绿盟数据安全白皮书 2.0

■ 文档编号		■ 密级	完全公开
■ 版本编号		■ 日期	2020 年 3 月 16 日

■ **版权声明**

■ **适用性声明**

本模板用于撰写绿盟科技内外各种正式文件，包括技术手册、标书、白皮书、会议通知、公司制度等文档使用。

目录

插图索引

引言

我国于 2017 年 6 月 1 日正式施行《中华人民共和国网络安全法》，规定了网络运营者对其收集的公民个人信息必须严格保密，不得泄露、篡改、毁损，不得出售或者非法向他人提供。个人信息安全得到真正的法律保护，从此确立了公民个人信息保护的基本法律制度，促进经济社会信息化健康发展。

依据《中华人民共和国网络安全法》第三十一条，阐明了保护范围是国家对公共通信和信息服务、能源、交通、水利、金融、公共服务、电子政务等重要行业和领域，以及其他一旦遭到破坏、丧失功能或者数据泄露，可能严重危害国家安全、国计民生、公共利益的关键信息基础设施。保护方法为在网络安全等级保护制度的基础上，实施重点保护。重点保护的对象及关键信息基础设施，包括设施保护、数据保护、产品和服务保护，而数据保护的对象为“个人信息”与“重要数据”。

以《中华人民共和国网络安全法》为核心，我国就数据安全依法出台多项新政策，就包括已提请审议草案的《数据安全法》，对外征求意见的《数据安全管理办法》《个人信息出境安全评估办法》，已发布的有《信息安全技术个人信息安全规范》《儿童个人信息网络保护规定》《网络安全等级保护制度》 2.0，所有新政的数据保护核心对象依然都是“个人信息”和“重要数据”。

从上述内容可以看出国家在数据安全保护层面的力度，以及对数据安全保护的重视。

一. 数据安全概览

一直以来，我国政府积极探索和研究云计算在电子政务中的应用，也鼓励党政部门使用云计算进行服务模式创新，并开展了一系列试点示范工作。现在，政务云已经在全国范围内建设和普及，为我国政府进行职能转变，构建为民务实高效政府，落实厉行节约工作，发挥了积极作用。目前，我国电子政务进入了改革的深水区，正在加快落实“互联网+政务服务”，数据交换和信息共享已经是信息化建设重点，各领域的政务信息系统和政务数据、公民个人信息，已经或正在迁移至政务云平台上，加强网络和信息安全保护工作已经成为必然，其中数据安全是重中之重。

1.1 数据安全建设工作难点

随着云计算、大数据、物联网、移动互联网、人工智能等新技术的发展，网络边界被不断打破，数字双生、敏捷创新、安全合规驱动快速转型，社会和企业都在面临数字化的转型带来的数据安全风险。

近年来数据泄露的安全事件频发，国家和机构对数据安全的重视程度不断提高，数据安全已经与关键信息基础设施一并成为影响国家稳定、民生安全及社会安全的关键因素。

1.1.1 企业的数据安全体系建设不完善

- **传统信息安全体系无法保护数据安全：**有别与传统信息安全防护体系，由于数据安全防护体系将保护对象聚焦在“数据资产”这样的无形资产上，数据资产的机密性、完整性以及可用性与硬件资产存在着巨大差别，这导致传统信息安全防护体系通常不具备对数据安全的有效保护能力。
- **静态防护策略无法保护数据安全：**通常一个信息系统中的硬件资产数量是有限的，且在没有重大的系统变更时不会发生显著变化，所以传统信息安全体系的安全策略的设计思路往往是静态的。而随着大数据技术的广泛应用，以及移动互联网应用的蓬勃发展，企业数据存储、处理平台所承载的数据量正在以极快的速度爆炸式增

长，若仍以静态的视角看待数据资产势必无法应对数据量急剧增长带来的数据泄漏、数据损坏、数据篡改以及对数据主体造成影响等安全问题。并且由于数据资产对流动性的要求，仅考虑当前主体的静态防护策略显然无法有效保证数据的安全。

- **数据资产的权责不一致**：数据通常来自于企业的业务部门，在业务部门使用，并且数据的所有权也常常属于业务部门，但由于数据安全策略有时会限制业务部门对数据使用的权力，而数据安全体系建设工作由安全部门主导，数据安全防护体系的建设会很有可能受到来自于业务部门的阻力，数据安全体系建设工作推动困难。

1.1.2 数据安全体系建设目标模糊、建设步骤不清晰

- **缺少数据安全体系建设指导方针**：数据资产在许多环境下对可用性的要求极高，并且由于数据资产对流动性的依赖，如何在保障数据可用性与流动性的前提下落实对数据机密性与完整性的保护是企业所面临的重要问题。
- **缺乏数据安全体系建设经验**：由于数据安全体系建设与传统信息系统安全建设存在着保护范围不同、保护对象不同、安全策略类型不同以及安全建设思路不同等差异，对于企业来说数据安全建设是一项全新的课题。
- **缺少合适的建设指导**：由于我国的数据安全研究正处在逐步推进的阶段，暂时缺少直接有效的指导标准或行业最佳实践帮助企业明确数据安全体系建设的方法与步骤。
- **缺乏有经验的数据安全人员**：由于我国数据安全建设工作正处在起步阶段，企业安全团队缺少有数据安全经验的人员，这导致数据安全建设难以有效的执行。

1.2 数据安全整体架构

在数据安全建设体系上绿盟科技提出“一个中心，四个领域，五个阶段”的顶层设计。一个中心是指以数据安全防护为中心。四个领域是指的数据安全建设的四个领域：组织建设、制度流程，技术工具和人员能力。五个阶段是指的数据安全建设的五个阶段：业务梳理，分级分类，策略制定，技术管控，优化改进。

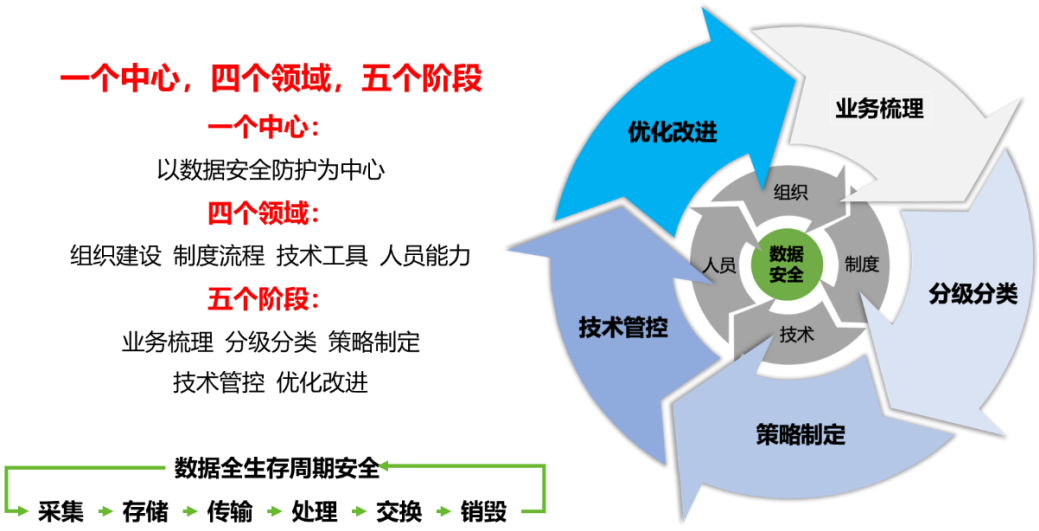

图 1.1 绿盟科技数据安全建设体系顶层设计

在数据安全建设体系中，组织建设、制度流程，技术工具，人员能力，4 个领域都需要同步开展建设工作，组织层面，决策层、管理层、执行层必须在数据安全建设领域达成一致，数据安全建设工作必须得到组织高层的支持。组织高层在数据安全领域的战略目标应该能够被管理层和执行层实现。

配套着在人员方面，要有相应的各级人员团队去进行相应的工作，数据安全相关工作人员应该依据岗位不同，具备管理能力、运营能力、技术建设能力，最关键的是要具备合规能力。目前数据安全法律法规对企业数据相关业务影响很大，一旦出现不合规情况也会造成很严重的后果。

在流程制度方面，从宏观的组织发展方针、组织战略，到中观的管理制度规范，一直到微观的计划报告表格日志等等，应该同步的进行建设。阐述清楚数据在组织中的战略地位，明确数据应该如何被管理，技术如何保障，进而逐步细化到日常企业的报告表格日志等运营工具的建设。通过流程制度建设指导“人”利用“工具”实现组织的数据安全战略目标。在技术层面，分级分类，数据防护，数据脱敏，流程审批，权限管理，数据标准等等，在各个领域都应该由技术工具予以支持。技术方面的内容我们后面也会详细介绍。

我们日常所说的“三分技术七分管理”也好“七分技术三分管理”也罢，都是在表明，管理是技术的运营依据、技术是管理的落地保障，两者要相辅相成。

二. 数据安全体系建设的步骤

数据安全体系建设的步骤，应借鉴 Gartner 的数据安全治理框架，其中定义了数据安全建设的五个阶段，及业务梳理、分级分类、策略制定、技术管控、策略优化。

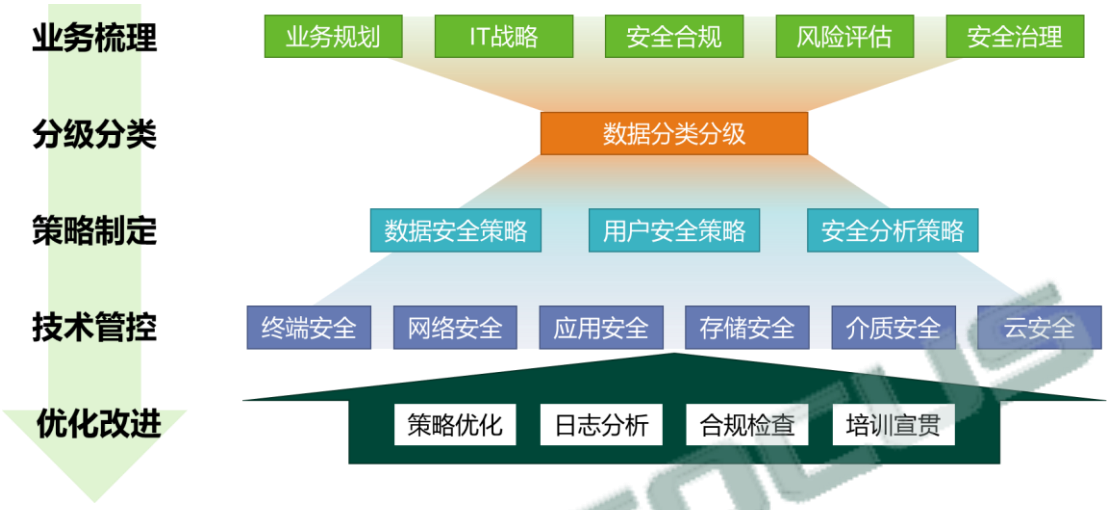

图 2.1 Gartner《How to Use the Data Security Governance Framework》

绿盟科技通过对国内法规的理解，以及对国内企业情况的研究，形成了一套具有中国特色的数据安全方法论。总结起来就是五个字“知”、“识”、“控”、“察”、“行”。

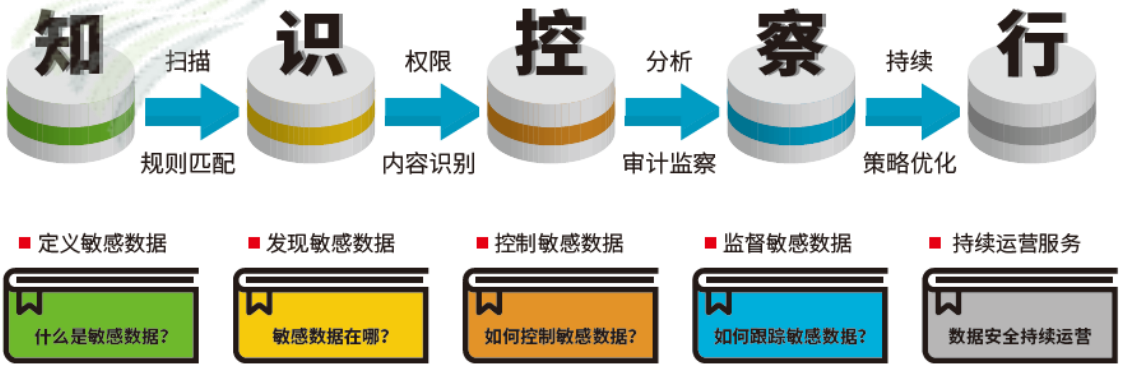

图 2.2 绿盟科技数据安全方法论

- **知：**分析政策法规、梳理业务及人员对数据的使用规范，定义敏感数据；
- **识：**根据定义好的敏感数据，利用工具对全网进行敏感数据扫描发现，对发现的数据进行数据定位、数据分类、数据分级。
- **控：**根据敏感数据的级别，设定数据在全生命周期中的可用范围，利用规范和工具对数据进行细粒度的权限管控。

- **察：**对数据进行监督监察，保障数据在可控范围内正常使用的同时，也对非法的数据行为进行了记录，为事后取证留下了清晰准确的日志信息。
- **行：**对不断变化的数据做持续性的跟踪，提供策略优化与持续运营的服务。

2.1 业务数据梳理

在组织与制度设计方面，传统网络安全均由 IT 部门负责，随着数据治理工作的深入开展，业务部门要深入参与数据资产梳理以及分级分类工作之中，因为只有业务部门是最了解数据价值与重要性的。因此需要自上而下形成高层牵头，横跨业务部门与安全部门的组织架构。由信息安全管理团队和数据业务管理团队共同商讨建立数据安全制度流程体系。制定好制度体系应该以文档化的方式进行落地管理。从最高级的方针战略，到最细节的表格日志，都应该由不同层级的团队负责进行文档化的落地，并严格执行。在相应的业务组织与管理制度指导下，企业才能更好的开展后续建设工作。

2.2 敏感数据的定义与识别

开展数据安全建设的第一步就是：定义什么是敏感数据，基于业务特点进行数据的识别、数据分类、数据分级。数据分类分级的准确清晰，是后续数据保护的基础。由于数据类型不同，对企业影响不同，我们建议根据《中华人民共和国网络安全法》要求对个人信息和重要数据分开进行评估与定级，再按照就高不就低的原则对数据条目进行整体定级。以个人信息为例，根据个人信息中数据的敏感程度，将数据分级分成四级，一级为低敏感级，二级为较敏感级，三级为敏感级，四级的极敏感级。对业务系统中的数据条目进行拆解，对其中每一个元数据进行分级，例如家庭住址为四级，身份证号码为三级，消费账单为二级，然后按就高不就低的原则对数据条目进行整体定级，那么这个数据条目应该定为四级。

为了更准确的对敏感数据进行定位与检索，绿盟科技用到了智能数据检索技术，此技术可以针对于当前大数据时代中常见的三类数据座检索，及结构化数据、半结构化数据、非结构化数据，这三类数据的最主要区别在于是否存在预先定义好数据模型，更确切的说是概念数据模型。

2.3 数据全生存周期安全风险评估

完成敏感数据分类分级后，就要到风险识别的步骤：发现哪里有敏感数据，并对敏感数据进行梳理与风险评估。敏感数据发现与数据风险评估的工作要结合人工服务和专业工具共同完成。

数据安全风险评估可以从数据的生存周期角度逐个考虑，这里引用国标 GB/T37988-2019《信息安全技术 数据安全能力成熟度模型》DSMM 架构图中的数据生存周期安全的步骤：数据采集安全、数据传输安全、数据存储安全、数据处理安全、数据交换安全、数据销毁安全。

数据在采集、存储、传输、处理、交换、销毁的数据生存周期中，会在 IT 系统的各种环境中存在，因此，环境的安全成为数据安全的一个重要因素。对于攻击者来说，IT 系统的方方面面都存在脆弱性，这些方面包括常见的操作系统漏洞、应用系统漏洞、弱口令，也包括容易被忽略的错误安全配置问题，以及违反最小化原则开放的不必要的账号、服务、端口等。IT 系统一旦出现安全隐患，都会导致系统环境中的敏感数据泄漏或丢失。

数据生存周期安全风险评估应从通用安全和各阶段安全两个方面进行数据环境风险检查，了解信息系统总体安全风险状况，对脆弱性的所有方面统一进行分析和评估，并提出整改意见，帮助客户建立快速响应机制，及时有效完成数据安全风险修复工作。

图 2.3 数据生存周期安全风险评估

2.4 数据安全纵深防护

针对数据安全的风险，应以数据为中心，向外对业务、网络、设备、用户采取“零信任”的态度，既然每个环节都不可信，那么管控手段就要覆盖全部环节，任意环节失信后都能实现熔断保护。

用户侧、终端侧、网络侧、业务侧，以及数据中心，都要做好安全防护措施，外向内防攻击、防入侵、防篡改，内向外防滥用、防伪造、防泄露。最关键的是，要对全部纵深防护环节进行整体控制，实现环境感知，可信控制和全面审计。整合多层次的纵深防御，及时发现问题，及时阻止安全问题。总之我们的防护宗旨是认证好人并允许其通过，识别坏人并阻断其访问。

- 构建统一身份认证管理平台，通过对用户、组织机构、权限管理和各类应用系统资源的规范化、标准化管理，在与各类已建业务应用系统授权管理功能有效对接的基础上，健全统一、灵活、多维度的门户访问控制机制，对应用访问权限进行全流程监管，对不同权限用户对数据的访问起到准确的定位与跟踪。
- 利用内容识别技术和网络协议识别技术实现网络数据的防泄漏。
- 利用加密技术和驱动技术实现终端数据的防泄漏。
- 在非生产环境中（测试、统计分析等），当敏感数据从生产环境转移到非生产环境时，这些原始数据需要进行统一的静态脱敏处理，然后可以直接使用这些脱敏数据；
- 在生产环境中，访问敏感数据应及时进行动态脱敏，根据访问需求和用户权限进行“更小颗粒度”的管控和脱敏。
- 为了确保敏感数据不被泄漏和篡改，应采用各种技术能力的联动来做全面的防护，并根据环境的变化智能的做出响应，从而降低数据被泄漏和篡改的可能性。

在人工智能、机器学习的大环境下，应该利用更加智能的方式实现数据风险的动态联防。

基于人工智能的数据安全管理系统作为整个 IT 系统的大脑，对整个网络的威胁、日志持续监控，通过内外部的态势数据、评估数据、情报数据的内容，利用大数据分析系统对事件进行快速分析，当发现敏感数据泄漏风险、黑客入侵风险、数据篡改风险、越权访问风险，等数据安全风险时，智能判断风险发生的位置、路径、方式等，从而快速向风险发生最近的或最直接的安全防御系统发出指令，安全防御系统根据指令调整安全策略，对链路、会话、行为等做出警告并阻断操作，实现自动化的响应能力。

2.5 敏感数据监察分析

敏感数据监察分析、发现安全问题与异常事件。可以考虑从用户、资产和数据的行为模式出发，利用 5W1H 分析模型来进行敏感数据行为分析，基于行为模式发现数据异常事件。也就是我们常说的 UEBA（User and Entity Behavior Analytics 用户及实体行为分析）。

将数据挖掘与分析技术应用到安全领域是目前安全分析重要手段，相关技术帮助安全分析人员节省了分析时间和分析量，提高分析数量和准确度。

在数据安全领域，传统的安全分析方法存在信息量大，有效信息少的问题，用户行为分析与机器学习技术能够优秀的解决上述问题帮助使用者更好的识别数据安全风险。

基于历史的可信访问行为提取访问规则，利用各类算法进行行为聚类，形成可划分的访问行为簇并可视化呈现。通过这种图谱分析与可视化展示让管理者对于敏感数据访问情况，由一无所知转变为可视可管。

2.6 优化改进与持续运营

当我们具备“知识控察”的能力后，不代表我们的数据是安全的。业务是在变的，数据也是在变的。因此我们的安全也是要不断变化的。为了应对变化，我们在“知识控察”的基础上提出了“行”，这是一个动词，代表着对数据安全的优化改进与持续运营。

在大的层面，合规要求指导安全策略的设置，安全策略支撑合规要求的落地，二者相辅相成，配合上持续优化改进运营的“知识控察行”体系，实现持续自适应的数据安全防护能力。

三. 数据安全新技术

当前，数据安全技术领域仍然有巨大的发展空间。在数据安全合规、数据利用与隐私保护等多重需求下，亟需发展和引入一批新型技术。本章节从企业应用需求出发，选择三类前沿技术进行介绍：(1) 数据脱敏效果的评估技术；(2) 数据发布/挖掘下的隐私保护技术；(3) 不可信环境下的安全计算技术。以期推动前沿技术在企业场景的成果转化与应用。

3.1 数据脱敏后的效果评估技术

企业应用的数据脱敏系统涉及的算法与策略种类较多，如泛化、屏蔽、截断、加噪和 FPE 加密等。用户通常可自行选择不同的脱敏方法与配置对个人隐私或企业敏感数据进行脱敏。在某个特定场景，比如企业间的数据共享，其脱敏数据后的数据是否达到了预期效果，即脱敏数据残余的风险是否在该场景下的可控范围内？为了检测和判定这个问题，亟需一套行之有效且统一的评估检测技术与辅助工具。对于数据脱敏的效果的评估，根据脱敏数据类型包括两类：针对个人隐私数据脱敏的效果评估技术，另一类是针对企业敏感数据脱敏的效果评估技术。

3.2 数据发布/挖掘下的隐私保护技术

通常来说，脱敏系统拥有的丰富脱敏方法与策略可以满足企业内部的产品测试与开发、数据统计分析等基本需求；然而，若企业需进一步扩大共享范围或者进行深层次的数据价值挖掘，发掘数据的最大化价值，那么产生了隐私保护前提下的处理与发布需求。这样的场景通常被称为隐私保护的数据挖掘（Privacy Preserving Data Publishing，PPDP）和隐私保护的数据挖掘（Privacy Preserving Data Mining，PPDM）。其中，匿名化是 PPDP 场景的关键技术；差分隐私是 PPDM 场景关键技术。

3.3 不可信环境下的安全计算技术

在一个开放的网络环境，常常被认为是不可信任的，因为存在潜在的各种安全威胁与攻击。对于不可信任环境下的数据安全存储，目前较为流行应用“去中心化”的区块链技术，通过工作量证明和密码学机制解决交易安全和存储不可篡改问题；而对于不可信任环境下的数据安全存储，比如公有云的外包计算、多个机构数据共享与联合计算，目前存在同态加密和安全多方计算两大类技术。

四. 总结

在大数据时代背景下，数据安全与数据利用不应是矛盾的、对立的关系，数据安全应该更好地为数据利用提供服务，采用新型技术手段，为企业提供新的思路和实践方案。

数据安全建设是一个长期和持续的过程，如何更好地进行数据安全建设，利用一套科学的、系统的数据安全建设体系对于一个企业来说十分重要且必要。数据安全建设需要技术和管理双管齐下，在建设的过程和环节中，需充分利用和发挥好各种关键技术的作用，在优化改进环节中，引入新技术来优化技术和管理流程，通过实现自动化和半自动化以降低运营成本，在安全的同时让数据价值最大化。