# 背景篇

在大数据、人工智能技术浪潮的冲击下，各行各业对历史积累与实时产生的多源、多构、多模数据，正形成日益迫切的分析、挖掘、发现和共享需求。此时，数据安全的内涵和外延都在持续扩大与深化，它不再是传统的文件加密、物理封存、档案管理等方法来保证“不被窃取、不被篡改、不被泄露”的静态实体，也不是通过责任把关、审批调阅、有限复制的管理机制所能覆盖。其保护对象、保护场景、保护要求、管理制度机制、安全责任划分等，都发生了深刻变化。从本质上看，数据安全已经从简单的数据状态属性转化成为数据生存全过程的动态属性。

本篇从具体的数据应用与管理场景切入，力图阐明数据安全这个命题产生的原因、过程和结果，进而剖析数据和数据安全是密不可分的，只有建立在数据安全的基石上，才能保证各业务过程所产生、使用及交换的数据是可信、可控、可见、可知和可增值的。

本篇通过跨越数据安全学术语境，从数据全生命周期各参与方的角度，探索数据安全在人工智能时代的呈现方式、特有属性、合规难点以及与业务相关的安全特征，并深入探讨数据安全与数据要素的相互作用及数据安全形态，进而全面把握当今数据安全在宏观、中观和微观层面的方方面面。

# 第1章

# 数据安全面面观

数据始终依附于其自身的生命周期而存在，数据安全也因此贯穿数据产生、收集、处理、加工、使用、管理和监管的全过程。在数据全生命周期的每一个阶段、每一个节点，数据安全都有着不同的界定和合规准则。长久以来，对于数据安全，似乎存在着一种认识：它“分散”在数据全生命周期的每一个环节，数据安全的控制与管理是离散化的，体系化的数据安全生态似乎只存在于学术研究领域中。事实真是这样吗？本章将从不同的维度来揭示数据安全的完整与离散的技术特征，以及二者之间的辩证关系。

## 1.1 我们与数据安全

### 1.1.1 突然面对数据安全

我们是在参加一个行业大数据应用重点项目中，第一次被要求全面、深入地面对数据安全问题，并以体系化思路系统性地解决大数据安全保障问题。在这个行业大数据应用项目中，数据是海量的、异构的、多模态的；存储与计算是基于私有云的；处理后的数据有极高价值。数据使用方式是多种多样的，业务访问环境包括桌面终端、移动终端和各类感知终端等多种类型；行业对零信任、数据分类分级、数据访问控制、敏感数据安全监测、数据安全运维以及敏感数据保护与合规使用等要求是明确且高标准的。这些要求对各参与方在信息安全、网络安全、应用安全、数据安全保护、风险管理、零信任的认识形成全面冲击。大家深感数据安全责任重大，积累的安全知识不够用了，已有的安全解决方案“失灵”了，一个个新问题摆在面前，大家都感到一片茫然、无所适从。

从行业发展角度看，需要严格遵守《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《中华人民共和国密码法》，以及《网络安全等级保护条例》《关键信息基础设施安全保护条例》《商用密码管理条例》《网络数据安全管理条例（征求意见稿）》等法律法规。同时，为支撑快速发展的业务数字化、智能化转型变革，行业还制定了一系列的制度、标准规范，提出了新的安全目标和要求，以指导和约束信息安全、网络

安全、数据安全的建设和发展。主要包括以下几个主要方面。

（1）安全要适应新型基础设施环境。原有的网络架构、基于 IOE 的技术架构正在向云计算、大数据、人工智能等方向发展，数据安全成为安全核心。在此背景下，需要解决多网络、多数据中心、多云环境下的整体安全问题，安全性已成为保障业务连续、稳定运行的内在需求。

（2）安全要服务新型数据应用方式。过去数据多采用分散采集及管理模式，或以多份拷贝形式随处流转，缺乏统一管控。现在要求实现数据的统一规划和集中管理，落实数据分类分级，实施按需授权、严格审批的使用机制，确保数据应用过程可控、可溯。

（3）安全要满足新型数据合规要求。在原有网络安全等级保护制度基础上，行业进一步提出数据安全保护、个人隐私信息保护、数据分类分级管理、数据合规出境等新的合规要求，需同步纳入安全体系建设。

（4）安全要应对新型数据安全要求。在传统信息安全和网络安全的基础上，要根据数据分类分级实施差异化保护，涵盖数据处理过程安全及端到端业务应用安全等新场景。不同安全等级的业务、应用、数据要按需隔离；不同的网络和多种终端要协同支撑云边端一体化应用；不同安全等级的应用要采用对应的安全部署拓扑和访问控制策略，实现精细化授权与策略的严格落地。同时，安全状况要实时监测，安全措施要动态调整，安全策略要持续运营，形成闭环管理机制。

更关键的是，上述安全合规的要求落地，需要与数据治理、数据管理、应用建设发展及业务模式创新深度融合，要保证大数据治理、业务应用建设及基础设施建设管理等各参与方在安全认识和工程路径上达成一致，密切协同、相向而行。

面对数据安全与大数据安全的深刻挑战，我们开始了新的学习和思考，并积极参与行业新型安全体系的设计和实践探索。

### 1.1.2　触碰零信任

零信任是这个行业为数据安全引入的第一个“新理念”，也是要啃下的第一块“硬骨头”。回顾我们的信息技术工作经历，从应用开发、数据治理到大数据应用，一路走过来，当第一次听到“零信任”时觉得挺有道理，但疑惑也随之而来，现有的应用系统早已配备了用户身份认证、权限管理、访问控制和日志审计等机制，应用安全一直就是遵循这一路径建设的，为什么还要再提出“零信任”呢？带着这个疑惑，我们开始了“零信任”的系统学习和实践探索。

**1. 零信任要改变什么**

对零信任最简单的描述就是“永不信任，持续验证”，但其内涵丰富，与原有的信息安全、网络安全理念既有一定交叉又存在区别，需要结合实践深入体会。以下是我们在学习过程中对该理念核心内容的系统梳理。

（1）身份是零信任的基石。一切访问主体和资源必须统一管理，主体必须有统一平台发放和管理的数字身份。“访问”包括网络访问、系统访问、应用访问及数据访问。

（2）数字身份要具备可鉴别性，且身份鉴别可按需发起，确保身份真实性始终可控。

（3）身份的核心是可信。需要通过持续的身份鉴别、行为分析及身份状态监测，动态评估身份的“信任度”，以此作为判断身份可信程度的依据。

（4）访问权限也需要“可信”保障，要通过统一的策略管理确保权限信息可信。

（5）访问权限是动态策略，可根据业务任务、工作角色、主体条件与属性、资源条件与属性、环境条件与属性等，灵活定义授权策略。

（6）访问策略决策是动态的，与访问主体与资源的安全状态、行为特征、信任情况以及整体安全态势等密切相关。

（7）网络需要按需隔离。

（8）策略需要灵活编排和统一管理，策略决策引擎是核心能力。

（9）立足安全数据、安全知识的持续安全态势感知、行为分析、风险分析与评估、事件应对是保证零信任运行的“中枢”。

（10）实现零信任有几项关键技术，包括身份与访问管理（IAM）、软件定义边界（SDP）、微隔离技术、策略编排与自动化机制，以及智能分析与评估等。

**2. 从身份鉴别、权限管理与访问控制切入**

在上述认识的基础上，我们按照客户要求迈出了实践的第一步：首先在应用层统筹身份管理、身份鉴别、权限管理与访问控制。其主要目标包括以下几点。

（1）实现用户、应用、功能与服务接口的统一资源管理和身份管理，实现用户统一身份凭证的发放和核验。同时，结合主体访问利用令牌等技术发放动态用户及应用身份凭证，作为应用访问有效期内的身份凭证。

（2）提供统一的权限管理，基于角色的访问控制（RBAC）与基于属性的访问控制（ABAC），实现动态授权。

（3）根据业务分类分级的访问策略要求，对应用、功能、服务接口及数据访问进行动态访问控制，并在访问控制过程中增加对业务合规性的验证。

（4）通过建设策略决策引擎和访问策略执行点，规范访问控制发起与执行。

（5）强化终端安全感知与控制能力建设。

（6）完成应用层日志采集与行为分析，建立异常发现机制，支撑动态访问控制，并与终端安全控制能力结合，实现风险管控策略的下发和执行，提高终端风险管控能力。

**3. 零信任与应用融合**

身份鉴别、权限管理与访问控制建设完成后，新建的大数据应用及业务应用均按照整体安全要求，实现了与身份鉴别、权限管理与访问控制的对接，基本达到了设计效果。但当推动原有已建成业务应用与身份鉴别、权限管理与访问控制对接时，遇到了很大阻力，甚至引发诸多质疑与不满。

（1）零信任到底不信任什么：当向业务部门提出，已建应用要与符合零信任要求的身份鉴别、权限管理与访问控制对接时，业务部门的第一个疑问就是：每个应用都有身份认证、授权、访问控制、日志审计，为什么还要重新建设？为什么必须采用统一的身份认证、权限管理和访问控制？数据归属于本部门，内部管控足矣，为什么要搞得这么复杂？此类质疑层出不

穷，导致工作进展缓慢。

（2）当业务应用的业务逻辑与身份鉴别、访问控制逻辑融合后，业务过程被安全机制“打断”。原本由应用侧统一处理的业务和管控逻辑，因插入了安全的逻辑过程，导致原有的业务过程连续性中断、可见性失效。一旦预期的业务请求访问过程出现问题，诊断、排错成为难点，业务部门与安全部门之间的责任扯皮成为常态。

（3）用户普遍反映多因素认证、多次认证影响了业务体验，制造了麻烦，业务与安全的冲突“被加剧”。

**4. 很多问题依然存在**

当推动应用层身份鉴别与访问控制后，还没有松口气，就遇到了一次蠕虫病毒爆发，导致大规模业务中断；没过多久，运维过程中又发生了一次较严重的数据误删除事故。面对接连发生的安全事件，大家质疑：费了这么大劲搞的新安全方案，到底能不能发挥作用？为什么很多安全风险和问题还是无法有效防范与控制？面对质疑，团队只能继续沉下心总结并认真复盘，理清问题根源。

（1）应用层访问控制虽有效管控了业务访问过程的数据安全，基本遏制了业务应用滥用数据、爬取数据等情况，但运维服务并没有纳入零信任管控，导致系统运维过程中出现的数据泄露、数据误操作危害依然严重。

（2）数据处理过程、数据生命周期管理过程未纳入零信任管控，数据处理环节、数据管理过程中依然可能发生数据泄露和数据滥用。

（3）零信任停留在应用层，未能与网络安全、云计算平台安全、终端安全形成联动。仿佛当前管控对象仅限于普通业务用户，而特权用户和潜藏的攻击者仍游离于体系之外。

（4）虽然搜集了全部认证、授权、访问控制以及应用访问日志数据，但仍然无法“看清”数据流向、业务访问过程，无法构建完整的访问链与日志链。风险发现能力并没有实质性提升，风险响应与管控能力也难以同步优化。

（5）零信任作为先进理念，并不是包治百病的“灵丹妙药”，要真正从整体视角出发考虑数据安全。

### 1.1.3　思考 AI 时代的数据安全

经历了零信任的实践探索与复盘，我们对客户提出的数据安全要求有了更为深切的“敬畏”之心，除了继续认真研究零信任理念，也开始了对数据安全的整体研究，逐步形成对数据安全体系的全新认识，如图 1-1-1 所示。

**1. 数据安全是数智化发展的生命线**

AI 时代是数字化、智能化技术与数据的深度融合的时代，其本质在于数据要素在全业务场景、全协同链条的流动与价值转化过程。数据安全是这一过程的“基础设施”，直接决定了数智化发展的稳定性与可持续性。因此，数据安全的核心价值体现在对发展基础的保障与价值释放的护航两个维度。

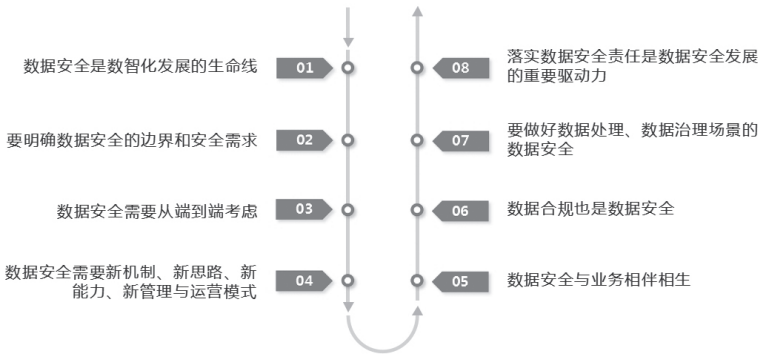


图 1-1-1　AI 时代的数据安全思考

1）保障数智化基础设施的安全

数字化构建了以云计算、大数据中心、工业互联网为核心的技术底座，智能化则通过 AI 大模型、物联网终端等技术实现底座能力的升级与延伸。这些设施的正常运转完全依赖于数据的安全流转。

2）保障数据要素价值的安全合规释放

数智化发展的核心目标是实现数据要素的最优化配置与价值最大化，而这一过程必须以安全为前提。安全保障机制缺失将导致数据孤岛与流转风险的双重困境，一方面，因担忧数据泄露不敢参与数据流转；另一方面，无序流转又可能引发行业或商业秘密外泄、个人信息滥用等问题。可以说，没有数据安全，数据要素的价值释放便无从谈起，数智化发展也将失去核心动能。

**2. 明确数据安全的边界和安全需求**

在法律法规、制度规范以及行业要求的双重约束下，需构建覆盖全部业务数据的全生命周期安全防护体系。应围绕数据分类分级与零信任理念，结合行业技术环境，构建以数据安全为中心的纵深防御体系。

**3. 数据安全需要从端到端考虑**

数据安全要覆盖数据流转的每个环节，贯穿数据从采集到销毁的全生命周期，绝非仅关注存储、使用或交换等单一环节。“端到端”就是一个比较准确的表达，既包括各种业务场景下从数据产生端到数据销毁端，也包括从业务请求端到数据服务端的完整链路。

**4. 数据安全需要新机制、新思路、新能力、新管理与运营模式**

（1）数据安全不是单点问题，而是系统性要求，需要体系化解决方案。原有针对网络实体对象的“各个击破”式安全保护模式不足以支撑当前要求，需要引入新的思路、构建新的能力。

（2）数据安全需要基础设施、应用、数据、安全等多方协同推进，需要重新划分安全角色与安全职责，需要构建安全责任链，保证数据安全能够结合保护场景逐环节落实并进行责任溯源，因此需要新的协同工作机制。

（3）供应链安全是数据安全的重要环节，需要对人员、产品、技术服务提供全面管理和持续检查。

（4）数据安全有其自身的运作体系，既包括按照“计划—执行—检查—处理”PDCA 循环运行的资产管理、身份管理、策略管理等业务内容，也包括按照“观察—判断—决策—行动”OODA 循环运行的监测与预警、风险评估与响应、事件管理与应对等业务内容。通过严格的闭环运行管理和持续优化，保证安全防护措施的有效运转与迭代，保证对已知风险的响应与处置及对未知风险的感知与应对，因此需要建立新的管理与运营要求和模式。

**5. 数据安全与业务相伴相生**

正在使用的数据或流动的数据是数据安全的重点保护对象，因此数据安全需要通过业务落实，并通过业务的持续、稳定、可靠运行来体现其成效和贡献。在数据安全实践工作中，我们提出了一个观点：管不好应用，就管不住数据。因此，强化业务应用全生命周期及业务访问全过程安全，是数据安全的重要抓手；推动数据安全与业务应用融合落地，是数据安全工程实施的关键内容。

数据安全与业务相伴相生，需要构建安全与业务同步推进的有效机制，需要既懂数据又懂安全的复合型人才。

**6. 数据合规也是数据安全**

从政策视角看，数据合规是在原有网络安全合规（如网络安全等级保护）的基础上，增加了隐私保护、数据出境等法律框架下的合规性要求。同时，不同的行业根据其业务特点及数据安全目标，提出了诸如“谁采集、谁管理、谁负责”“谁使用、谁负责”等规范化管理要求。从技术视角看，数据合规是数据安全的重要内容，也是数据安全的基本要求。

**7. 做好数据处理、数据治理场景下的数据安全**

数据加工与处理、数据治理与管理场景既是数据安全的重要阵地，也是数据安全落地的一个难点。在实践中，我们将上述场景的数据安全细化为以下几个方面。

（1）使数据可知：通过安全可控的手段使用数据资源目录或数据管理元数据；具备发现、识别新数据、新数据关系的能力。

（2）保证数据可信：保证数据来源、数据收集手段及渠道可信；保证数据加工与处理过程所采用的计算、存储资源和能力可信。

（3）使数据可用：结合数据分类分级和数据脱敏、加密使用策略，提供精细化数据访问授权和访问控制，保证数据按最小权限使用。

（4）数据可动态分类分级：在数据收集、存储、加工处理和使用过程中，按照数据分类分级规则与策略要求实现动态数据分类分级，并按照业务需要对敏感数据进行相应处理或保护。

（5）数据可按需脱敏：在上述各环节中，按照相应的数据脱敏及隐私信息处理策略要求，对数据进行脱敏处理。

（6）保证数据可溯源：严格规范记录数据收集、加工处理以及数据使用的日志记录，分析并形成血缘关系，实现过程流溯源与数据流溯源。

（7）保证数据合规：按照敏感信息（含隐私信息）合规要求，在数据收集、存储、使用、交换、公开、销毁等环节提供安全控制措施，确保操作合规；按照国家及行业对数据出境的合规性要求，严格管控数据出境行为。

（8）使数据被安全保护：按照数据分类分级保护要求，结合数据收集、加工处理、存储、使用等环节，对数据本身及其承载实体与承载环境实施差异化的安全保护措施，并通过安全运营保证安全保护措施和策略的匹配度。

数据处理与数据治理场景安全要求如图 1-1-2 所示。

| | |
|---|---|
| 使数据可知 | • 通过安全可控的手段使用数据资源目录或数据管理元数据<br>• 具备发现、识别新数据、新数据关系的能力 |
| 保证数据可信 | • 保证数据来源、数据收集手段及渠道可信<br>• 保证数据加工与处理过程所采用的计算、存储资源和能力可信 |
| 使数据可用 | • 结合数据分类分级和数据脱敏、加密使用策略，提供精细化数据访问授权和访问控制，保证按数据最小权限使用 |
| 数据可动态分类分级 | • 按照数据分类分级规则、策略要求实现动态数据分类分级，并按照业务需要对敏感数据进行处理或保护 |
| 数据可按需脱敏 | • 在数据收集、存储、加工处理和使用过程中，按照响应的数据脱敏及隐私信息处理策略要求，对数据进行脱敏处理 |
| 保证数据可溯源 | • 严格规范记录数据收集、加工处理以及数据使用的日志记录，分析并形成血缘关系，实现过程流溯源与数据流溯源 |
| 保证数据合规 | • 按照敏感信息（含隐私信息）合规要求，在各环节提供安全控制措施<br>• 按照国家及行业对数据出境的合规性要求，严格管控数据出境行为 |
| 使数据被安全保护 | • 按照数据分类分级保护要求，实施差异化的安全保护措施，并通过安全运营保证安全保护措施和策略的匹配度 |

图 1-1-2　数据处理与数据治理场景安全要求

**8. 落实数据安全责任是数据安全发展的重要驱动力**

一个组织的数据安全往往涉及多部门、多岗位、多角色的协同参与，绝不是安全部门的“独角戏”，明确数据安全责任是落实数据安全要求的重要抓手。实践表明，系统梳理保护场景中各阶段、各环境数据安全责任是数据安全的首要工作；结合安全责任制定安全机制、分解安全任务、形成安全逻辑是落实数据安全的不二法门。随着业务的持续发展及业务模式的不断创新都会带来安全责任的变动和调整，因此，数据安全责任不是一成不变的，优化并调整安全责任也是安全运营的重要工作内容。

## 1.1.4　数据安全深化了我们的数智化认识

数据安全与业务密不可分，是我们对数据安全的关键认识之一。当认识到这一点，我们的数智化认识与数据安全之间便发生了“化学”反应，再看业务数智化建设发展工作，也发生了一些“质”的变化。

（1）数据安全是业务信息化、数字化、智能化的内生需求。数智化全过程都要融合数据安全与数据合规。

（2）数据合规是业务合规的构成内容。原有业务合规主要是业务规则约束，其实也包括数

据合规，比如个人信息采集要事先告知，或者非业务负责人不能越权查看非本人负责的业务内容等。

（3）信息化、数字化、智能化的全技术要素都遵从数据安全要求。基础设施建设、应用开发、数据治理、应用开放、系统运维等各环节都离不开安全与合规的框架，只是环节不同、角色各异、职责有别以及要求分殊。

（4）数据安全的技术性要求将持续挤压并重塑数智化发展的技术路线，包括分层架构、应用与数据分离、云原生开发技术等。

（5）IT 架构与安全架构要实现“五同步”。即同步规划、同步设计、同步建设、同步实施和同步运营。

（6）推动数智化必须学习安全合规、理解安全合规，培养既懂业务又懂安全的复合型人才。

### 1.1.5　义无反顾地拥抱数据安全

在数据安全的实践过程中，从安全认识的建立、安全规划设计，到安全解决方案的制定、安全产品的选择都遇到了太多的问题，也看到了数据安全产业的巨大发展机会。本着“绕不开就加入”的原则，我们义无反顾地进入了数据安全领域，结合团队自身的特点，逐步形成了自上而下理解数据安全、自下而上实践数据安全的方法路径，构建了涵盖数据安全架构思想、能力体系设计思路以及实战解决方案，并在多个重大项目中得到了检验和提高。

**【感悟】数据安全，唯有迎难而上**

回顾我们与数据安全“同行”的过程，历经混沌、迷茫、清醒，再度陷入混沌、迷茫、清醒的螺旋上升过程。在这个过程中，我们看到了不同思路和方法，学习了不同理论和技术，实践了不同的安全目标与业务场景，也合作了不同的组织和产业团队，最深的感悟就是“难”，但也必须迎难而上，需求侧与供给侧都没有退路。当然，这个过程也让我们收获满满，希望通过本书与大家分享，以期互相促进、共同进步。迎难而上的数据安全之路，希望你我成为同路人！

## 1.2　提供数据者说数据安全

### 1.2.1　谁是数据的提供者

要看清楚谁在提供数据，可以从两个视角分析。

**1. 从数据种类视角看**

（1）个人信息：个人在政务办理、电信服务、金融服务、互联网服务等过程中所提供、记录及留存的个人基本信息与业务办理记录等。个人信息的提供者是个人，其权属归属于个人。

（2）组织信息：组织在政务办理、电信服务、金融服务、互联网服务等过程中提供、记录

及留存的组织基本数据与业务办理记录等。组织信息的提供者是组织，其权属归属于组织。

（3）业务信息：个人或组织作为业务或服务主体，在为他人或组织办理业务、提供服务过程中留存的各类数据，包括业务活动主体数据、物联网感知端采集数据、业务活动过程数据、业务活动结果数据、业务日志数据，以及为特定业务目的而在原始业务数据基础上加工形成的新数据。业务信息的提供者是业务办理方或服务方，其权属关系比较复杂，需要根据具体的业务特征进行差异化分析，这也是目前数据要素流通中遇到的难题之一。需要特别说明的是，我们所说的业务既包括组织的主营业务，也包括运营、运维、安全管理及安全监管业务。

（4）互联网公开信息：个人或组织在互联网发布的各类公开信息。此类信息的提供者即为发布者，虽可供各方共用，但可能存在版权等相关法律约束。

**2. 从数据全生命周期视角看**

在数据全生命周期中，可能会有多个角色参与数据的提供。

（1）原始数据提供方：原始数据提供方是提供原始数据的责任方，是数据的源头。数据提供方既可以是个人也可以是组织，提供的数据既可以是个人信息、组织信息、业务信息，也可以是互联网公开信息。

（2）数据收集方：具有数据收集职能的数据收集方也可以按照合规性和业务性要求将收集到的数据提供出去。数据收集方既可以是个人也可以是组织，提供的数据也是多种类，底线要求就是安全合规，要安全合规地收集数据，并按照需要和授权安全合规地提供数据。

（3）数据加工与处理方：具有数据加工与处理职能的数据加工方或处理方，按照合规性和业务性要求将加工处理后形成的数据对外提供。数据加工与处理方一般是组织，提供数据的底线同样是安全合规，要安全合规地加工处理数据，并按照需要和授权安全合规地提供数据。

（4）数据使用方：数据使用是一个宽泛的概念，包括通过业务访问使用数据、通过分析工具分析挖掘数据、为人工智能提供计算数据等场景。参与其中的数据工程师、数据科学家、人工智能训练人员等角色，都有可能将工作中获取的、工作过程中产生的中间数据以及工作结果数据提供出去，当然使用和提供数据的底线同样是安全合规。

（5）数据管理与治理方：在数据生命周期中，数据管理与数据治理是一个重要的环节（在这里我们不纠结数据管理和数据治理的概念，有兴趣的读者可以找专业资料进行学习）。在数据管理与数据治理中会形成各类管理数据、元数据、数据标准、数据目录等，这些数据是组织的重要数据，需要有针对性的安全保护，也需要根据管理、运营、运维要求安全合规地提供使用。

多方参与的数据提供关系如图 1-2-1 所示。

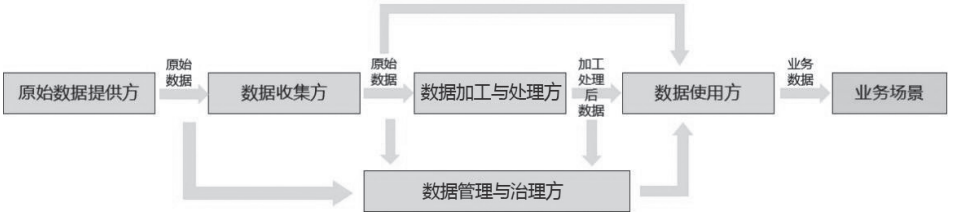


图 1-2-1　多方参与的数据提供关系