流量反欺诈体系建设

**导读：流量反欺诈技术是一项重要的技术，可以帮助平台和企业识别和阻止欺诈流量，从而保护平台和企业的利益，提高用户体验，打击黑灰产，维护互联网的健康发展。流量反欺诈技术包括人机验证、规则引擎、机器学习、复杂网络和多模态集成等。这些技术可以从不同的角度识别欺诈流量，并将其拦截在平台和企业之外。流量反欺诈技术在各个领域都有着广泛的应用，包括广告、电商、金融、游戏等。通过采用流量反欺诈技术，平台和企业可以有效地减少欺诈流量带来的损失，提高自身的收益。流量反欺诈技术是一项不断发展的技术，随着黑灰产对抗手段的不断升级，流量反欺诈技术也在不断地更新迭代。未来，流量反欺诈技术将朝着更加智能化、自动化、协同化的方向发展。**

* 顺丰数据治理总体介绍
  + **体系建设演进路线**
  + 顺丰数据治理框架
  + 数据治理各领域能力介绍
* **顺丰数据治理实践分享**
  + 顺丰数据整理工作的关键要素
  + 顺丰主数据治理分享

**▌流量反欺诈背景**

**顺丰做数据治理十多年，数据治理体系的模块是逐步来建设的。十年前，我们就已经建了数仓，同步做了元数据管理，数据质量管理，以及数据安全的管理。顺丰数据治理的演进路线分3个阶段。**

**第一阶段：**

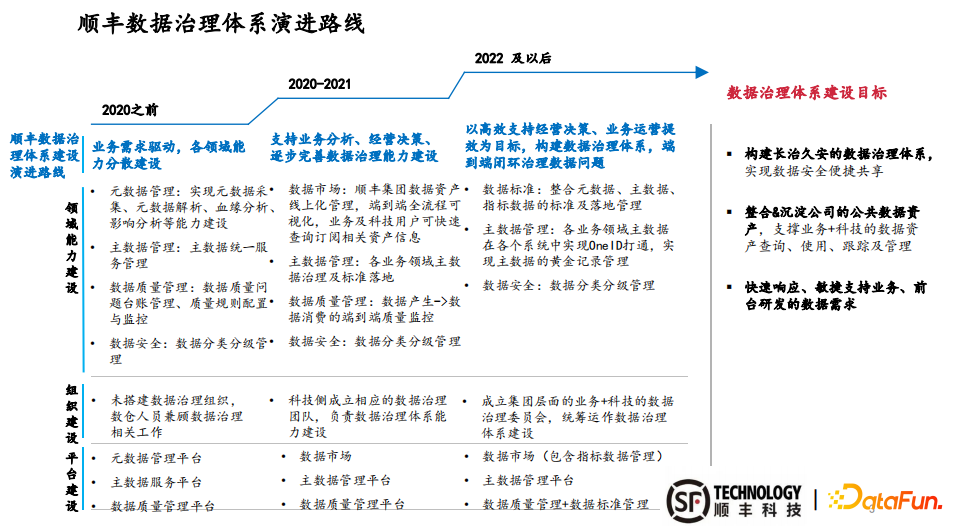
**2020年前，我们主要在进行数据平台的搭建，关键领域能力的建设。关键领域能力建设方面， 包括元数据管理、主数据管理、数据质量管理、数据安全。**

**第二阶段：**

**2020到2021年以及22年上半年，在加强领域能力建设的同时，我们更多的把数据治理的一些工作和业务更紧密的结合在一起，我们发现，在数据治理体系建设的过程中，如果没有业务的参与，该工作就变成了一个纯科技工作工作，难以有效的落地。因此，在2020到2021年，我们成立了专职的数据治理团队，来做数据治理体系能力的建设。 这个专职的团队，有科技方和业务方共同的参与，业务方包含了顺丰科技各条业务线上的业务，财务、采购、人资，市场等。**

**第三阶段：**

**22年及以后，更加深层次的进行整个数据治理体系的完善。这个阶段，工作聚焦在资产管理、数据标准的建设，还有主数据OneID的打通。组织建设上，我们成立了集团层面即顺丰数运层面的数据治理委员会，来统筹运作数据治理体系的建设。**

****

**我们数据治理体系建设的目标是：**

1. **构建长治久安的数据治理体系，能够实现数据安全的便捷共享。**
2. **把公司的公共数据资产沉淀下来。**
3. **快速响应、便捷支持业务、支持前台研发的数据需求。**

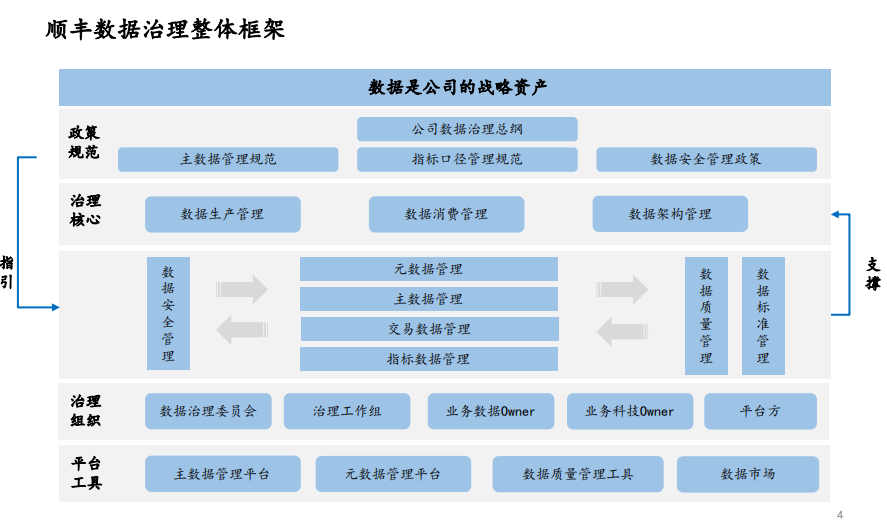
**▌顺丰反欺诈基础**

**顺丰数据治理的整体框架，最顶层的设计战略，包括政策规范，顺丰有顺丰集团的数据治理总纲，我们建立了各类的规范，主数据管理规范、指标口径管理规范、数据安全管理政策。**

**我们主要的治理领域包括，元数据管理、主数据管理、交易数据管理、指标数据管理，以及数据安全，数据质量，数据标准。我之前参与过华润、华为、碧桂园的数据治理工作，在数据治理领域方面，不同的公司大同小异，有些公司数据治理中不包含交易数据，有些公司会包含数据服务，数据模型管理。 在顺丰主要关注元数据、主数据、交易数据、指标数据的安全、质量和标准的管理上。**

**在数据治理组织层面，我们有顺丰的数据治理委员会、专职的数据治理工作组，也有相应的业务方和平台方。**

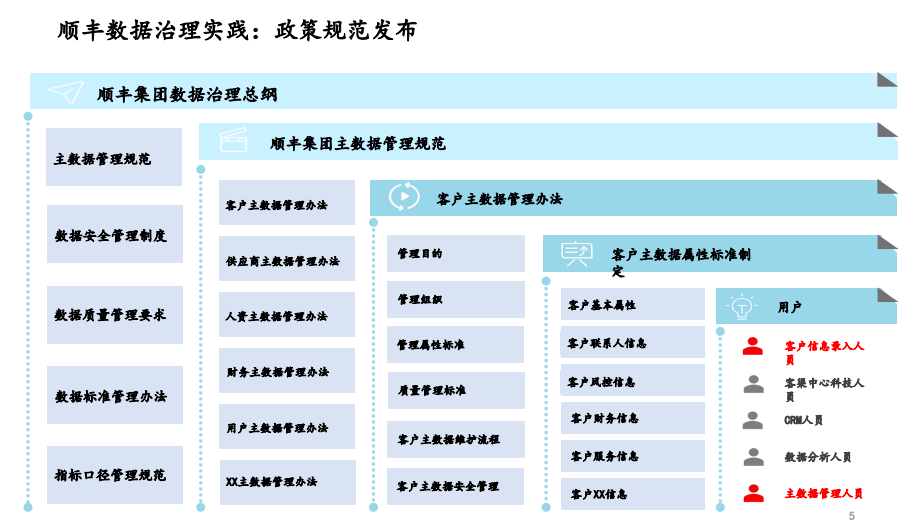
**在平台工具层面，我们有主数据管理平台、元数据管理平台、数据质量管理工具和数据市场。和其他的公司也类似。**

****

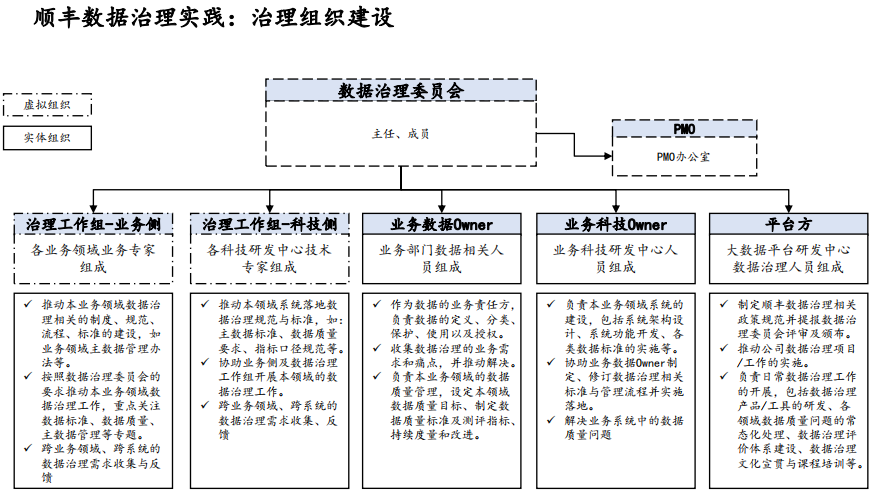
1. **政策规范**

**顺丰数据治理总纲是数据治理体系的纲领文件，顺丰数据治理体系的相应的规范政策都是在数据治理体系总纲的指导下，来进行编写的，包括主数据的管理规范，数据安全的管理制度，数据质量的管理要求，数据标准的管理办法，及指标口径的管理规范。**

**拿主数据的管理规范来看，基于顺丰集团主数据管理规范的要求，针对每一类主数据，我们要制定出相应的主数据管理办法：如客户主数据管理办法、供应商主数据管理办法、人资主数据管理办法、财务主数据管理办法，用户主数据管理办法等。管理办法涉及主数据的管理目的，管理组织，管理的属性标准等，以此来构建主数据管理规范的内容。**

****

1. **治理组织建设**

**我们成立了集团层的数据治理委员会。**

**首先它包含了业务侧的治理工作组，科技侧的治理工作组，这是两个虚拟组织，各由各业务领域的业务专家，各科技研发中心的技术专家组成。同样它包含了3个实体组织， 业务数据Owner，业务科技Owner，平台方。业务数据Owner参考了我们的GPO（流程Owner）的概念，业务部门的核心的领导就是业务数据Owner， 由业务数据Owner来承载业务数据的定义、分类、保护、使用及授权。业务科技Owner是业务数据Owner对应的科技研发中心，比如供应商数据的业务数据Owner是集团采购供应链中心，业务科技Owner是采购与综合解决方案研发中心。平台方是由大数据平台研发中心的专职人员组成，主要是负责协助业务数据Owner，业务科技Owner来做数据治理体系的整体的建设。**

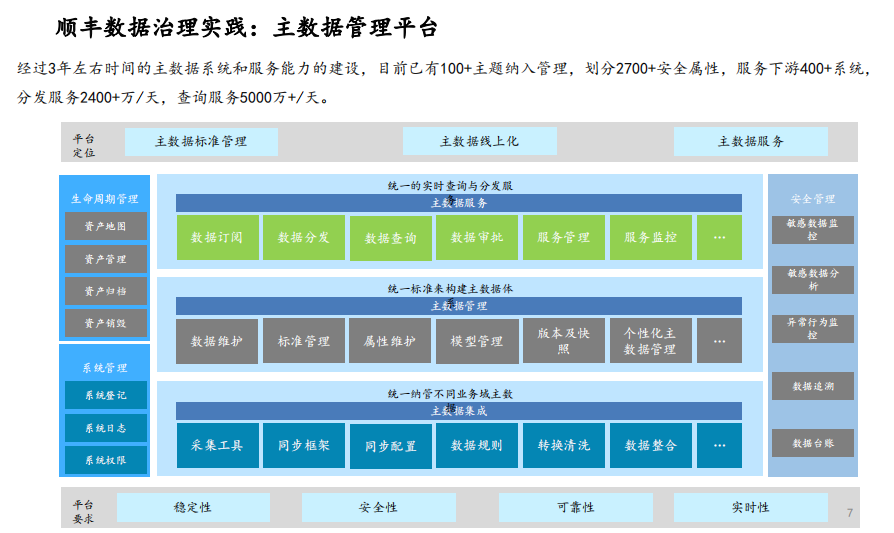
**这是我们在组织层面做的事情。 这么做是因为，关于数据治理的很多建设工作，包括标准的制定，如果仅仅是科技侧来制定这个标准的话，标准有时会慢慢的变形；同样，数据质量要达到闭环的话，从业务走到科技，从科技再回到业务，实现质量的闭环，这需要多方的参与，否则数据质量问题就会重复的出现；更重要的一个原因是，建设数据治理体系，我们需要一个组织，有一个一把手来统筹运作，规划指引。**

**顺丰数据治理委员会的主任是顺丰整个集团的CIO，它的成员包括顺丰各条线的业务的领导，科技的领导，我们是成立这样一个组织，来推进数据治理体系的工作。**

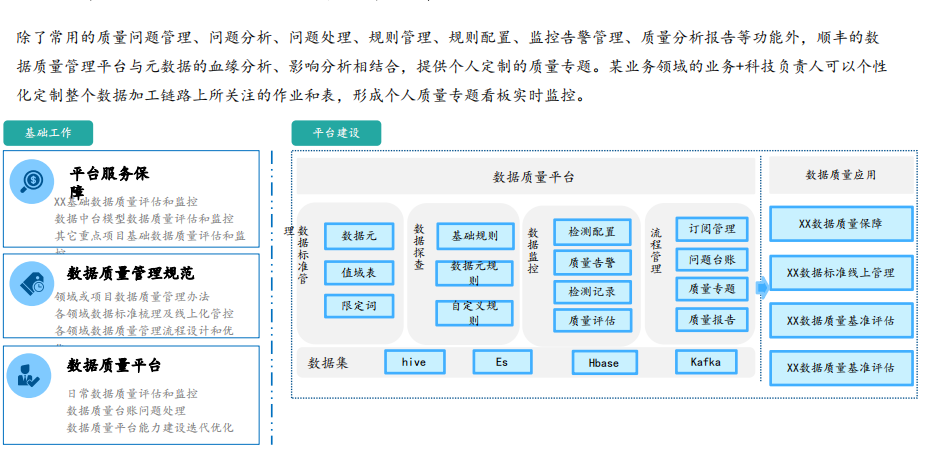
**▌流量反欺诈技术**

**1、主数据管理平台**

**主数据管理平台，从19年开始，经过3年的建设，平台的定位在于做主数据标准的管理，主数据线上化管理，统一的主数据标准的服务。主数据标准管理用于管理制定好的各类主数据的管理。IT数字化建设过程中会存在一些没有源头， 没有Owner的主数据，均需要做线上化管理。目前已经由100多个主题纳入主数据管理平台，包含了2700多个安全属性，服务对接给下游400多个系统，分发服务大概2400多万/天，查询服务大概5000多万/天。**

****

**2、数据质量管理平台**

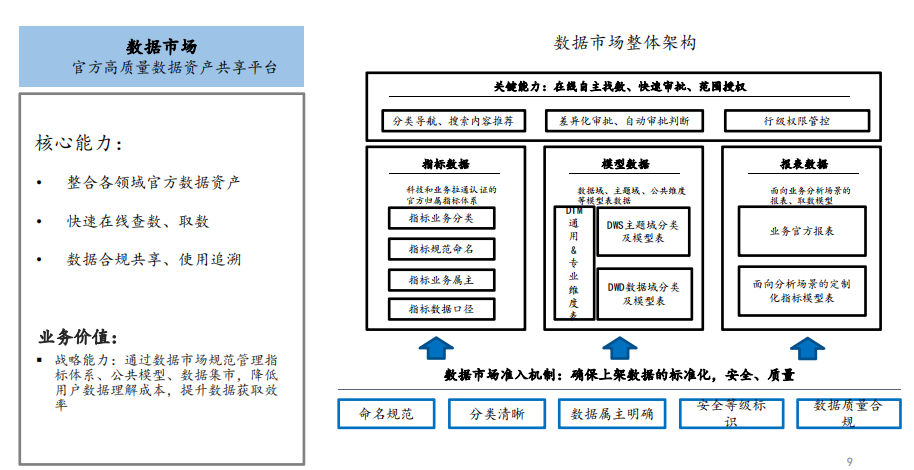
****

**数据质量平台管理，主要常用的功能包括数据质量问题的管理、问题分析、问题处理、规则管理、规则配置、监控告警管理、质量分析报告功能等。**

**管控规则分为了强依赖和弱依赖。质量平台的监控规则和调度的作业完全绑定在一起，质量平台会监控关键作业的运行情况。平台可对关键作业产生的表的数据波动，数据置信，平衡性校验，唯一性校验，空值校验等配置监控作业，并基于配置的监控作业进行监控。如果监控作业触碰强依赖的规则，系统就会对作业进行阻断，否则触碰弱依赖的规则，系统不做干预，作业将继续执行。这是一个数据质量平台监控系统和调度管理、以及元数据管理的血缘分析和影响分析结合的一个应用。**

**质量平台还提供了个人定制的质量专题看板。平台支持根据每个人关心的作业，进行个性化的配置，并提供实时、日、周、月等时间维度的监控数据。**

**3、数据市场**

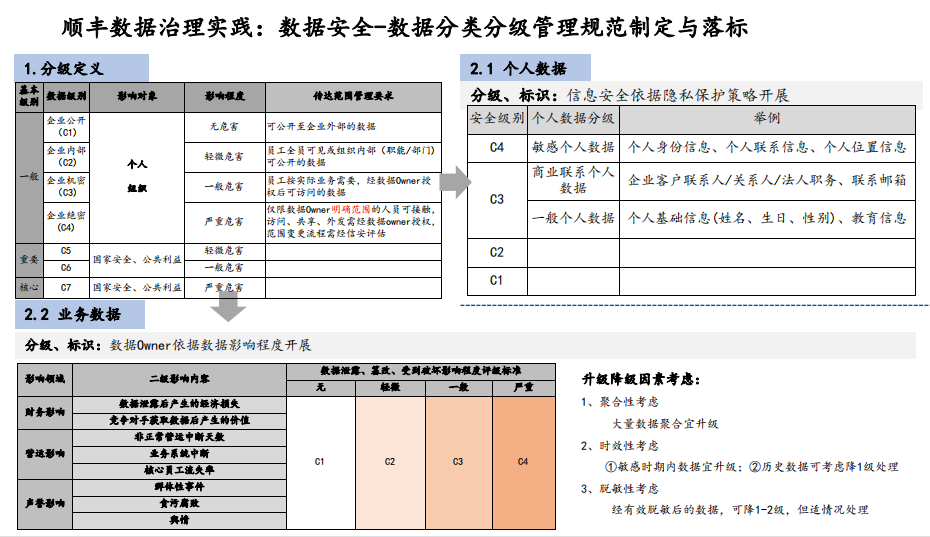
****

**在顺丰的数据市场，在其他公司称为数据字典、数据地图，这是一个数据资产线上化的工作。我们把公司所有的数据，指标、模型、报表还有最初的源系统来的ODS数据，进行分类，按照数据治理架构， 从主题域、主题、业务对象、实体、属性，进行划分和展示，可以说是技术元数据和业务元数据的整合展示。系统在顺丰的应用情况很好，科技侧的同事用该系统来查看模型数据、表数据，业务方的同事可用来查询指标数据，指标的业务分类、规范命名、数据口径、数据从哪张源表来，在哪张报表中展示，用途等。**

****

**可以看到示例。指标板块可以看到指标分类、技术负责人、业务负责人、指标口径、分析维度等。点开详情，可以看到整个的业务属性、技术属性，管理属性三部分；报表板块可以看到其归属组织、技术负责人、业务负责人等以及使用情况。模型表偏科技侧，可以查看表是谁建的，使用情况，包含字段，可以进行详情查询，系统会显示该模型表上下游的血缘分析，关联的表都查出来。报表和模型表，支持用户申请权限。数据市场涉及的用户非常的广，有科技侧的用户，也有业务用户，用户也需要一些跨业务领域的数据共享，数据市场提供了数据共享和权限申请的流程。**

**4、数据安全—数据分类分级管理规范制定与落地**

****

**数据分为两类，个人数据和业务数据。个人数据完全遵循个人隐私保护条例和个人数据安全保护法的规定。个人数据安全级别分为C1、C2、C3、C4级别。敏感个人数据对应C4级别、商业联系个人数据C3级别、一般个人数据C3级别。**

**其中个人身份信息、个人政治面貌，政治倾向、个人生理上的一些信息，个人联系信息、个人位置信息属于敏感个人信息C4级别，会得到严格的管控。业务数据从财务影响，营运影响、声誉影响3个方面进行划分，同样分为C1、C2、C3、C4级别，规则支持安全级别的升降级。数据的业务数据owner负责针对数据的标准为数据的安全制定相应的级别以及安全的管控策略，对数据进行有效的管理。**

**以上是对顺丰数据治理体系框架的介绍，以及数据治理各领域能力的介绍，以下是数据治理实践分享。**

**▌运营和情报体系**

**1、数据治理工作的关键要素**

**以下介绍顺丰做数据治理的实践。顺丰做数据治理工作进行多年，踩过坑，走过弯路，有了些心得和总结，跟大家做一下分享。首先介绍数据治理工作进行的关键要素。**

**我们总结了四点。基于我们的经验，结合了美的、华为、阿里和腾讯的一些数据治理专家的访谈资料，参考《华为数据治理之道》《DAMA数据管理知识体系指南》一些文献，并结合了顺丰科技多年的数据治理工作实践，我们将数据治理工作开展和推进的关键要素，总结为4点：**

**第一，高层领导的支持是特别关键的。整个数据治理体系的建设，要想有效的落地以及得到很好的推进，必须有公司高层领导的支持，并且需要有公司高层领导的持续关注和支持。在顺丰，数据治理委员会的主任由顺丰集团的CIO担任。**

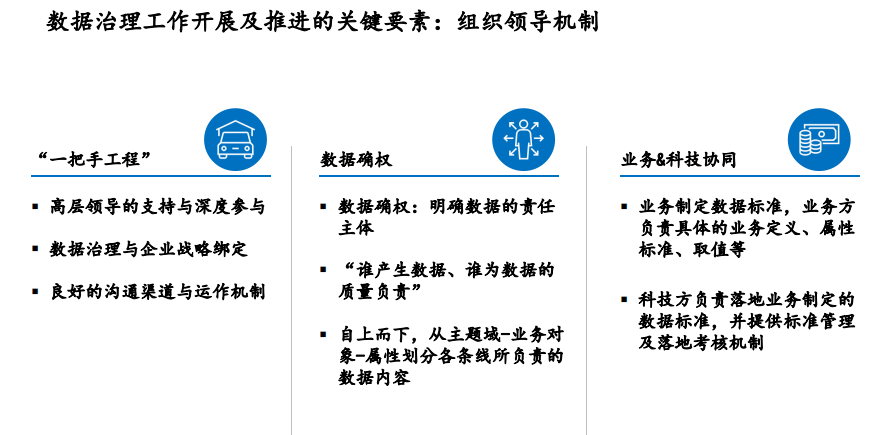
**第二，比较关键的是运营组织保障常态化治理，包含业务和技术。虽然我们在看数据治理的一些产品， 数据质量、元数据、数据安全等是偏科技的，但是实际上要想数据治理工作有效推进，就要有业务的参与。**

**第三，要考核与激励双结合。顺丰的数据质量是有数据质量健康度的考核评价。针对各个主题域的数据质量，顺丰都进行了健康度的评价打分，并对结果进行晾晒。**

**第四，长短结合，综合推进，分阶段实施。**

**以上是我们认为的四大关键要素，接下来分享顺丰数据治理工作的组织领导机制。**

**2、数据治理工作的关键要素：组织领导机制**

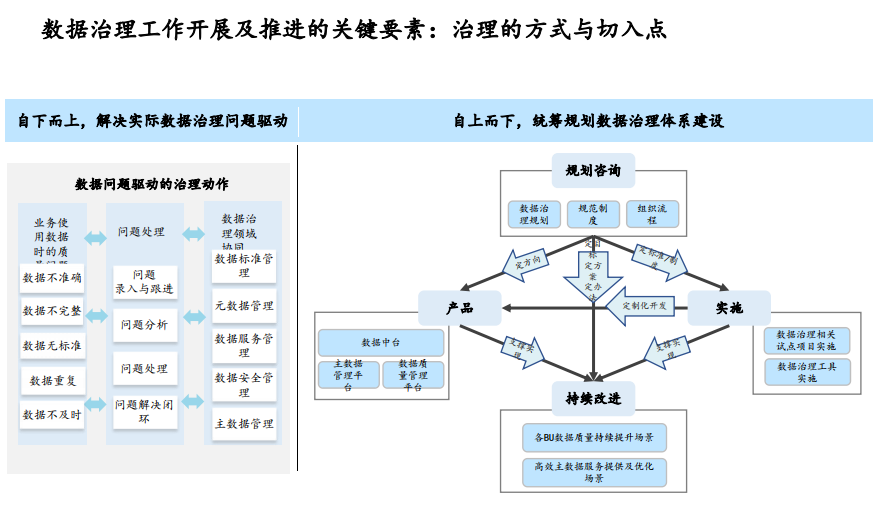
****

**第一，一把手工程，需要得到高层领导的参与。**

**第二，数据确权。在华为、碧桂园，说的是数据Owner，意思是要明确数据的责任主体。在顺丰，数据的Owner，最开始是科技侧，现在是业务方。然后还要明确的是”谁产生数据，谁为质量负责”,这句话喊了多年后，最近才逐步落地和得到尊崇。产生数据的业务部门并不使用数据，而是下游在使用数据。下游使用数据的时候， 才会真正的关心数据的质量，会发现数据质量没办法闭环的情况。因此，就产生了“谁产生数据，谁为质量负责”的口号。**

**第三，业务侧与科技侧协同。以上提到了“长短结合，综合推进”。顺丰也是这样做的。**

**3、数据治理工作的关键要素：治理的方式和切入点**

****

1. **自下而上，解决数据领域的问题。从实际的解决数据应用过程中的数据质量问题入手，然后具体去看它具体属于数据治理领域当中的哪些问题，识别具体是数据标准的问题、或者是元数据、数据服务，亦或者主数据管理的问题。 这样自下而上解决实际的数据问题，来推进顺丰数据治理各个领域能力的建设。**
2. **自上而下，统筹规划数据治理体系建设。单纯的自下而上，数据治理工作，便成了构建解决数据问题的产品，各个产品间应用的关联度不高。因此，需要自上而下的统筹规划的工作模式。**

**▌顺丰主数据治理分享**

**顺丰在主数据管理方面做的工作，主要做了4步：**

**第一，识别主数据，明确主数据的业务数据Owner, 业务科技Owner, 确定下责任。**

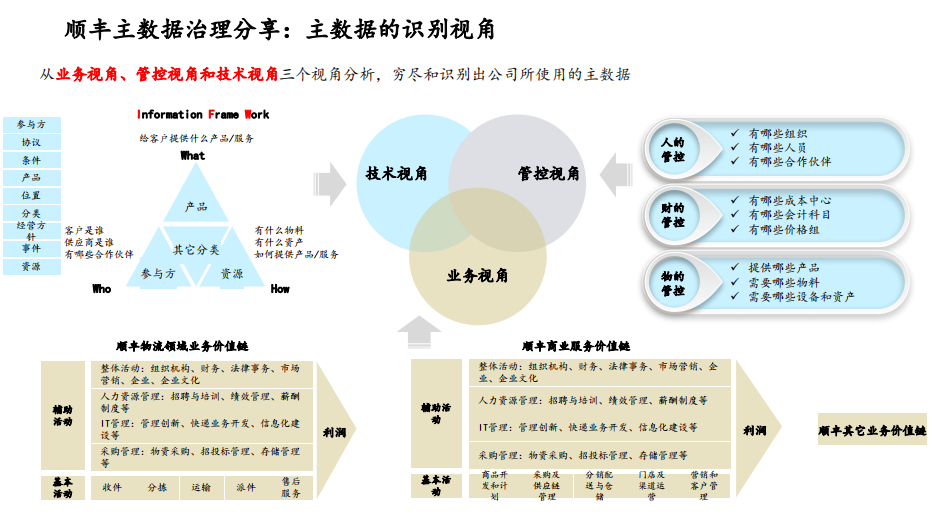
**第二，梳理主数据的属性，制定主数据的属性标准。**

**第三，确定可信的业务源系统，客户主数据，在顺丰叫CDM，合同主数据，在顺丰叫CMDM，供应商主数据，在顺丰叫SRM。**

**第四，我们对业务源系统，标准的落地、信息录入的管控、数据服务的提供，进行质量监控。同时，质量平台针对相应的数据做了质量监控。**

**基本上做完这4步之后，主数据的质量就能得到显著的提升。**

**顺丰的主数据治理：**

****

1. **主数据的识别视角和管理范围**

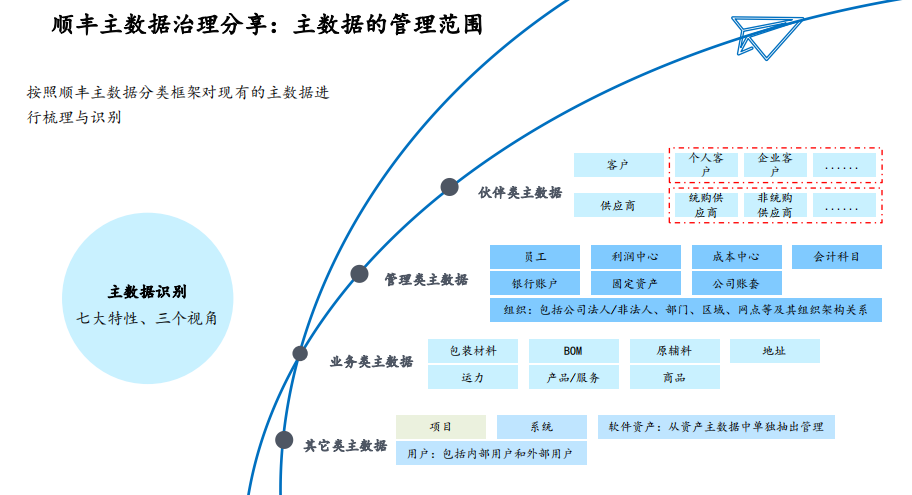
**要确定主数据的Owner，首先要识别主数据的范围。顺丰从业务视角、管控视角、技术视角，三个视角来识别顺丰集团的主数据。**

**从技术视角，参考IBM提出的企业数据管理模型，指从参与方、协议、条件、位置、分类、产品等方面确定企业数据划分的主题域。**

**从管控视角，指从人、财、物的管控三个方面来确定企业数据划分的主题域。人的管控如有哪些人员哪些组织，财的管控如成本中心会计科目等，物的管控指提供哪些产品、有哪些物料、提供了哪些服务。**

**从业务视角看，指从公司的多个业务领域的价值链来确定主数据。顺丰有多个BU，如物流领域业务价值链，基于其基本活动收件、分拣、运输、派件和售后等，及相应的辅助活动的人资、IT、采购、法务等确定主数据。另一个价值链，如商业服务价值链，基于活动商业的开发、采购、仓储、门店、配送、营销来确定主数据。**

**用以上3个视角穷举公司的主数据，识别出顺丰的主数据有。伙伴类的主数据：客户、供应商等；管理类的主数据：员工、利润中心、成本中心、会计科目；业务类的主数据：包装材料、BOM、辅料、商品、产品服务等，以及其他类的主数据，用户、资产、项目系统等。**

****

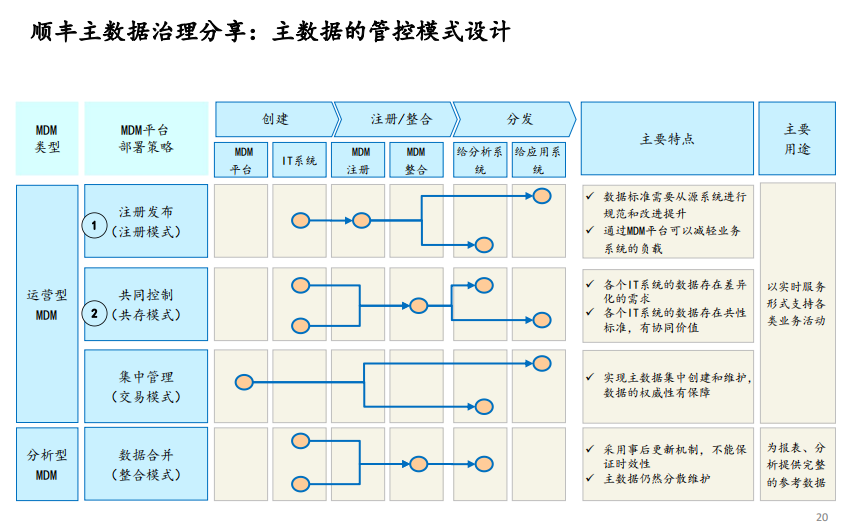
1. **主数据Owner及职责分工**

**我们针对这些主数据来进行业务Owner的确定，图中基于主数据分类，确定其业务数据Owner和业务科技Owner，划分明确的标准和职责。**

****

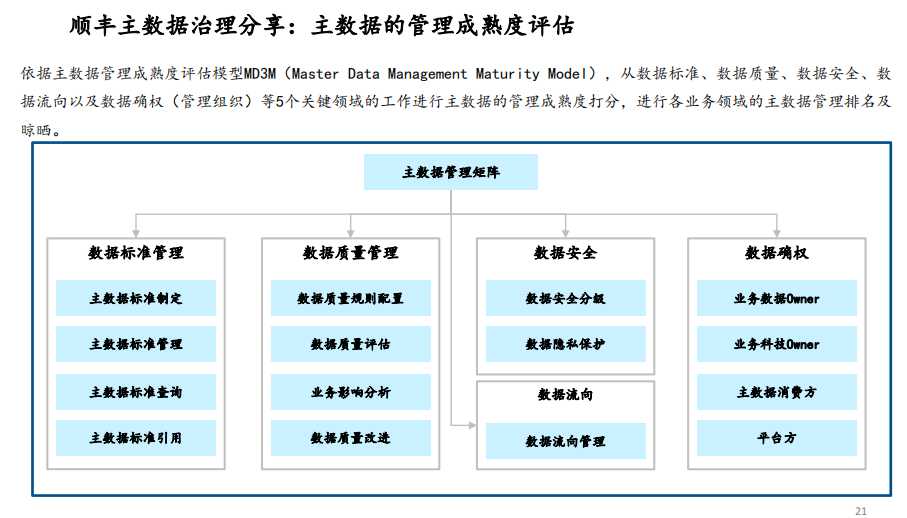
**3、主数据的管控模式设计**

**在主数据管理划分职责之后，对主数据的管控模式进行设计。在顺丰，主数据的管控模式，主要是注册发布式，和共同控制式两种模式。 注册发布式，用于有唯一的业务源系统的主数据的管理，像客户、员工数据，有唯一的业务源系统。客户主数据有CDM系统、员工主数据有HR系统，在业务源系统把数据的标准定好落地，把数据提供给MDM，给到主数据管理平台，然后我们给到下游400多家的系统使用。共同控制式，用于有多个业务源系统的主数据的管理。像合同数据，我们有销售类合同、非销售类合同、成本的合同、销售类的合同等，合同分散在不同的系统中，它在主数据管理平台上，首先进行统一标准的整合和标准的Mapping，然后再提供给下游的系统使用。顺丰目前主要使用以上两种模式。除了以上两种模式外，还有集中管理、数据合并。如华润万家的门店主数据管理使用了集中管理式。**

****

**4、主数据的管理成熟度评估**

**针对主数据的管理成熟度评估，从数据标准管理、质量管理、数据安全、数据流量、数据确权5个方面进行。如标准管理包含标准制定、标准管理、标准查询、标准引用，质量管理包含质量评估、影响分析等。评估即判断某个主数据在这5个方面各部分做到哪些，基于评价体系，对其打分。 比如客户主数据，有明确的业务数据Owner，有业务科技Owner，有平台方，做了主数据属性的安全分级和隐私保护，也做质量评估，会有一个整体的晾晒排名。**

****

**通过以上几个部分，完整主数据管理的工作。**

**今天的分享主要是以上这些，介绍了顺丰的数据治理体系，各领域的一些能力，以及主数据的实践。**