创新，提升数据治理能力将为国家治理能力现代化做出重要贡献。

（三）如何提升数据质量是数据治理过程中的重点难点

数据质量是奠定数据价值的基础，也是数据治理的一大难点和重点。当前很多数据问题都与数据质量相关，脏数据、死数据、不完整数据层出不穷，在影响数据质量提升的同时，也限制了数据的交换共享、数据模型梳理、信息资源整合、不同业务系统软件集成、业务协同等。如何提升数据质量，避免因劣质数据而带来的消极影响，已成为数据治理中亟待解决的基础性问题。

（四）一站式需求呼唤数据治理监管系统（数据母体）

市场因需求而存在。随着数据的重要性日益突出，数据质量检测、数据清洗比对、标准数据采集加工、数据模型架构等服务需求正呈现爆发式增长趋势。然而，由于当前市场缺乏专项产品和专业数据治理服务公司等，使得“如何对数据的质量进行又好又快的检测”、“数据质量问题怎样用可视化方式进行展现”、“数据质量问题如何进行整改”等问题无法得到及时、有效解决，人们迫切需要能满足一站式需求的数据治理监管系统。由此，国脉集团开发运营的“数据母体”应运而生。

二、产品概述

（一）产品定义

数据母体以标准体系为支撑，通过规则与算法，检测定位数据质量问题，给出整改意见与整改措施，以此来孕育高质量数据，形成一套完整的数据质量管理与监控体系。产品提供海量标准数据元采集/检索/下载、值域标准规范(规则库)、权威的国家/行业/地方标准，数据质量检测、清洗比对、标准化、诊断报告、数

据建模、知识图谱等系列服务。是政府、企业强化数据治理能力，提高数据质量水平，规范数据标准的重要工具和首选产品。

（二）主要模块

1、数据治理：

①表单标准检测：对数据库中的表单字段进行检测，把字段名称与公共数据元池中数据元作比对，检测表单的字段命名是否符合标准数据元命名规范，并对检测结果进行统计，用数据可视化进行显示。

②实体数据异常监测：该模块提供了重复率、空项率、位数、增量和脏数据出现率这几种检测类型，从全方位的视角来为用户解析数据质量。一套规则能在多套方案中复用，在保证多角度准确数据质量分析的前提下，大大减少了用户投入的精力与时间，为用户提供了一种灵活而全面的数据质量分析方式。

重复率：用于检查一个字段下的实体数据是否有重复；

空项率：用于检查非空字段是否存在空值；

位数：用于检查字段的实体数据是否符合位数规范；

增量：用于检查数据的及时性，衡量数据上传是否及时；

脏数据出现率：用于检查字段实体数据的取值范围和格式是否规范，支持数值型、字符型、日期型字段检查。

③综合诊断

把进行过表单标准检测与实体数据异常监测过后的部门，通过排行榜的形式进行展示，用来体现各部门数据质量管治情况，设置不同的诊断周期，通过部门排名的高低，促使各部门对数据进行整改，提升数据质量。

2、标准体系：

①公共数据元池

公共数据元池通过对数据元的中文名称、数据格式、值域等要素进行清洗比对、智能标注关联，能快速构建各种领域、行业、主题的标准数据元基础；通过统一数据标准定义依据，匹配数据统计口径，解决数据混乱，打破数据壁垒，促进信息跨地区、跨行业融合。

②标准资源库

标准资源库对国家层面、地方层面及行业层面的标准文件、术语库、分类编

码等进行多维度分装入库,提供该类数据的一站式服务,通过对标准文库的共享、转换、互操作、质量管理、内容加工等服务提升文件使用价值。

③规则库

数据质量规则是数据质量审核的逻辑校验标准,是数据质量检测的基础。规则库作为数据质量检测过程中必不可少的一部分,包含了数据质量检测所需要用到的各种规则。通过对值域的梳理,设定好需清洗数据的取值范围,对不符合规则的实体数据进行判断,识别出脏数据,以此来进行专业的数据规范、标准数据比对等标准化处理,提升数据质量、规范数据交换底层。

④综合主题词库

综合主题词库从基本主题、职能主题、行业主题、服务主题、地址主题、领域主题等六个角度构建出六套标准语义数据资源模型。用户可通过主题词典系统选购主题语义资源,生成支持各类政务活动的资源架构。

(三) 产品价值

1、助力数据问题定位和数据质量升级

依托数据母体,对数据质量进行在线检测和可视化统计分析,有助于从底层数据层面进行排查、追溯和定位数据质量问题发生的症结所在,包括字段名称规范性,实体数据存在重复项、空项、脏数据等问题。把有问题的数据诊断后得出检测报告发给数据管理部门进行整改,以此提高数据管理部门的数据质量。

2、快速抽取实体数据的同时保障实体数据安全

数据母体通过连接数据库或前置机的方式抽取实体数据,支持海量数据的高吞吐量。另外数据母体具备完整的角色权限体系,提供数据保护功能,在快速抽取实体数据的同时保证安全性与保密性。

3、灵活的数据可视化分析报告

数据母体能够快速得出数据质量问题的可视化分析报告。页面提供了丰富的展示形式,轻松地把静态数据转化成动态的图表,将检测结果进行有效的整合,快速准确的反应问题所在并提出整改的依据。

4、实现海量数据元的存储管理与应用

数据母体支持海量数据元存储,用户可根据需要批量导入数据元于其个人专项数据库或系统,对其数据元资产进行存储,并利用平台提供的数据元资产统计、

智能关联等功能，有效管控数据元。用户还可以根据自身需求，选取组合标准数据元，快速构建出各种领域的标准数据元基础库。

5、推动数据标准规范的落地与优化

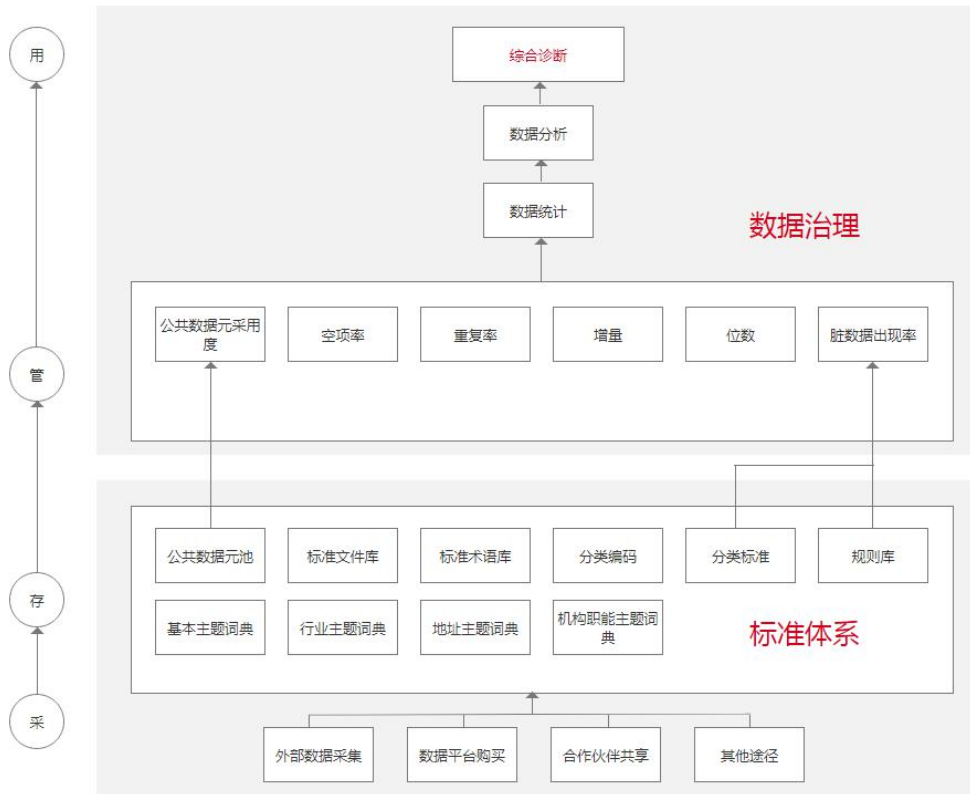
数据母体收录了千余份涵盖权威国际、国家、行业、地区的标准文件、数据元及数据元值域标准规范、标准术语等，并持续扩充，借助数据标准化处理，推动数据标准落地；用户对数据标准的选择、反馈等，也有助于各行业数据标准的优化改造，并进一步提升“上层建筑”数据、信息资源、事项等的标准化。

三、产品数据体系与技术体系说明

（一）数据体系

数据母体从数据获取、数据加工、数据关联、数据存储、数据应用服务、数据统计分析等方面构建标准、高效、智能的数据体系，从“采、存、管、用”全面处理数据。

目前已搜集标准文件 1764 份，梳理数据元 372883 条，梳理各种规则（值域）5329 条，另外梳理了行业主题词典（国民经济行业分类）16180 个，标准术语 1909 条，分类编码（ICS 国家标准分类号）1304 条，另外整理了江浙沪地区的地址主题词典（确定到社区）63895 条。



数据母体数据体系示意图

(二) 技术体系与技术框架

1、技术体系

数据母体采用当前主流的体系结构，支持跨平台、跨数据库应用。数据库采用 Mysql 关系型数据库作为数据库服务。平台兼容 Linux，中标麒麟等多种操作系统平台，应用开源 Web、应用服务器等成熟中间件产品进行系统设计，保证系统技术路线先进、可行、合理、成熟。

技术路线



数据母体技术路线图

2、技术框架

(1) 系统基于 Spring Cloud 框架，采用微服务开发模式，支持后端服务分布式部署，支持跨服务调用；简化配置管理、服务治理、断路器、动态路由、集群状态管理等操作。

(2) 支持 OAuth2.0 统一认证协议，结合网关等授权形式，保证访问的安全性。提供多种流程，不同客户端按照实际情况选择不同的流程获取 access_token。支持移动设备、第三方设备，具有较强地拓展性。

(3) 采用 Docker 容器为基础的部署方式和虚拟化技术，简化部署流程与操作，提升系统的稳定性；提升系统的总体的管控能力和后续维护的便利性，提升按服务独立维护的灵活性。

(4) 采用 vue 框架，进行前后端分离，支持前端服务独立部署，支持组件化的开发工作。

四、产品应用场景

(一) 快速检验数据库表单字段的规范性

基于公共数据元池，对数据库表单字段的规范性做检测，对于采用公共数据元的字段，系统打上标记，并显示采用的公共数据元名称及来源，对于没有采用

公共数据元的字段，系统也会标记出来。检验的数据库表名称、字段总数、检测字段数量、采用公共数据元的字段、没有采用公共数据元的字段、公共数据元采用度等信息都会有可视化的图表统计。

（二）自定义规则配置检测数量庞大的脏数据

XX市大数据中心累计梳理全市61个单位的政务信息资源,1205个数据集,2.69万余个有效数据项。在已经完成梳理的单位中,已接入公安局、人社局、卫计委、教育局、环保局、安监局等61个单位632个数据集1.25万个数据项,总数据量达10.78亿余条。面对如此庞大的数据量,客户需要有系统可以对数据库的数据进行快速抽取,并在系统中选择需要检测的字段,然后配置规则来对脏数据进行检测,最后得出相关的报告和整改意见。数据母体可以通过连接数据库或者连接前置机的方式,快速抽取数据,用户也可以在系统中选择需要检测的字段,然后通过预先建设的规则库,来检测字段的实体数据中是否存在脏数据,系统会有可视化的信息统计与脏数据错误的原因。

（三）快速检测实体数据的重复率

数据库当中很多数据存在重复的情况,如何把重复的实体数据快速检测出来,重复的数量有多少,以何种方式进行展现。这个时候就需要有专业的系统去帮助客户进行检测,并通过可视化的信息统计帮助客户直观的、形象地了解数据的情况,以便于客户能及时、准确地把握和调整数据治理监管的方向。数据母体就可以在这个场景下得到应用。

（四）快速检测实体数据的空项率

一些数据库表字段填写数据的时候因为相关属性缺乏,在数据库中只能用Null来表示。空值的存在会使数据质量大打折扣,降低数据价值。因此我们需要系统专门对实体数据的空项率进行检测。先把数据库中要检测的表梳理出来,再选择每张表需要检测空项率的字段,然后开始检测空项率,最后得出哪些字段存在空值,空值的数量有多少等数据,形成可视化统计报告。

（五）快速检测实体数据的位数长度是否符合要求

XX市之前对全市身份证号码进行了检测,发现了500万个身份证号码中,位数长度只有17位的,就有20多万个,这个只是其中一个错误,需要技术人员专门写程序来做检测,费时费力。单单500万个身份证号的数据量,光检测就花

了 2 天的时间，这还不算开发的时间。如果有一个已经开发好的系统，既能省去开发的时间，又能使检测的时间大大缩短，同时保证了数据的安全性。

（六）对于数据新增是否及时的监测

随着时间的推移，每个部门或多或少都会有数据新增，有的更新周期长，有的更新周期短。对于数据管理部门来讲，他们希望能够随时了解自己每个周期内数据新增的情况，对于数据新增不及时的问题，他们能及时做出应对。而数据母体可以通过实时监测技术来实现这个需求。

（七）快速建模生成数据元基础库

基于场景需求定制，特别是各地政府对数据元采集、整理、建库的规划需求，通过标准数据元组合，可快速建模形成法人、人口等数据元基础库。用户也可以选择不同领域的数据元，按照自身的需求来定制数据元基础库，领域涵盖了医药卫生、公共安全、环境保护、城镇建设、财经金融、新闻出版、教育、海关、交通、旅游、粮食、邮政、文化、通信、石油化工、航天、农业、林业、民政、烟草、船舶、汽车、气象等领域，可支持通过以 XLSX、CSV、API 接口的方式给予客户数据库授权服务。

（八）辅助构建标准核心数据元池

核心数据元池建立，是各地政府信息资源标准化的基础。根据实际需求，梳理出核心数据元，并通过数据元清洗比对、标准化处理和智能标注关联，构建标准的核心数据元池。

（九）行业数据元评估清洗标准化

行业数据质量不高和存在交换壁垒的一大原因在于数据元不规范、不统一。使用数据母体，可对各行业数据元标准化程度进行检测评估与认证；同时根据需求，可对评估质量低的行业数据元进行去重、同义词解析等清洗操作和基于专业数据元规范、标准数据元比对等进行标准化处理，助力提升数据元质量、规范数据交换底层。如国家有关部委、地方及中小企业等信用库、法人库、基础库、材料库、政策库数据元标准化建设等。

（十）基于数据服务创建应用模型

数据元作为数据的基本单位，与数据服务质量好坏息息相关。围绕不同的应用主题设计数据模型，并支持数据模型可扩展，满足不同业务需求，统一数据接口

径，从而构建以应用场景为维度的标准数据模型库。

五、成功案例

国脉软件已为 100+政府客户所信赖和应用，其中：

- 国防科工局数据治理平台建设项目；
- 浙江省公共数据目录系统项目；
- 海南省政务信息资源共享采集云项目；
- 吉林省政务信息资源编目项目；
- 山东省政务服务标准化项目；
- 广东省交通运输厅一体化数字平台项目；
- 杭州市政府数字化转型督查考核系统；
- 德州市“一次办好”事项数据化处理和流程优化项目；
- “i 厦门”运营监督系统建设项目；
- 深圳市政务服务事项数源供需对接项目；
- 滨州市营商环境第三方评估服务；
- 江阴市大数据中心政务数据质量治理平台；
- 广州市工业和信息化委员会政府数据资产登记服务项目；
- 上海市“一网通办”三清单；
- 绍兴市数据资源目录管理系统项目；
- 珠海市政务服务数据管理政务数据资产普查及珠海“数字政府”改革建设项目；
- 佛山市目录普查项目；
- 上海徐汇区“高效办成一件事”建设项目；

- 北京西城区政务服务办“一网通办”信息化建设项目；
- 上海浦东区政策服务平台建设项目；
- 深圳南山区政务服务流程再造动态管控平台、政策精准送与项目一次办受理系统；
- 佛山禅城区事项治理系统.....

六、公司介绍

（一）关于国脉

国脉，是数字政府、智慧城市、数字经济、营商环境、政务服务、（大）数据治理专业提供商。创新提出“软件+咨询+数据+平台+创新业务”五位一体服务模型，拥有一网通办核心支撑系统（GDBOS）、数据基因（DNA）、营商通（MAT）、我能平台（IEP）四大系列百余模块的软件产品，长期为中国智慧城市、智慧政府和智慧企业提供专业咨询规划和数据服务，运营电子政务、物联网、智慧城市、数邦客、中船通、蟠桃会、仓颉数源、国策数据等系列行业专业平台，广泛服务于信息中心、大数据局、行政服务中心等政府客户、中央企业和金融机构。

从2004年12月成立以来，已在全国七大区域设立20余家分支机构、5大技术研发基地、国脉研究院（智慧城市研究中心、数字化转型研究中心、数字政府研究中心、营商环境评价中心、互联网+政务评价中心、数字经济研究所、政策数据研究所等），投资建设国脉智慧产业城；服务客户2000余家，涵盖国家部委、省/市/区县级政府、上百家央企及部分世界500强企业等；执行项目5000余个，并帮助客户积极打造智慧城市、数字政府、“最多跑一次”、“互联网+政务服务”等项目创新标杆，引领组织变革、数据治理、流程再造新浪潮；连续多年开展中国政府网站评估（自2006年）、智慧城市发展水平评估（自2011年）、互

联网+政务评选（自 2015 年）、营商环境评估评选（自 2018 年）、数字政府评估评选（自 2019 年）等公益活动。

被业界誉为中国信息化民间智库知名品牌、电子政务优选咨询机构，国内首倡智慧政府理念，首创智慧城市、数据治理、互联网+政务服务评价体系，首推数据资产普查、全口径数据资源目录、数据元标准化、数源确认与供需对接、“最多跑一次”事项梳理、业务流程再造动态管控系统等产品，信息资源编目、公共数据普查等业务全国占有率和影响力名列前茅。在积极拓展外延发展的同时，国脉也正加速自身内生变革和组织数字化转型，争做赋能共益组织；并聚焦面向未来的能力建设和“价值+协同”，打造“我能平台”新物种，积极引领并赋能智慧组织构建。

（二）服务优势

1、专家·专业

——**中国最高端信息化与互联网研究专家网络。**国脉坚持“专家、专业、专注”的公司定位，恪守“网络价值构建师、数据治理架构师”的企业愿景，信奉“智慧就是力量”。国脉创始人均为信息化、电子政务和智慧城市领域知名专家，并担任多个部委、省市及行业协会的专家委员，来自国家信息化专家咨询委员会、国家行政学院、国家信息中心、中国社会科学院、中国互联网协会、中国信息化百人会、清华大学、浙江大学等机构和高校的专家常年为国脉智库提供外脑支持，成熟完备的国脉内部专家培养体系为行业源源不断地输送精英咨询研究团队。

——**专业、强大的本地化实施团队。**公司在北京、上海、广州等全国 20 多个地方开设了分支机构，并组建了具备专业、一定实力的本地化团队，致力于为政府机构、城市园区、大型企业等客户提供科学、专业、创新的高品质服务。

2、高端·领先

——**信息化顶层规划优势。**国脉受国务院办公厅、中央网信办委托参与十三五信息化规划、电子政务及政府网站未来发展思路、规范编制；也是参与《政府网站未来五年发展指引》规划研究、中央政府门户网站规划及未来三年规划研究、国办发[2014]57号文《关于加强政府网站信息内容建设的意见》起草的咨询机构。

——**“互联网+政务服务”咨询案例国内最优。**2018年1月，习近平总书记主持召开中央深改组第二次会议，对浙江“最多跑一次”改革给予了肯定。国脉作为打造国内最优秀的政务服务平台和国内最优秀的“互联网+政务服务”浙江案例，不仅参与了浙江政务服务网规划与设计、政务服务产品优化研究与推广、政务服务事项标准化与事项梳理，还参与了“最多跑一次高频事项”数据串梳理、事项认责、数源确认以及供需对接平台产品的提供。除浙江省之外，为北京市、广东省、山东省、海南省等省级政务服务平台以及多个信息惠民国家试点城市及区县政务服务项目提供“咨询+产品+标准+数据”等服务。

3、创新·可靠

——**“咨询+产品”一体化的政务服务与数据服务。**政务服务、数据治理领域涉及顶层设计、目录构建、应用建设、标准规范保障等一套理想的解决方案，需对各地区、各行业的数据业务、标准有着深刻的理解，国脉基于十余年咨询经验，率先推出数据基因等产品，并提供“咨询+产品”一体化数据服务，提升信息资源管理规范和使用水平。

——**自主研发、安全可靠的数据产品。**数据基因系统等产品按照实现系统与数据分离、字段级按需共享、灵活数据开放等理念，在吸收最新的国家标准、行业标准以及最新的数据管理理念和思想后，充分考虑到政府、行业业务开展和自身本身的特点研发而成，具有极强的创新性、针对性、标准性和实用性，能够快速提升本地区、本行业的数据治理水平。

——**获得市场的广泛认同。**国脉数据整体解决方案成熟、建设经验丰富，落地案例覆盖省、市、县三级，目前已为浙江、海南、江苏、贵州等多地客户提供数据相关服务。建设

时间短，3-4周快速建立地区信息资源体系；运用效果好，实现数据流动、动态管理，得到客户的一致认可。

七、联系我们

北 京：010-82617379 / 68882870

广 州：020-37958159

上 海：021-50899737

佛 山：0757-82915007

南 京：15951611566

深 圳：15011758853

杭 州：13777819956

重 庆：023-67271746

宁 波：0574-27830655

海 南：13807673398

舟 山：0580-2299661 / 2299669

长 春：13944125008

长 沙：15273455628

石家庄：0311-67501282

衡 阳：0734-8158822 / 8158833

青 海：0971-6285860

江 西：0791-83833559 / 83833669

云 南：15211837779