

# 统一的治理与整合：深入洞察与合规管理的基础



## 要点

- 合规工作看似是费力费钱的苦差事，但实际上它能带来显著效益：有助于促进企业实施数据治理战略。
- 得益于最近的技术进步，机器学习克服了人类实践所面临的重大局限性。
- 为一个项目建立“统一的治理与整合”(UG&I)的基础后，该项目便可成为企业或业务单位中其他用例的基础。
- 在企业经历数字化转型之际，企业领导逐渐认识到：对整个数据领域（无论是在本地部署环境还是公有云和私有云环境中）实施数据治理将会产生显著效益。

在形形色色的企业中，数据种类不断增加，数据总量呈爆炸式增长。在多云环境中，物联网、社交媒体、移动设备、虚拟现实实施和光学跟踪等各种新数据源产生的信息流以指数级激增。

与此同时，相关法规规定了不容忽视的合规要求。在颁布的各项相关法规中，欧盟的[通用数据保护条例 \(GDPR\)](#) 无疑是企业领导最关注的一个。不过，始终遵守金融服务和零售行业的 PCI-DSS、医疗保健行业的 HIPAA 以及其他法规也同样重要。

## 基于“了解数据”、“信任数据”和“使用数据”三原则，UG&I 共包含七个构建模块。

随着数据的激增，遵守法规以及利用自由流动数据的重要性就凸显出来。如果不借助采用统一的治理与整合解决方案的可信分析战略，对数据进行控制，企业不仅可能被如潮的信息所淹没，还可能面临违反法规的风险。幸运的是，诸如机器学习和人工智能等新兴技术可帮助消除障碍，健全监管与合规。有远见的企业领





导逐渐认识到，数据治理不仅对于合规管理至关重要，同时也为敏捷业务和创新奠定了基础。

### 可信分析基础的构建模块

合规工作看似是费力费钱的苦差事，但实际上它能带来显著效益：有助于促进企业实施数据治理战略。统一的治理与整合 (UG&I) 计划应该应用于数据——无论是结构化还是非结构化数据，无论是公有云还是私有云中的数据。实施 UG&I 不仅可以为合规工作本身带来显著效益，还能为企业中的其他领域（尤其是分析能力）产生可观的附加价值。无论企业中所部署的分析工具的成熟度如何，任何分析计划都需要有可信的数据才能有效开展工作。数据可信是指这些数据来源可靠，而且您了解企业中使用数据的方式，并且知道数据如何增强分析项目。通过分析，还可能提出新的产品设计和营销方案，有助于改善销售、供应链或客户服务计划。分析还可发现运营流程中的低效环节，一旦消除，便可增强企业的敏捷性，提高收益。基于“了解数据”、

“信任数据”和“使用数据”三原则，“统一的治理与合规” (UG&I) 共包含七个构建模块。

- 1. 数据发现** 多年来，许多企业的数​​据都随意存储，到处乱放。因此，他们可能并未意识到企业中存储着海量数据。数据治理的第一步就是对企业的​​所有数据进行盘点。一种比较好方法是：首先重点关注某个特定项目，然后逐步扩展到更广泛的企业范围。选择一个可以快速获益的领域作为起点，可以起到示范作用，提高大家对于这项工作的积极性。
- 2. 编目** 发现数据后，就应对其进行编目，使用元数据标记来标识数据类型、用途、所有权、数据沿袭等。由于某些行业中的企业具有同样的需求，因此为这些行业预先构建数据模型有助于加速编目过程。
- 3. 质量** 冗余、过时或不重要的数据（统称 ROT 数据）的存储和管理成本高昂。ROT 数据还会使合规工作更加困难，并且严重影响分析工作。数据必须符合

一定的质量标准，并始终保持质量水平，以便确保后续使用顺利进行。

**4. 整合** 需要时，可轻松组合与转换多个来源的数据，并与其他有需求的系统共享。这涉及到将结构化和非结构化数据汇集在一起，并利用 Apache Atlas 和 Hadoop 等开源技术进行整合。

**5. 复制** 通过创建与原始数据自动同步的数据流，有助于确保数据湖、数据仓库、数据集市和影响点解决方案中拥有最新数据。随着数据量的增加，复制功能也应与时俱进，支持低延迟的大容量复制。

**6. 主数据管理 (MDM)** 提供有关企业客户或产品等关键实体的全面可信的视图，对现代企业而言是一项基本要求。现代 MDM 解决方案包含基于图形的分析探索功能、敏捷的自助式访问、治理功能以及用户友好型仪表板功能。

**7. 数据保护和生命周期管理** 战略信息资产和敏感信息资产应受到妥善保护。从创建到处置的整个数据生命周期都应受到有效管理，可采取记录管理、诉讼数据管理和归档存储等做法。

### 机器学习加速治理流程

得益于最近的技术进步，机器学习克服了人类实践所面临的重大局限性。机器学习可以大规模实现数据治理与整合计划的自动化，克服海量数据带来的困难，在整个企业范围建立有效的数据治理体系。比方说，企业拥有约两万个数据项，通常需要由六人组成的团

队花上六个月时间对这些数据项进行人工分类，以便确保各业务部门能够以可靠可信的方式使用这些数据项，推进分析工作。而在机器学习的帮助下，同样的过程只需数天甚至数小时即可完成，时间的长短取决于数据资产的规模。处理速度的极大提升解除了治理流程的沉重负担，减少了相关的合规工作，并为高效的分析计划铺平了道路。

### 单一基础满足多种用途

为一个项目建立“统一的治理与整合”(UG&I)基础后，该项目便可成为企业或业务单位中其他用例的基础，例如：

**可管控的数据湖** 各行各业的许多企业已开展了大数据预测性分析项目。关键的第一步是将海量的结构化和非结构化数据存储到数据湖中。过去，低质量的数据经常会影响数据湖的实施。**不受管控的数据**导致许多数据湖退化为包含 ROT 数据的数据沼泽，这些具有初始加密格式的 ROT 数据通常无法投入使用，并且往往不符合监管准则。然而，在实施 UG&I 后，数据湖中的数据可以和业务术语对应起来，不仅易于数据用户理解，而且整体一致，受到管控。将此类数据提供给数据用户，有助于加快数据科学、数据探索和人工智能项目实现价值的速度，并提供宝贵的洞察。

**企业数据仓库卸载** 随着时间的推移，许多企业不断将 ETL (提取、转换、加载) 作业添加到数据仓库中，形成了成本不菲的 ETL 执行环境。此外，数据仓库无法扩展为处理非结构化或半结构化数据，因此不适合执行数据探索任务。从数据仓库中**卸载数据**，既降低了成

本，还使数据能够与数据湖等环境中的其他数据类型组合，从而支持动态数据探索。

**客户体验** 当服务代表与客户互动时，他们必须拥有关于客户、产品或其他实体的可信、最新、准确的**单一视图主信息**。错误或过时的信息可能会破坏与客户的互动，影响客户的信任，最终导致客户流失。遵循 UG&I 流程的数据可确保每次客户互动都能成为深化客户关系、增强信任、积极提升品牌忠诚度和品牌价值的载体。

## 合规

对于监管要求而言，可信的分析基础有助于推动积极的合规管理。在应对 GDPR 等法规时，每个构建块都会发挥重要作用。最重要的是，要保护个人数据，首先要对个人数据进行定义，这样组织才能确定自己所拥有的个人数据。基础的数据目录之所以重要，是因为它包含有关数据质量、数据扩充与分析以及积极合规工作的所有监管规则。

金融服务行业的 PCI-DSS 和医疗保健行业的 HIPAA 等法规早已生效数年。两者都旨在保护个人信息免受侵害。相关保护措施尤其重要，它们建立了基于角色的数据访问权限，以防数据落入坏人之手。

## 结束语

在企业经历数字化转型之际，企业领导逐渐认识到：对整个数据领域（无论是在内部环境还是公有云和私有云环境中）实施数据治理将会产生显著效益。通过重点关注 UG&I 的七个治理实践——发现、编目、质量、整合、复制、主数据管理以及数据保护和数据生命周期管理，企业的数据准备工作不仅可以满足分析处理和洞察的要求，还有助于积极遵守相关法规。虽然数据量庞大，但机器学习和人工智能实践可帮助克服在数据映射、编目、海量数据匹配以及保持数据质量等任务中人工操作的局限性。

有远见的企业领导明白，现在花些时间建立稳固的 UG&I 基础，会在今天和不久的将来带来显著的回报。他们认识到，如果将数据治理作为推进业务优化、开拓创新以及积极合规的引擎，企业就将如虎添翼，无往不胜。

**要了解更多信息，请访问：**

<https://www.ibm.com/analytics/cn/zh/technology/agile/>

