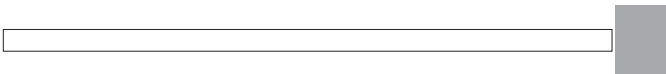


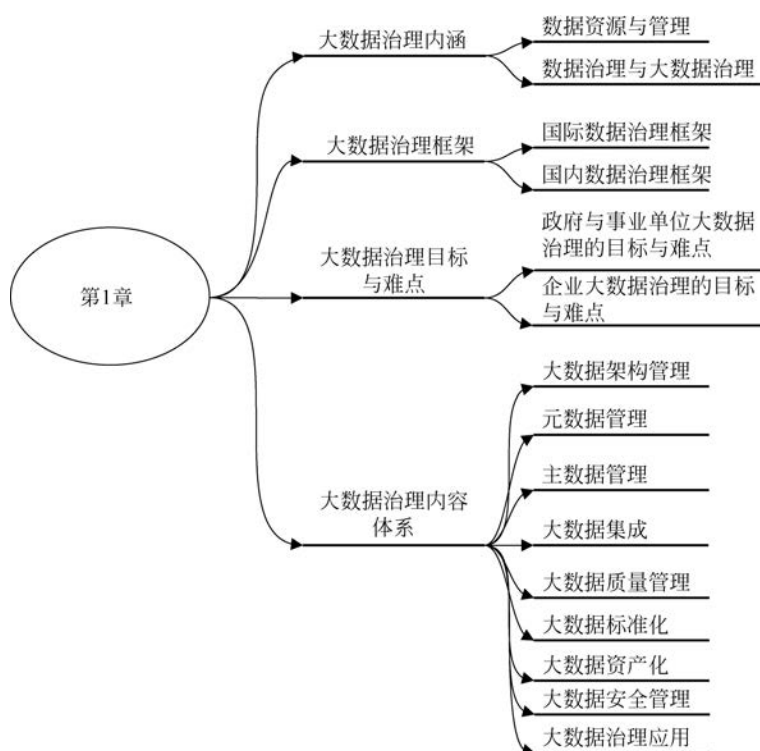


第一篇 大数据治理理论





思维导图



内容提要

随着大数据时代的到来,数据规模急剧扩大,数据治理方式也在不断地革新与变化。大数据给传统数据治理工作带来了重大挑战,需要与之相适应的数据治理理念、治理框架、治理标准、治理模式、治理理论方法等。大数据治理是智慧社会治理的基础,也是组织实现数字化战略的前提。本章将重点介绍大数据治理的内涵、框架、目标与难点以及内容体系。



本章重点

- ◆ 掌握大数据治理的内涵。

- ◆ 理解大数据治理的框架。
- ◆ 厘清大数据治理的目标与难点。
- ◆ 了解大数据治理的内容体系。

1.1 大数据治理内涵

大数据治理是针对大规模、跨组织、异构多源、多模态数据资源和资产进行标准化、融通、关联、解析、聚合等一系列活动的集合,其目的是在保障大数据安全的基础上有效提高大数据资源和资产的可用性、易用性和安全可靠性,实现大数据资源和资产价值的显著提升。大数据治理是实现数据驱动管理决策的基础,也是智慧社会治理的重要组成部分。大数据治理是一个体系工程,往往涉及不同组织、不同部门、不同系统间的信息共享、交换、融合、组织和管理控制行为活动,需要法律法规、理论、方法、技术等一系列支撑。通过多维多层次跨域大数据治理,为政府、企事业单位的精细化管理和精准决策提供有力支撑,方便人们对信息的获取、管理和利用,为数智化社会治理和经济社会高质量发展奠定重要基础。

1.1.1 数据资源与管理

1. 数据与数据资源

数据是事实或观察的结果,是对客观事物的逻辑归纳,是用于表示客观事物的未经加工的原始素材。在计算机科学中,数据是所有能输入计算机并被计算机程序处理的符号的介质的总称,是用于输入电子计算机进行处理的具有一定意义的数字、字母、符号和模拟量等的通称。数据经过加工后就成为信息。

数据资源是可供人类利用并产生效益的一切记录信息的总称,属于一种社会资源。资源是动态的,数据资源是人类从工业社会进入信息社会的产物。但数据资源并非单一数据,而是可利用或可能被利用的数据集合。作为数据资源的数据,应当具有一定的数量和可用的质量,才能够满足特定的需要。

数据资源作为新型资源,具有区别于传统资源的特性:一是无形性。非物质性和无形性使得数据资源被传统物权所排斥,因而无法成为传统物权的客体。基于此,数据资源可以被他人近乎零成本、快速地、无次数限制地复制,可以跨越时空限制而为社会公众所共享、共用,且不会发生有形的损耗。二是可变性。数据资源形成和流通的过程意味着数据资源总是处于变化之中。同时,数据流通过程中的每一个事物特征和活动状态也都可能形成新的数据资源。此外,数据资源也会基于市场主体的不同需求,或者数据生命周期而发生相应变化。三是社会性。传统意义下的自然资源具有社会性,意味着自然资源参与到整个社会关系中,为全体人类所共享,自然资源的开发利用及消耗最终追求的都是社会福利的增加。而数据资源尽管也参与到了整个社会关系中,但数据资源的获取、处理及利用总是与对数据资源有需求的社会主体密切相关,人类认识和掌握数据资源也是一个社会过程。基于此,当讨论数据资源归属时,更多的是需要考虑数据资源的持有、使用和经营,而非所有。四是共享性。数据资源在使用上具有非竞争性和非排他性,一个人使用特定数据资源客观上并不会妨碍另一个人使用相同的数据资源。因此,由于额外用户使用它们的边际成本为零,福利最

优的解决方案是免费授予对这些数据资源的一般访问权限,在这个意义上,数据资源可以说是一种公共物品,其本质就是分享与流通。

2. 数据资源管理

数据资源管理(data resource management)是应用信息技术和软件工具完成组织数据资源管理的文件处理方法,在这种方法中,数据根据特定的组织应用程序的处理要求被组织成特定的数据记录文件,只能以特定的方式进行访问。这种方法在为现代企业提供流程管理、组织管理信息时显得过于麻烦,成本过高并且不够灵活。因此出现了数据库管理办法,它可以解决文件处理系统存在的问题。

1.1.2 数据治理与大数据治理

1. 数据治理

数据治理在概念上有狭义和广义之分。

狭义的数据治理是指对数据质量(data quality)的管理、专注于对数据本身的分析,包括数据资源及其应用过程中相关管控活动、绩效和风险管理的集合,以保证数据资产的高质量、安全及持续改进。狭义的数据治理的驱动力源自两个方面:①内部风险管理的需要,风险包括数据质量差影响关键决策等;②满足外部监管和合规的需要。随着全球越来越多的企业认识到信息资产的重要性和价值,数据治理的目标也在发生转变。除满足监管和风险管理要求外,如何通过数据治理来创建业务价值也备受关注。

广义的数据治理是对数据资产管理行使权力和控制的活动集合(规划、监控和执行),指导其他数据管理职能如何执行,在高层次上执行数据管理制度。对数据的全生命周期进行管理,包含数据采集、清洗、转换等传统数据集成和存储环节的工作,同时包含数据资产目录、数据标准(data standards)、质量、安全、数据开发、数据价值、数据服务与应用等,整个数据生命期开展的业务、技术和管理活动都属于数据治理的范畴。

2. 大数据治理

大数据(big data, mega data),亦称“巨量资料”,指的是所涉及的资料量规模巨大到无法通过主流软件工具,在合理时间内达到摄取、管理、处理并整理成为帮助目标主体达成管理目的的信息。IBM(国际商业机器公司)提出了大数据的“5V”特征,即规模性(volume)、多样性(variety)、高速性(velocity)、价值性(value)和真实性(veracity)。规模性是指其数量大,能够容纳大量人员和物资的信息。多样性表示数据来源及数据形式的多样化。高速性反映了信息在时间维度上的即时即达。价值性体现为价值密度低,能够从不同来源的海量信息里抓取到有价值的细节。真实性是数据准确性以及可信赖度的表现,即数据质量高且能够解释和预测现实事件的过程。

大数据治理的工作重心在于与大数据相关的数据优化、隐私保护与数据变现的政策,其主要特点在于治理对象和治理结构的复杂性。大数据治理对象的复杂性是指大数据以规模化、多样化、异构、多模态的形式出现在各类信息系统中,数据本身的复杂度较高,需要新的法律法规、理论、管理制度以及更高级的技术方法进行数据共享、数据集成、数据分析处理和数据服务。大数据治理结构的复杂性是指主体的多元化、跨域跨组织,以及组织内部部门之

间、部门上下级之间形成纵向和横向的复杂网络结构所带来的沟通与协调成本往往较高,无形中为不同组织间的数据共享提升了难度和不确定性。

1.2 大数据治理框架

随着人类社会数字化水平的提升,数据规模不断扩大,数据治理的难度不断提升。为有效地进行数据治理,国内众多组织基于数据治理理论和实践,制定了多种数据治理框架和标准,对于组织数据治理体系的建设和数据治理实践有着重要的参考意义。本节将重点介绍国内外主要的大数据治理框架与标准。

1.2.1 国际数据治理框架

1. ISO 数据治理标准

国际标准化组织(ISO)于2008年推出第一个IT(信息技术)治理的国际标准——ISO 38500,随后在2015年巴西会议上形成决议,将数据治理国际标准分为两个部分:ISO/IEC 38505-1《基于ISO/IEC 38500的数据治理》(以下称ISO 38505-1)和ISO/IECTR 38505-2《数据治理对数据管理的影响》。

ISO 38505-1 标准的核心内容包括:①数据治理的目标是促进组织高效、合理地利用组织数据资源。②数据治理的基本原则包括:职责、策略、采购、绩效、符合和人员行为。③数据治理的核心任务包括:一是明确了数据治理的意义、治理主体的职责、数据治理的监督机制;二是对治理准备和实施的基于“价值、风险和约束”的数据治理方针和计划指导;三是进一步明确数据治理的“E(评估)—D(指导)—M(监督)”方法论,提出了数据责任矩阵表及其应用方法。

ISO 38505-1 数据治理标准所提出的EDM模型如图1.1所示。该图展示了数据治理主体与数据治理之间的关系,以及组织实施数据治理时的内外部环境 and 数据利用的压力等。利益相关方包括客户、员工以及所有利益关联的监督者。同时,图1.1展示了“评估—指导—监督”循环模型需要的与数据相关的输入类型。

1) 内部需求

管理机构将制定业务的总体战略,审查数据的潜在用途,根据组织本身需求或竞争对手情况,调整战略方向,以支持预期的结果。企业将围绕组织的战略目标塑造数据文化,以确保数据治理策略达到其总体目标。由于数据与决策一样有价值,因此这种数据文化需要的数据访问、良好数据相关的组织行为处理,依赖于相关环境中的所有的做法和决策过程。

2) 外部压力

企业可能需要调整战略和政策,以确保其符合外部市场的压力对其的作用,这样的市场压力包括:客户对可用数据的可用性、治理和交互的期望;竞争对手使用数据来改进或扩展其产品、服务或流程;如何收集数据,包括有关收集和使用个人信息的隐私通知和同意的要求;数据保留和处置要求;适当处理偏见、歧视和定性的决策义务;有关共享或重用数据的自身产权问题。

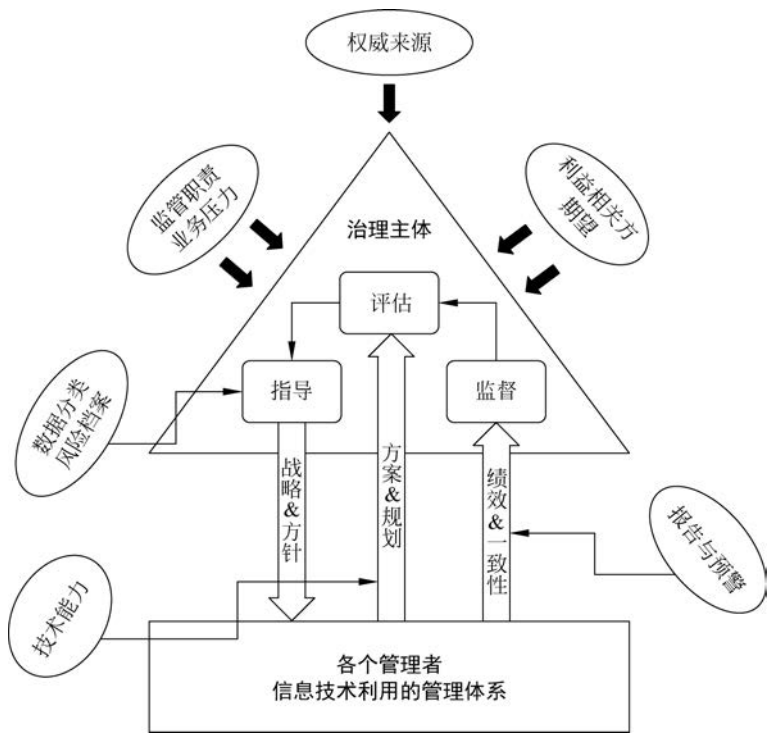


图 1.1 ISO 38505-1 数据治理标准所提出的 EDM 模型

3) 评估

在评估企业数据治理时,应考虑到组织的内部要求和外部压力。此外,应审查和判断目前和未来数据的管理和使用情况,例如:数据和相关技术与流程的内部使用情况;竞争对手、其他组织、政府和个人使用的数据;评估不断发展的一系列立法、法规、社会期望;控制并影响数据使用的其他因素。同时,历史机构还应理解组织的数据管理能力,例如:企业可以从数据泄露中恢复到什么程度;以正确的格式轻松传递正确的信息,以帮助决策的制定;组织是否利用云计算等新技术来增强自己的能力。

4) 指导

组织应当负责并指导数据战略的政策制定和执行,并且遵循以下原则:首先,应最大化企业对数据投资的价值。企业对数据的投资包括对数据的采集、管理、存储等。数据是一种有价值的资产,作为有用的信息可以进行分发和销售,通过订阅、出版物或网站等渠道馈送销售并获得货币价值。其次,应根据数据风险偏好管理与数据相关的风险。某些数据(如产品研究或未披露的股票市场相关数据)具有很高的业务价值,需要使用资源来利用和保护此数据。最后,应确保组织的数据管理水平。组织应对数据的管理、使用,以及根据数据作出的决策等进行全过程治理,数据治理问责机制应被委派给适当的组织。

5) 监督

组织应通过适当的系统测量,监测数据的使用情况。确保数据被有效地融入企业战略实施过程中,并确保数据的使用和管理符合内部测量、外部法规和数据管理的要求。由于战略或条例的要求,组织的监督可能包括其他非常重要的领域:隐私问题、数据使用的透明

度；有效的信息安全管理系统；数据的保留和处置要求；数据的再利用、共享或出售及其相关权利、许可或版权；在决策中适当考虑文化规范、偏见、歧视或相貌。

2. DGI 数据治理框架

国际数据治理研究所(DGI)是业内最早、世界上最知名的研究数据治理的专业机构。DGI 在 2004 年推出了 DGI 数据治理框架,为企业数据管理的战略决策和行动提供最佳实践与指南。该框架认为,企业决策层、数据治理专业人员、业务利益干系人和 IT 领导者可以共同制定决策和管理数据,从而实现数据的价值最大化和成本最小化,管理风险并确保数据管理和使用遵守法律法规与其他要求。

DGI 框架是一个十分具有实践指导意义的数字治理模型,它是从组织数据治理的目标或者需求出发进行设计的,采用 5W1H 法则,即 why、what、who、when、where、how,描述了谁可以采取什么行动来处理什么信息以及何时在什么情况下使用什么方法。该框架将数据治理分为人员与治理组织、规则、流程三个层次,共 10 个组件:数据利益相关方、数据治理办公室、数据治理愿景,数据治理目标、评估标准和推动策略,数据治理规则与定义,数据治理决策权,数据治理问责制,数据治理管控措施;数据治理流程。DGI 数据治理框架如图 1.2 所示。

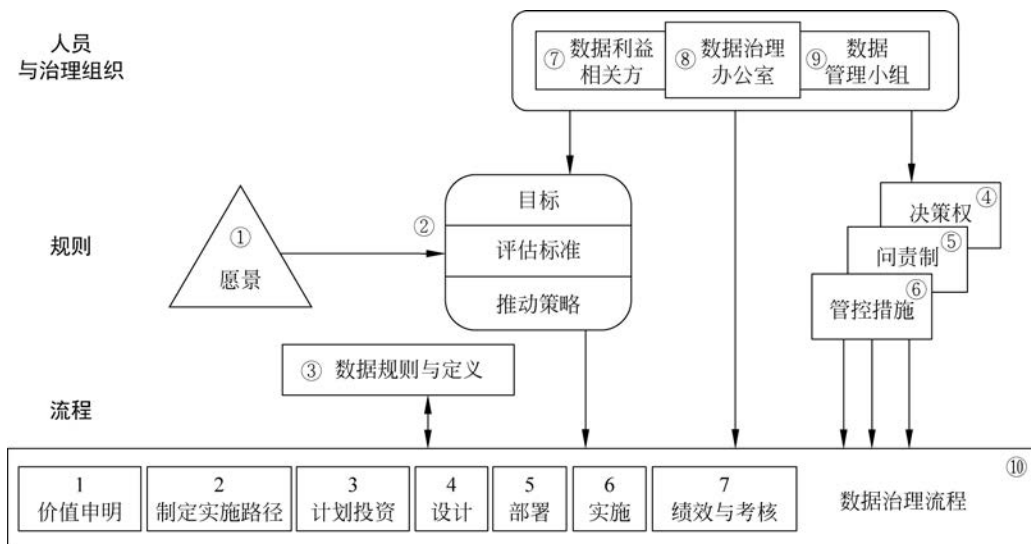


图 1.2 DGI 数据治理框架

1) why: 数据治理原因

DGI 框架中的第 1~2 个组件: 数据治理愿景,数据治理目标、评估标准和推动策略。这两个组件用来定义企业为什么需要数据治理,为企业数据治理指明了方向,是其他数据治理活动的总体策略。

(1) 数据治理愿景。这是企业实施数据治理的最高指引,为数据治理活动指明方向。其一般包含三个部分: 定义数据管理和使用规则; 为相关干系人提供数据安全保护和服务; 对不遵守规则引起的数据问题作出反应并解决。这三个部分定义了数据治理的使命——数据治理干什么,是最高数据治理立法。

(2) 数据治理目标、评估标准和推动策略。数据治理从来都不是为治理数据而治理数据,而是为了解决数据利益相关方的业务痛点或实现某些业务和管理目标。企业数据治理目标一般包含但不限于:致力于政策、标准、战略制定的数据治理;致力于数据质量的数据治理;致力于隐私/合规/安全的数据治理;致力于架构/集成的数据治理;致力于数据仓库与商业智能(BI)的数据治理;致力于支持管理活动的数据治理。此外,数据治理的实施需要获得数据利益相关者的支持,参与数据治理的每个人都应该知道数据治理目标是什么,以及如何衡量是否达成目标。

2) what: 数据治理内容

DGI 框架中的第 3~6 个组件,包括数据规则与定义、数据的决策权、数据问责制、数据管控措施,这 4 个组件定义了数据治理到底治什么。数据规则与定义,侧重业务规则的定义,例如:相关的策略、数据标准、合规性要求等;数据的决策权,侧重数据的确权,明确数据归口和产权为数据标准的定义、数据管理制度、数据管理流程的制定奠定基础;数据问责制,侧重数据治理职责和分工的定义,明确谁应该在什么时候做什么;数据管控措施,侧重采用什么样的措施来保障数据的质量和安全性,以及数据的合规使用。

(1) 数据规则与定义。定义数据相关的策略、标准、合规性要求、业务规则等。数据规则与定义即定义数据标准,是实施数据治理的基础,其主要任务包括:首先,收集数据标准需求,包括现行的数据标准、业务规则以及数据使用过程中默认的规则——厘清现状。其次,对每个数据域的实体数据定义数据标准,包含数据的业务含义、质量规则、存储标准、合规性要求等。再次,需要对齐并优先处理冲突的规则,如果数据标准存在冲突,应协调数据的利益相关方协商解决。最后,发布数据标准,在企业范围内达成共识。

(2) 数据的决策权。决策权就是数据的确权,也就是明确数据的归口管理部门/岗位。在创建任何数据标准或作出任何与数据相关的决策之前,必须先解决什么时候、使用什么流程、由谁来作出决策等问题。

(3) 数据问责制。数据治理规则/标准一经发布,就要着手实施,数据治理团队需要将相关活动执行任务分配至日常工作中。

(4) 数据管控措施。数据治理需通过建立风险管理策略,来控制 and 预防数据安全风险。管控措施可以是预防性的,也可以是检查或纠正性的。同时,还可以建立或改进现有数据控制措施,例如:变更管理、策略、培训、项目管理等,以支持数据治理目标的实现。

3) who: 数据治理的利益干系人

DGI 框架中的第 7~9 个组件,定义数据治理的利益干系人,主要包括:数据利益相关方、数据治理办公室和数据管理小组。DGI 框架对数据治理的主导、参与的职责分工定义给出了相关参考。

(1) 数据利益相关方。数据利益相关方是可能会影响或受到所讨论数据影响的个人或团体,如某些业务组、IT 团队、数据架构师等,他们对数据治理项目的目标会有一个更加准确的定位。

(2) 数据治理办公室。数据治理办公室的职责就是促进并支持数据治理和数据管理的相关活动,例如:执行数据治理程序;阐明数据治理和管理活动的价值;提供与业务和数据计划的关系,如数据质量、法规遵从性、隐私、安全性、体系结构和 IT 治理;从数据利益相关方收集并调整政策,标准和指南;支持和协调数据管理的相关会议;为 IT 项目提供数据和分析;为利益相关方提供数据治理政策的培训、宣传等活动。

(3) 数据管理小组。很多企业的数数据管理委员会可能会分为几个团队或工作小组,以解决特定的数据问题。数据管理小组负责特定业务域的数据质量监控和数据的安全合规使用,根据数据的一致性、正确性和完整性等质量标准检查数据集,发现并解决问题。同时,数据治理小组也有义务向 DGO 提出数据治理策略、数据标准等方面的建议或意见。

4) when: 数据治理的时机选择

DGI 框架中的第 10 个组件,用来定义数据治理的实施路径、行动计划。

5) how: 数据治理流程

DGI 框架中的第 10 个组件: 数据治理流程,描述了数据管理的重要活动和方法。该部分强调主动、应对、持续的数据治理流程,描述数据治理的方法。数据治理的流程如下: 价值申明; 制定实施路径; 计划投资; 设计; 部署; 实施; 绩效与考核。

6) where: 组织数据治理的成熟度级别

DGI 框架外的组件,虽没有在 10 个组件之列,但却十分重要,强调明确当前企业数据治理的成熟度级别,找到企业与先进标杆的差距,是定义数据治理内容和策略的基础。

3. DAMA 数据管理框架

国际数据管理协会(Data Management Association International,DAMA)是一个由全球性数据管理和业务专业的志愿人士组成的非营利协会,致力于数据管理的研究和实践。其出版的《DAMA 数据管理知识体系指南》(简称 DAMA-DMBOK)一书目前已出版第 2 版,即 DAMA-DMBOK2。

DAMA-DMBOK2 理论框架由 11 个数据管理职能领域(图 1.3)和环境要素(图 1.4)共同构成“DAMA 数据管理知识体系”。每项数据职能领域都在七个基本环境要素约束下开展工作,按照一定的逻辑结构进行分析,保证数据治理的目标和实际商业过程的贡献。其用于指导组织的数据管理职能和数据战略的评估工作,并建议和指导组织去实施与优化数据管理。



图 1.3 DAMA 数据治理框架 11 个数据管理职能领域

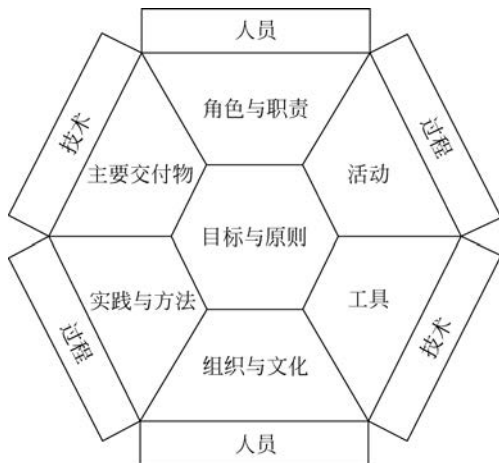


图 1.4 DAMA 数据治理框架七个基本环境要素

(1) 数据管理职能领域。其包括：数据治理,数据架构,数据建模和设计,数据存储和操作,数据安全,数据集成和互操作,文档和内容管理,参考数据和主数据(master data),数据仓库和商务智能,元数据,数据质量。

数据治理：通过建立一个能够满足企业数据需求的决策体系,为数据管理提供指导和监督。

数据架构：定义了与组织战略协调的管理数据资产蓝图,以建立战略性数据需求及满足需求的总体设计。

数据建模和设计：以数据模型的精确形式,发现、分析、展示和沟通数据需求的过程。

数据存储和操作：以数据价值最大化为目标,在整个数据生命周期中,从计划到销毁的各种操作活动。

数据安全：确保数据隐私和机密性得到维护,数据不被破坏,数据被适当访问。

数据集成和互操作：包括与数据存储、应用程序和组织之间的数据移动和整合相关的过程。

文档和内容管理：用于管理非结构化媒体数据和信息的生命周期过程,包括计划、实施和控制活动,尤其是指支持法律法规遵从性要求所需的文档。

参考数据和主数据：包括核心共享数据的持续协调和维护,使关键业务实体的真实信息,以准确、及时和相关联的方式在各系统间得到一致使用。

数据仓库和商务智能：包括计划、实施和控制流程来管理决策支持数据,并使知识工作者通过分析报告从数据中获得价值。

元数据：包括规划、实施和控制活动,以便能够访问高质量的集成元数据,包括定义、模型、数据流和其他对理解数据及其创建、维护和访问系统至关重要的信息。

数据质量：包括规划和实施质量管理技术,以测量、评估和提高数据在组织内的适用性。

(2) 基本环境要素。其包括：目标与原则,活动,主要交付物,角色与职责,实践与方法,工具,组织与文化。

目标与原则：每个职能在自己主题领域里的方向性目标,以及职能指标量化的基本原则。

活动：每个职能都由一个或多个活动组成,其中有部分活动能被细化为子活动。

主要交付物：信息、物理数据库即各职能在管理过程中最终输出的文档。

角色与职责：参与执行和监督职能的业务角色和 IT 角色,以及其各自职能中承担的具体责任。

实践与方法：包含常见和流行的实践方法,以及交付物的执行过程和步骤。

工具：各种配套支撑技术的类别、标准和规范、产品选择的标准和常见的学习曲线。

组织与文化：主要包括管理度量指标和标准、成功和商业价值的度量指标与标准等因素。

1.2.2 国内数据治理框架

1. GB/T 34960 规定的的数据治理规范

《信息技术服务 治理 第5部分：数据治理规范》是我国信息技术服务标准(ITSS)体系

中的“服务管控”领域标准,属于 GB/T 34960《信息技术服务 治理》的第 5 部分,该国标系列共包含以下五个部分:通用要求、实施指南、绩效评价、审计导则、数据治理规范。

根据 GB/T 34960.1—2017《信息技术服务治理第 1 部分:通用要求》中的治理理念,《信息技术服务 治理 第 5 部分:数据治理规范》在数据治理领域进行了细化,提出了数据治理的总则、框架,明确了数据治理的过程,可对组织数据治理现状进行评估,指导组织建立数据治理体系,并监督其运行和完善。

数据治理框架包含顶层设计、数据治理环境、数据治理域和数据治理过程四大部分,如图 1.5 所示。

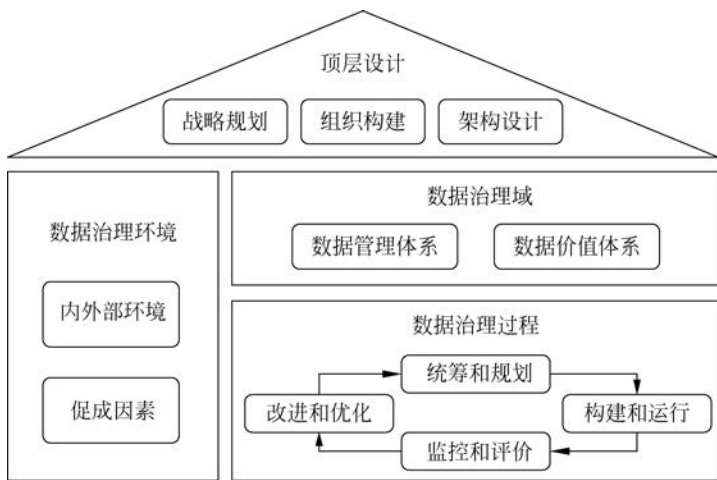


图 1.5 GB/T 34960 数据治理框架

(1) 顶层设计。其包含数据相关的战略规划、组织构建和架构设计,是数据治理的基础。

战略规划。数据战略规划应与业务规划、信息技术规划保持一致,并明确战略规划实施的策略,应该包括:①理解业务规划和信息技术规划,调研需求并评估数据现状、技术现状、应用现状和环境;②制定数据战略规划,包括但不限于愿景、目标、任务、内容、边界、环境和蓝图等;③指导数据治理方案的建立,包括但不限于实施主体、责权利、技术方案、管控方案、实施策略和实施路线等,并明确数据管理体系和数据价值体系;④明确风险偏好、符合性、绩效和审计等要求,监控和评价数据治理的实施并持续改进。

组织构建。组织构建应聚焦责任主体及责权利,通过完善组织机制,获得利益相关方的理解和支持,制定数据管理的流程和制度,以支撑数据治理的实施,应该包括:①建立支撑数据战略的组织机构和组织机制,明确相关的实施原则和策略;②明确决策和实施机构,设立岗位并明确角色,确保责权利的一致;③建立相关的授权、决策和沟通机制,保证利益相关方理解、接受相应的职责和权利;④实现决策、执行、控制和监督等职能,评估运行绩效并持续改进和优化。

架构设计。架构设计应关注技术架构、应用架构和架构管理体系等,通过持续的评估、改进和优化,以支撑数据的应用和服务,应该包括:①建立与战略一致的数据架构,明确技术方向、管理策略和支撑体系,以满足数据管理、数据流通、数据服务和数据洞察的应用需

求；②评估数据架构设计的合理性和先进性，监督数据架构的管理和应用；③评估数据架构的管理机制和有效性，并持续改进和优化。

(2) 数据治理环境。其包含内外部环境和促成因素，是数据治理实施的保障。

内外部环境。组织应分析业务、市场和利益相关方的需求，适应内外部环境变化，支撑数据治理的实施，应该包括：①遵循法律法规、行业监管和内部管控，满足数据风险控制、数据安全和隐私的要求；②遵从组织的业务战略和数据战略，满足利益相关方需求；③识别并评估市场发展、竞争地位和技术变革等变化；④规划并满足数据治理对各类资源的需求，包括人员、经费和基础设施等。

促成因素。组织应识别数据治理的促成因素，保障数据治理的实施，应该包括：①获得数据治理决策机构的授权和支持；②明确人员的业务技能及职业发展路径，开展培训和能力提升；③关注技术发展趋势和技术体系建设，开展技术研发和创新；④制定数据治理实施流程和制度，并持续改进和优化；⑤营造数据驱动的创新文化，构建数据管理体系和数据价值体系；⑥评估数据资源的管理水平和数据资产的运营能力，不断提升数据应用能力。

(3) 数据治理域。其包含数据管理体系和数据价值体系，是数据治理实施的对象。

数据管理体系。组织应围绕数据标准、数据质量、数据安全、元数据管理和数据生存周期等，开展数据管理体系的治理，应该包括：①评估数据管理的现状和能力，分析和评估数据管理的成熟度；②指导数据管理体系治理方案的实施，满足数据战略和管理要求；③监督数据管理的绩效和符合性，并持续改进和优化。

数据价值体系。组织应围绕数据流通、数据服务和数据洞察等，开展数据资产运营和应用的治理，应该包括：①评估数据资产的运营和应用能力，支撑数据价值转化和实现；②指导数据价值体系治理方案的实施，满足数据资产的运营和应用要求；③监督数据价值实现的绩效和符合性，并持续改进和优化。

(4) 数据治理过程。其包含：统筹和规划，构建和运行，监控和评价，改进和优化，是数据治理实施的方法。

统筹和规划。明确数据治理目标和任务，营造必要的治理环境，做好数据治理实施的准备，包括：①评估数据治理的资源、环境和人员能力等现状，分析与法律法规、行业监管、业务发展以及利益相关方需求等方面的差距，为数据治理方案的制订提供依据；②指导数据治理方案的制订，包括组织机构和责权利的规划、治理范围和任务的明确以及实施策略和流程的设计；③监督数据治理的统筹和规划过程，保证现状评估的客观性、组织机构设计的合理性以及数据治理方案的可行性。

构建和运行。构建数据治理实施的机制和路径，确保数据治理实施的有序运行，包括：①评估数据治理方案与现有资源、环境和能力的匹配程度，为数据治理实施提供指导；②制订数据治理实施的方案，包括组织机构和团队的构建，责权利的划分、实施路线图的制定、实施方法的选择以及管理制度的建立和运行等；③监督数据治理的构建和运行过程，保证数据治理实施过程与方案的符合性、治理资源的可用性和治理活动的可持续性。

监控和评价。监控数据治理的过程，评价数据治理的绩效、风险与合规，保障数据治理目标的实现，包括：①构建必要的绩效评估体系、内控体系或审计体系，制定评价机制、流程和制度；②评估数据治理成效与目标的符合性，必要时可聘请外部机构进行评估，为数据治理方案的改进和优化提供参考；③定期评价数据治理实施的有效性、合规性，确保数据及其

应用符合法律法规和行业监管要求。

改进和优化。改进数据治理方案,优化数据治理实施策略、方法和流程,促进数据治理体系的完善,包括:①持续评估数据治理相关的资源、环境、能力、实施和绩效等,支撑数据治理体系的建设;②指导数据治理方案的改进,优化数据治理的实施策略、方法、流程和制度,促进数据治理体系和数据价值体系的完善;③监督数据治理的改进和优化过程,为数据资源的管理和数据价值的实现提供保障。

2. 数据管理能力成熟度评估模型

数据管理能力成熟度评估模型(Data Management Capability Maturity Assessment Model,DCMM)由全国信息安全标准化技术委员会大数据标准工作组(工业和信息化和软件服务业司主导,多家企业和研究机构共同组成)研发,并于2018年3月15日正式发布,是我国数据管理领域最佳实践的总结和提升。

DCMM按照组织、制度、流程、技术对数据管理能力进行了分析、总结,提炼出组织数据管理的八大过程域,即数据战略、数据治理、数据架构、数据应用、数据安全、数据质量、数据标准、数据生存周期(图1.6),共包含28个过程项、441项评价指标。



图 1.6 数据管理能力成熟度评估模型

数据战略:数据战略规划、数据战略实施、数据战略评估。

数据治理:数据治理组织、数据制度建设、数据治理沟通。

数据架构:数据模型、数据分布、数据集成与共享、元数据管理。

数据应用:数据分析、数据开放共享、数据服务。

数据安全:数据安全策略、数据安全治理、数据安全审计。

数据质量:数据质量需求、数据质量检查、数据质量分析、数据质量提升。

数据标准:业务数据、参考数据和主数据、数据

元、指标数据。

数据生存周期:数据需求、数据设计和开发、数据运维、数据退役。

与CMMI(能力成熟度模型集成)类似,DCMM将组织的数据能力成熟度划分为初始级、受管理级、稳健级、量化管理级和优化级五个发展等级,帮助组织进行数据管理能力成熟度的评价,如图1.7所示。

(1)初始级。组织没有意识到数据的重要性,存在大量的“数据孤岛”,经常由于数据的问题出现低下的客户服务质量、繁重的人工维护工作等,具体的表现如下:当用户不相信数据的时候,业务管理者和IT管理者不知道问题的根源在于数据;组织在制定战略决策的时候,没有获得充分的数据支持;没有正式的数据蓝图规划、数据架构设计、数据管理组织和流程等;业务系统独自管理数据,各个业务系统之间的数据存在不一致或者冲突的现象。没有人意识到数据管理或者数据质量的重要性;数据的管理是根据项目实施的周期来进行的,没有人知道针对数据的维护、管理的成本到底是多少。

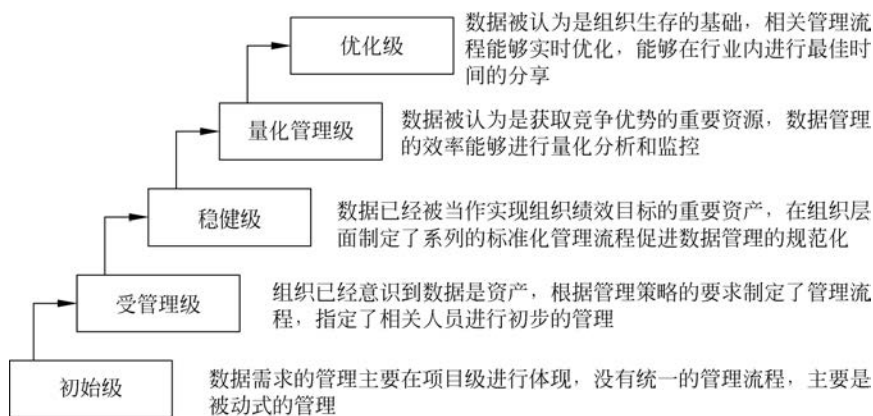


图 1.7 DCMM 数据管理能力成熟度等级

(2) 受管理级。组织已经识别了数据管理、应用相关的干系人，具体的特征如下：管理者已经意识到数据的重要性，已经制定、设立了一些数据管理的规范和岗位，想要促进数据管理相关工作的规范化；已经意识到数据质量和数据的“孤岛”问题是一个重要的管理问题，在进行数据分析的过程中，发现大量的数据不一致和重复的问题，但是找不到问题的根源或者为此要负责的人；组织进行了一些数据集成的工作，尝试整合分散于各个业务系统的数据，也设计了一些数据模型和管理的岗位；开始进行一些重要数据的文档工作，对重要数据的安全、风险等方面进行一些考虑，并且设计相关的管理措施。

(3) 稳健级。数据的管理者可以快速地满足跨多个业务系统、准确、一致的数据要求，有详细的数据需求响应处理规范、流程。其具体的标志如下：管理者已经意识到数据的价值，在组织的层面明确了数据管理的规范和制度；数据的管理以及应用能够充分地参考组织的业务战略、经营管理需求以及外部监管需求；建立了规范的管理组织、管理流程，能够推动组织内各部门/子公司来按照流程开展工作；组织在日常的决策、业务开展过程中能够获取充足的数据支持，显著提升了工作效率；能够定期开展数据管理、应用相关的培训工作。

(4) 量化管理级。组织认识到数据在流程优化、工作效率提升等方面的作用，针对数据管理方面的流程进行全面的优化，针对数据管理的岗位进行 KPI(关键绩效指标)的考核，规范和加强数据相关的管理工作，并且应用相关的业务对 KPI 考虑的工作进行支撑，具体的标志如下：管理者已经认识到数据是组织的战略资产，已经了解数据在流程优化、绩效提升等方面的作用，在制定组织业务战略的时候可以获得相关数据的支持；在组织层面建立了可量化的评价指标体系，可以准确测量数据管理流程的效率，并且可以及时进行流程优化；在数据管理、应用过程中充分借鉴了行业最佳实践，国家标准、行业标准等外部资源，促进组织本身的数据管理体系、应用体系的优化。

(5) 优化级。其具体的标志如下：整个组织可以把数据作为组织的核心竞争力，可以利用数据创造更多的价值和提升组织的效率；能够参与国家、行业等方面相关标准的制定工作；能够把组织自身数据能力建设的经验作为行业的最佳实践进行推广，成为行业的标杆。

1.3 大数据治理目标与难点

当前,随着信息化的进一步推进,特别是人工智能、大数据等新一代信息技术的广泛应用,数字化和智能化进程进一步加快,大数据治理正在嵌入政府部门、企业等各个组织并推动经济社会变革。不同类型的组织有不同的大数据治理目标,数据治理的难题也不尽相同。下面以政府与事业单位和企业为例,介绍大数据治理的主要目标和面临的难点。

1.3.1 政府与事业单位大数据治理的目标与难点

1. 政府大数据治理目标与难点

党的二十大报告指出:“完善社会治理体系。健全共建共治共享的社会治理制度,提升社会治理效能。”科学、精准的社会治理离不开大数据的有力支撑。正如“大数据之父”牛津大学教授维克托·迈尔-舍恩伯格(Viktor Mayer-Schönberger)所言:“大数据对社会的好处将是无穷的,它在一定程度上将解决迫在眉睫的问题,如气候变化、疾病,以及促进多元共治和经济发展。”

大数据为政府治理工作带来了更多、更有力方法和手段,如公安机关对关键办案信息及犯罪嫌疑人精确肖像的获取、高校人事档案管理等。2022年6月《国务院关于加强数字政府建设的指导意见》的出台,标志着我国政府进一步加大在数据治理方面的投入,将有利于助推中国智慧政府建设、提升政府治理效率和治理能力、提高政府公共服务水平,同时也有利于科学决策和利用信息资源进行社会管理。当前,政府数据治理的主要目标包括以下几个方面。

(1) 助力提升政府的决策能力。运用网络技术、大数据更加全面地获取决策所需信息,增强决策的科学性和精准性,提高决策质量。

(2) 优化政府公共服务,助力服务型政府建设。党的十九大提出,“社会主要矛盾已经转化为人民日益增长的美好生活需要和不平衡不充分的发展之间的矛盾”。服务型政府强调以人民为中心,数字政府以用户服务为基点,这体现了两者的一致性。通过数字赋能的政府公共服务治理,以期更好地从公民需求角度出发,实现公共服务精准供给与对接,提供高质量、便捷、多样化、多层次的公共服务。

(3) 助力政府更好地履行各项职能。数字政府建设驱动着政府职能转变特别是履行职能方式创新,推动简政放权、全流程优化与重塑、全过程监管、全方位服务的“放管服”改革向纵深发展。除提供公共服务外,市场监管、社会治理与环境保护等领域的职责履行,都在大数据及其技术中得以赋能增效。

(4) 推进数字政府建设。政府治理的变革,需要大数据技术基础,需要技术与制度的双轮驱动。大数据时代的数字政府建设应引导政府全方位现代化转型,更加需要大数据治理的支撑。数字政府的升级发展,应该从为政府治理赋能的工具与手段的定位,转向从深层次推进政府治理体系的变革。数字政府治理效能的发挥,因受到治理结构、激励制度等方面约束,与技术的快速发展之间还存在着矛盾与差距。为此,数字政府要从运用技术为政府治理赋能增能,发展到推动政府治理体制和治理体系变革的层次。制度影响着技术的作用,也影响着数字和大数据技术作用发挥的方向与程度。数字政府的发展,必须驱动政府治理体制

及制度性配套改革。

数字政府建设升级要促进政府职责体系优化。党的十九届四中全会提出,“坚持和完善中国特色社会主义行政体制,构建职责明确、依法行政的政府治理体系”。“十四五”规划提出的主要发展目标之一是,“国家行政体系更加完善,政府作用更好发挥,行政效率和公信力显著提升”。这对数字政府建设提出了更高的要求。数字政府发展最为关键的挑战并不是大数据技术本身,而是改革政府与其内部和外部多元主体之间的权力与责任、权利与义务关系,构建起符合现代发展目标的政府治理体系。

政府层级体系的进一步优化。党的十九届四中全会提出,赋予省级以下政府更多的自主权,“按照权责一致原则,规范垂直管理体制和地方分级管理体制”。大数据技术发展使时空高度压缩,数字政府建设提供的技术支撑平台,在驱动政府横向部门之间数据共享与协同共治的同时,要以技术的强大力量推动政府内部纵向体系的协同,减少政府层级。政府层级体系改革与行政区划改革密切关联。

政府大数据治理的难点在于内部管理、外部环境和数据本身三个方面。在内部管理方面,需要转变思维理念、提升数理能力等,面临着巨大挑战。在外部环境方面,政府需要解决如何打造一个开放、公平、可信任的数据环境的问题。在数据本身方面,需要解决数据质量、安全、权限等难题。这些挑战对于政府大数据治理的成功具有重要影响。

(1) 政府内部数据运用存在思维和能力上的问题。数据思维是数据治理战略定位的核心。政府数据的资产价值已得到高度认可,但在治理实践中,数据思维的运用仍存在不足。虽然所有政府都需要数据,但数据共享会使组织失去部分数据所有权,从而可能削弱组织的竞争力。因此,部分组织会担心共享数据被曲解、破坏机密性或增加自己的工作量。政府在数据治理中必须克服这些组织的阻力。总的来说,这些问题涉及组织间的利益冲突。为了消除部门数据私有化的陋习,必须改变过分强调数据收藏性和专业性的数据管理思维,开展具有前瞻性和系统性的数据管理,强调数据的连续性和综合集成,以此为基础进行数据整合、开放共享和开发利用等工作。同时,一些决策者过分追求技术至上,认为只要通过数据技术手段就可以实现社会公共服务的精准性。然而,在实际工作中,大部分的数据问题通常不是由现有的业务规则或技术本身引起的,而是数据治理不足导致的。数据治理不仅仅是技术问题,过分强调技术反而可能导致获取数据方面的不平等,造成额外的障碍,如知识可及性和“数据鸿沟”等。

此外,虽然公共部门将收集和存储大量数据视为监管职能的一部分,但并不意味着它们有能力处理这些数据。政府部门并不缺乏数据,政府数据治理的关键在于“人”。几乎所有部门都反映数据治理人才不足的问题。因此,当前,我国政府数据治理急需一支高质量、高水平的数据人才队伍。此外,政府部门还存在数据角色设置混乱的问题。要实现政府数据治理的成功,需要领导人和技术专业之间的密切协作。然而,在大数据时代和“开放政府数据”运动的推动下,政府部门在收集和处理政府数据方面存在人员不足、协同不足等问题。这些问题在一定程度上限制了政府的数据治理能力。

(2) 开放、公平、可信任的数据环境缺乏。政府与社会不同主体参与数据开发利用的广度、规模和深度,实际上反映了政府与社会不同主体之间数据利益的平衡。为增强各主体参与政府数据治理活动的积极性,政府部门需要制定和实施有效的规划和管理,利用政策手段鼓励和约束各利益相关者的增值利用行为。另外,也需要建立一个开放、公平的数据环境,

以满足多元治理主体的需求和表达他们的数据诉求与治理能力。然而,大多数国家的政府缺乏全面的愿景和战略方针来构建与维护良好的政府数据治理环境,即如何以可持续、包容和信任的方式利用公共部门的数据来改善公民福祉。因此,建立社会信任的数据治理环境的关键是解决数据安全与保护的问题。

(3) 政府数据治理面临数据质量、真实性、安全和权限等问题。政府数据来源多元且数据量巨大,但由于缺乏规范的数据管理标准,政府在数据治理过程中经常出现数据割据、数据冗余等问题影响数据使用的现象。政府数据开放是政府数据得以开发、利用的基础,但如果没有准确、可靠和真实的数据或信息,就不可能有一个成功和可靠的开放政府。此外,隐私安全与数据权限问题也是一个重要挑战。人们在日常生活和使用电子设备的过程中创造了大量的数据,数据平台掌握了大量的个人隐私数据,政府在进行相关公共决策时也会涉及公民大量的个人隐私数据。因此,如何保证政府以及相关平台不泄露、不滥用个人隐私数据是政府数据治理的一个重要难题。治理主体应该从根本上思考什么是信息隐私以及如何有效地治理这一问题。

(4) 数据所有权问题一直是一个备受争议的话题。许多人认为,个人数据的所有权应完全归个人所有,但也有人认为这种思考方式有缺陷,并不能解决现有的问题,反而可能引起新的问题。他们认为,单独的个人数据并不是很有用,只有融合大量的个人数据进行分析,才能发挥数据的最大效用。面对这样的争论,一些学者提出建立一个框架,让人们拥有规定使用个人数据的权利,而不要求他们拥有数据的所有权。美国的数据保护法案已朝这个方向迈出了重要的一步。此外,政府部门合规和风险管理流程尚未正式化等问题,也将增加数据泄露和敏感信息泄露的风险。政府需要同时保证规范采集和使用数据,并有效地保护数据安全和隐私,这是政府数据治理创新的一个难点。

2. 事业单位大数据治理目标与难点

事业单位在日常运营中产生了大量数据,包括人员管理、物资采购、财务管理等方面的数据。事业单位大数据治理的目标是通过日常运营中产生的大量数据进行有效采集、存储、分析和共享,提高工作效率、降低运营成本、提升服务质量等,以带来诸多好处。为了实现这一目标,需要建立完善的数据管理机构和管理流程,明确数据治理责任人,并对数据的安全保障、共享和开放、标准化、分析和挖掘等方面进行全面规划和实施。同时,需要加强对数据分析和挖掘的能力培养,建立数据分析团队和技术平台,实现对数据的深度分析和挖掘,为决策提供更准确和科学的依据。

事业单位大数据治理的难点在于面临多个方面的挑战。首先,事业单位内部的数据多样性和分散性导致数据的采集和整合存在困难;其次,由于数据涉及个人隐私等方面的保护,数据安全和隐私保护是大数据治理的一大难点;最后,由于数据分析需要高端技术和专业知识,缺乏专业人才也是事业单位大数据治理的一大瓶颈。此外,由于机构之间独立性和管理体制上的差异,跨机构数据共享和协同也是难以克服的挑战。因此,需要在政策和制度上加强对大数据治理的引导和规范,建立统一的数据标准和管理机制,同时加强对数据分析和挖掘的人才培养与技术支持,从而推进事业单位大数据治理的顺利实施。

学校存在大量的数据源,如学生的个人信息、学业成绩、教学活动记录、校园设施管理等。对这些数据的收集、存储、管理、分析和挖掘,可以实现学校数字化、智能化和信息化的

转型,提高教学质量和管理效率,优化资源配置,进而提升学校竞争力。当前,学校大数据治理主要有以下目标。

(1) 教学质量管理。通过对学生学业成绩等数据进行分析,了解学生学习情况,优化教学计划,改进教学方法,提高教学质量。

(2) 资源优化配置。通过对校园设施使用情况等数据进行分析,了解资源利用率,优化资源配置,节省资金成本,提高资源利用效率。

(3) 学校安全管理。通过对校园安全监控设备等数据进行分析,了解安全状况,预测安全风险,及时处理安全事件,保障校园安全。

然而,在学校实现大数据治理过程中,也存在一些难点和挑战。

(1) 数据来源和采集。学校存在多种数据源,如学生信息管理系统、学生成绩管理系统、图书馆管理系统等,如何收集这些数据,整合并建立学校大数据平台,实现数据的一体化管理是一项难点。

(2) 数据质量管理。学校的数据存在数据质量问题,如数据的准确性、完整性、一致性和时效性等,如何进行数据质量管理,保障数据的可靠性和准确性是一项难点。

(3) 数据共享和协同。学校的各个部门之间存在“信息孤岛”,数据难以共享和协同,如何实现数据共享和协同,提高数据的利用效率,避免“数据孤岛”造成资源浪费是一项难点。

(4) 数据安全和隐私保护。学校的数据存在安全问题,如数据泄露、损毁或篡改等,如何保护学校数据的安全和隐私,避免数据安全问题导致信息泄露或其他问题是一项难点。

(5) 数据分析和挖掘。学校的数据量庞大,如何通过数据分析和挖掘,提取有价值的信息,为学校提供决策支持,推动学校数字化、智能化和信息化的转型是一项难点。

1.3.2 企业大数据治理的目标与难点

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,数据基础制度建设事关国家发展和安全大局,要维护国家数据安全,保护个人信息和商业秘密,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。数据基础制度体系体现在国家、行业、企业、个人等不同层面,需要政府和企业协同作用,提高数据治理效能。当前,企业大数据治理主要有如下目标。

(1) 推进高效的管理决策和业务协同。一致性的大数据环境让系统应用集成、数据清理变得更加自动化,可以有效降低过程中的人工成本。同时,高质量的数据服务允许企业员工方便、及时地查询到所需的数据,无须在部门与部门之间进行协调、汇报等,有效提升了业务处理效率。大数据治理是有效管理企业数据的重要举措,是实现数字化转型的必经之路,对推进高效的管理决策和业务协同具有重要意义。

(2) 助力企业创新与转型。基于数据实现企业管理的升级和业务的创新,通过数据拓展新业务、构建新业态、探索新模式是企业的目标。对于传统制造企业,利用数据治理,可以加速管理创新、产品创新、销售模式创新,例如:利用数据治理加强集团管控、基于数据实现供应链协同和优化、基于市场预测实现创新产品设计 with 快速上市等。对于服务行业,利用大数据探索服务的新模式,可以拓宽服务的视野,实现模式领域的横向拓展、服务精度的纵向延伸。

(3) 增强数据安全及控制数据风险。大数据治理有利于建立基于知识图谱的数据分析

服务,例如 360°客户画像、全息数据地图、企业关系图谱等,帮助企业实现供应链、投融资的风险控制,更好地管理公共领域的风险。有效的大数据治理可以更好地保证数据的安全防护、敏感数据保护和数据的合规使用。通过数据梳理识别敏感数据,再通过实施相应的数据安全处理技术,例如数据加密/解密、数据脱敏/脱密、数据安全传输、数据访问控制、数据分级授权等手段,实现数据的安全防护和使用合规。

许多企业在迈向数字化和推进数字化转型过程中,面临一系列大数据治理难题,包括以下几个方面。

(1) 对大数据治理的业务价值认识不足。传统数据治理是以技术为导向的,注重底层数据的标准化和操作过程的规范化。尽管以技术驱动的数据治理能够显示数据的缺陷,提升数据的质量,但由于传统以技术驱动的数据治理模式没有从解决业务的实际问题出发,企业对数据治理的业务价值普遍认识不足。为快速实现数据价值和成效,最直接的方式就是以业务价值为导向,从企业实际面临的数据应用需求和数据痛点需求出发,满足管理层和业务人员的数据需求,以实现数据的业务价值、解决具体的数据痛点和难点为驱动来推动治理工作。企业数据治理的业务价值主要体现在降低成本、提升效率、提高质量、控制风险、增强安全和赋能决策上。企业应从管理层和业务部门的痛点需求出发,将数据治理的业务价值量化,以增强管理层和业务人员对数据治理的认知与信心。数据治理必须着重于业务需求,并着重于解决让业务人员感到痛苦或他们无法解决的问题。

(2) 缺乏企业级大数据治理顶层设计。当前企业普遍都认识到数据的重要性,很多企业也开始探索数据治理。然而目前企业大量的数据治理活动都是项目级、部门级的,缺乏企业级数据治理的顶层设计以及数据治理工作和资源的统筹协调。大数据治理涉及业务的梳理、标准的制定、业务流程的优化,数据的监控、数据的集成和融合等工作,复杂度高,探索性强,如果缺乏顶层设计的指导,则在治理过程中出现偏离或失误的概率较大,而且出现偏离或失误又不能及时纠正,其不良影响将难以估计。大数据治理的顶层设计属于战略层面的策略,它关注全局性和体系性。在全局性方面,从全局视角进行设计,突破单一项目型治理的局限,促进企业主价值链各业务环节的协同,自上而下统筹规划,以点带面实施推进。在体系性方面,从组织部门、岗位设置(用户权限)、流程优化、管理方法、技术工具等入手,构建企业数据治理的组织体系、管理体系和技术体系。企业数据治理的顶层设计应站在企业战略的高度,以全局视角对所涉及的各方面、各层次、各要素进行统筹考虑,协调各种资源和关系,确定数据治理目标,并为其制定正确的策略和路径。

(3) 高层领导对大数据治理重视度不够。大数据治理是企业战略层的策略,而企业高层领导是战略制定的直接参与者,也是战略落实的执行者。数据治理的成功实施不是一个人或一个部门就能完成的,需要企业各级领导、各业务部门核心人员、信息技术骨干的共同关注和通力合作,其中企业高层领导无疑是数据治理项目实施的核心干系人。企业高层领导对大数据治理的支持不仅在于财务资金方面,其能否对数据战略的细化和实施充分授权、所能提供的资源是决定数据治理成败的关键因素。为了保证数据治理的成功实施,企业需要成立专门的组织机构,例如数据治理委员会。

(4) 数据标准不统一和数据整合困难。一方面,企业内部的数据标准不统一。随着大数据的发展,企业数据呈现出多样化、多源化的发展趋势,企业必须将不同来源、不同形式的数据集成与整合到一起,才能合理有效地利用数据,充分发挥数据的价值。然而由于缺乏统

一的数据标准定义,数据集成、融合难度较大。另一方面,企业之间的数据标准不统一。各行业、各企业都倾向于依照自己的标准采集、存储和处理数据,这虽然在一定程度上起到了保护商业秘密的作用,但阻碍了企业(如位于同一产业链上的上下游企业)之间的协同发展。

(5) 大数据治理的职能部门不清晰。多数企业的业务人员普遍认为数据治理属于 IT 部门的工作,因而对数据治理持“事不关己,高高挂起”的态度。尽管 IT 部门的确对数据负有很大责任,但不包括数据的定义、输入和使用。数据的定义、业务规则、数据输入及控制、数据的使用都是业务人员的职责,而这些恰恰是数据治理的关键。因此,在数字化时代,IT 和业务更应当紧密融合在一起,让业务人员与 IT 人员一起定义数据标准、规范数据质量及合理使用数据,朝着共同的目标努力。有效的数据治理策略是实现数据驱动业务、业务融入 IT 的重要举措,这些举措包括数据治理的规划应与业务需求相匹配,数据治理的目标应围绕业务目标的实现而展开。

1.4 大数据治理内容体系

大数据治理内容体系主要包括大数据架构管理、元数据管理、主数据管理、大数据集成、大数据质量管理、大数据标准化、大数据资产化、大数据安全管理、大数据治理应用等。下面提供部分主要章节概要,更全面和详细的内容将在后续章节中进一步介绍。

1.4.1 大数据架构管理

大数据治理实践表明,仅仅依靠技术部门来推动和开展数据治理工作是无法取得项目立项时所期望的成果的,其原因在于数据治理涉及企业各个部门的业务和资源,只有来自更高管理者的驱动力,才能保证企业内部的高效协作。数据治理是一项需要企业通力协作的工作,而有效的组织架构是企业数据治理成功的有力保障。为达到数据战略目标,企业有必要建立体系化的组织架构,明确职责分工。

大数据架构便于企业运营决策者以数据视角分析整个企业的数据分布与业务域之间的关系,是企业了解自身价值和制定战略决策的最重要依据。大数据架构是从跨业务、跨层级、跨应用系统的视角出发,统一组织和规划数据,提高数据集中存储和跨系统数据共享的效率。大数据架构的设计,可以满足企业信息化各层级的要求:首先,经营管理更加集中,实现业务级别的集约化管理;其次,业务更加融合,按照业务主线深度集成所有业务流程,实现企业整体资源共享与业务协作;最后,决策更加智能,实现经营决策的智能分析、管理控制的智能处理、业务操作的智能作业,进一步深化数据分析利用能力,提高管理决策支持能力。

1.4.2 元数据管理

元数据管理作为大数据治理的核心基础设施,是有效管理这些海量数据的基础和前提,在数字化转型中发挥着重要作用,并且得到企业的重点关注。元数据是数据的说明书,有了完善的元数据,可通过可视化的方式展现数据上下游关系图,快速定位问题字段,可以帮助企业降低数据问题定位的难度。同时,以元数据为核心的大数据治理也被企业广泛认可和

实施部署。数字化时代,企业需要知道它们拥有什么数据、数据在哪里、由谁负责,数据中的值意味着什么,数据的生命周期是什么,哪些数据安全和隐私需要保护,以及谁使用了数据、用于什么业务目的、数据的质量怎么样等,这些问题都需要通过元数据管理解决。

1.4.3 主数据管理

主数据是满足跨部门业务协同需要、反映核心业务实体状态属性的基础信息,是企业的核心数据资产。管好、用好主数据是实施企业数据治理的重要内容。构建完整的主数据管理体系可以对主数据实施统一、规范、高效的管理,确保分散的系统间数据的一致性,改进数据合规性。不仅如此,主数据还是数据标准落地的关键载体,是企业实施全面数据治理的核心基础,成功实施主数据管理可以很好地推动企业全面建设数据治理体系。主数据是企业的核心数据资产,在一定程度上,主数据质量的好坏决定了数据价值的高低。主数据不仅是实现各部门之间、各信息系统之间、各企业之间数据互联互通的基础,也是数据分析、数据挖掘的基础,这个基础如果打得不牢,企业的数字化转型将举步维艰。主数据被誉为企业数据中的“黄金数据”,是业务应用、数据分析、系统集成的基础。主数据管理是企业数据治理工作中的核心内容,构建完善的主数据管理体系,统一企业核心主数据是支撑企业数字化转型的基石。

1.4.4 大数据集成

随着 IT 的不断发展,数据集成技术也在不断演进。从数据库之间的点对点(peer-to-peer,P2P)集成到基于企业服务总线(enterprise service bus,ESB)的总线型集成模式,从主数据的中央式数据集成到微服务架构分布式的集成模式,从数据仓库的结构化数据集成到数据湖的融合各类数据源的数据集成模式,每一次集成架构的变革,每一次集成技术的升级,其背后的驱动力都是越来越高的数据应用需求和越来越复杂的数据集成环境。大数据集成的核心任务是将相互关联的分布式异构数据源集成到一起,使用户以透明的方式访问这些数据源。数据集成是对企业内部各系统之间、各部门之间,以及企业与企业之间数据移动过程的有效管理。通过数据集成来横贯整个企业的异构系统、应用、数据源等,实现内部企业资源规划(ERP)、数据仓库等各异构系统的应用协同和数据共享。通过数据集成来连通企业与企业之间的数据通道,实现跨企业的数据共享,发挥数据价值。

1.4.5 大数据质量管理

2020 年 4 月发布的《中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见》中提出“提升社会数据资源价值。培育数字经济新产业、新业态和新模式……探索建立统一规范的数据管理制度、提高数据质量和规范性”,说明建立统一规范的管理制度、提高大数据质量已成为数字经济发展的首要任务。大数据质量管理是针对大数据在运行生命周期中可能产生的数据质量问题,通过一系列处理和管理工作不断改善问题,从而提高大数据质量的管理水平。其主要目的是通过提高大数据的准确度和规范性,提高其价值,为企业事业单位、各类社会组织 and 政府带来更高的效益。实际上,大数据质量管理是一个综合性解决方案,需要结合各种技术手段、业务能力和管理方法。通过有效的控制手段,对大数据进行有

效的管理控制,解决大数据质量存在的问题。大数据质量管理贯穿数据生命周期的过程,覆盖大数据质量需求、大数据探查、大数据诊断、大数据质量评估、大数据监控、大数据清洗、大数据质量提升等方面。数据源不断增多,数据规模不断扩大,数据的复杂度也不断提升,这些都给大数据质量管理带来了前所未有的挑战。因此,基于大数据质量管理框架的大数据质量管理方法应包含事前计划(大数据质量需求)、事中检测(大数据质量检查和大数据质量分析)、事后评估及处理提升(大数据质量提升和大数据质量评估)四个阶段。

1.4.6 大数据标准化

大数据标准化可以提高数据的可携带性和互操作性,可能成为促进和改进数据使用的关键,有助于支持更具有竞争力和分布式的数据收集生态系统。事实上,对于将跨企业和跨行业的数据交换视作至关重要的行业而言,标准化是其运营的先决条件。同时标准化还可以创造实质性的效益。以医疗行业为例,当使用相似的指标记录患者对治疗的反应时,可以促进数据的整合,从而为临床护理、公共卫生和生物医学研究提供信息,为不同地点的人员提供更快、可互操作的病例查阅等,通过这些方式来改善医疗保健服务。大数据标准化可以让数据流动更顺畅、机器学习更优,在数据输入算法造成权利被侵害或不合理伤害发生时更容易得到监管。而对于政府而言,大数据标准规范是强化政务数据全生命周期管理、流动、应用、归档,实现融合汇聚、价值挖掘、高效协同、安全有序的重要保障。

1.4.7 大数据资产化

大数据资产化是指数据在市场中流通交易从而为使用者或所有者带来经济利益的过程。它本质上是数据交换价值的形成,而实现数据价值的核心在于资产化。从数据资产化的角度,大数据治理应更多地聚焦于数据的资产属性,要求政府以战略性方式管理数据,实现数据资产的保值、增值。因此,数据治理应从供给环节开始考虑数据质量、数据流通、数据增值加工、数据基础服务、数据安全保障等关键问题,并建立起覆盖数据采集、存储、流通、应用等全生命周期的数据优化结构。由于数据不同于通常意义上的有形实物和无形知识产权,因此大数据资产化仍然面临许多技术挑战,例如大数据资产的形态、计量和计价,数据资产流通的数据支撑等。在国家实施大数据战略的背景下,着手解决大数据资产化进程中的技术问题是推动大数据资产从概念走向实践的关键环节。

1.4.8 大数据安全管理

党的二十大报告强调,“推进国家安全体系和能力现代化”,“强化经济、重大基础设施、金融、网络、数据、生物、资源、核、太空、海洋等安全保障体系建设”。大数据安全管理不仅是大数据治理的重要内容,也是利用大数据赋能社会治理水平提升的重要基础。大数据安全管理贯穿数据采集、传输、存储、处理、交换、销毁等各个阶段。从数据安全管理的全局考虑,企业需要引进数据安全风险评估方法论和技术措施,制定数据应急保障流程和方法,以便在发生数据安全事件时很好地进行风险控制。传统网络安全问题解决的基本思想是“隔离”,即通过在每个边界设立网关设备和网络流量控制设备,守住边界。然而,随着移动互联网、

云服务的出现,网络边界实际上已经逐渐消失。数据安全的威胁正在进一步升级,在高级持续性威胁(advanced persistent threat,APT)、用户行为异常风险、网络漏洞等数据安全威胁下,传统的防御性、检测型的安全防护措施已经难以满足要求。因此,需要采用新的安全机制、措施和技术来满足新形势下大数据安全管理的要求。

1.4.9 大数据治理应用

大数据治理已经应用到社会治理的各个方面,如政府大数据治理、交通大数据治理、应急大数据治理、石油化工企业大数据治理、医疗大数据治理、养老大数据治理、药物大数据治理等。在交通拥堵和交通安全等问题普遍存在的情景下,交通大数据技术治理有效提升了城市交通系统效率,推进了城市交通绿色化、智能化发展;在应急治理备灾阶段,应急大数据治理技术可以智能、高效、精准地对危机事件进行预警监测,同时及时向政府和社会公众传递可靠的情报信息,摆脱原有应急治理模式信息收集和传递效率低下的困境;在石油化工企业运营过程中,大数据治理技术已经渗入运营的各个层面,如原油选购、排产、库存管理、设备检维修及运行;在医疗健康领域,大数据治理推进信息资源的整合、对接和共享,从而提升医疗机构信息化水平,充分发挥信息化作用;在人口老龄化问题日益严重的社会背景下,大数据治理有效促进智慧养老模式的发展;在医药行业,药物大数据技术可以有效地处理不断增长的药物数据,为整个医药行业的发展提供有效的决策支持。

本章小结

本章首先对大数据治理的内涵进行了阐述,突出了大数据治理对象和治理结构高复杂性的特点;在理解大数据治理的基础上,详细梳理了国内外大数据治理总体框架、各个组成部分以及治理标准;并以政府与事业单位和企业为例分别介绍了政府与事业单位大数据治理和企业大数据治理的主要目标以及面临的难点,简明扼要地介绍了大数据治理的内容体系。大数据技术正以前所未有的广度和深度渗透到政府与企事业单位的治理,运用大数据推动组织创新和数字治理已成为当下社会各界的共识。通过大数据治理和深度运用大数据内蕴含的信息与知识,可以充分实现大数据的价值,推动国家创新和经济社会高质量发展。

思考题

1. 数据治理与大数据治理在内涵上有何联系和区别?
2. 对数据管理能力成熟度评估模型进行简要描述。
3. 当前我国大数据治理面临的难点和挑战主要有哪些?
4. 如何理解大数据治理的业务价值?
5. 简要概括企业和政府大数据治理目标的异同。

即测即练

