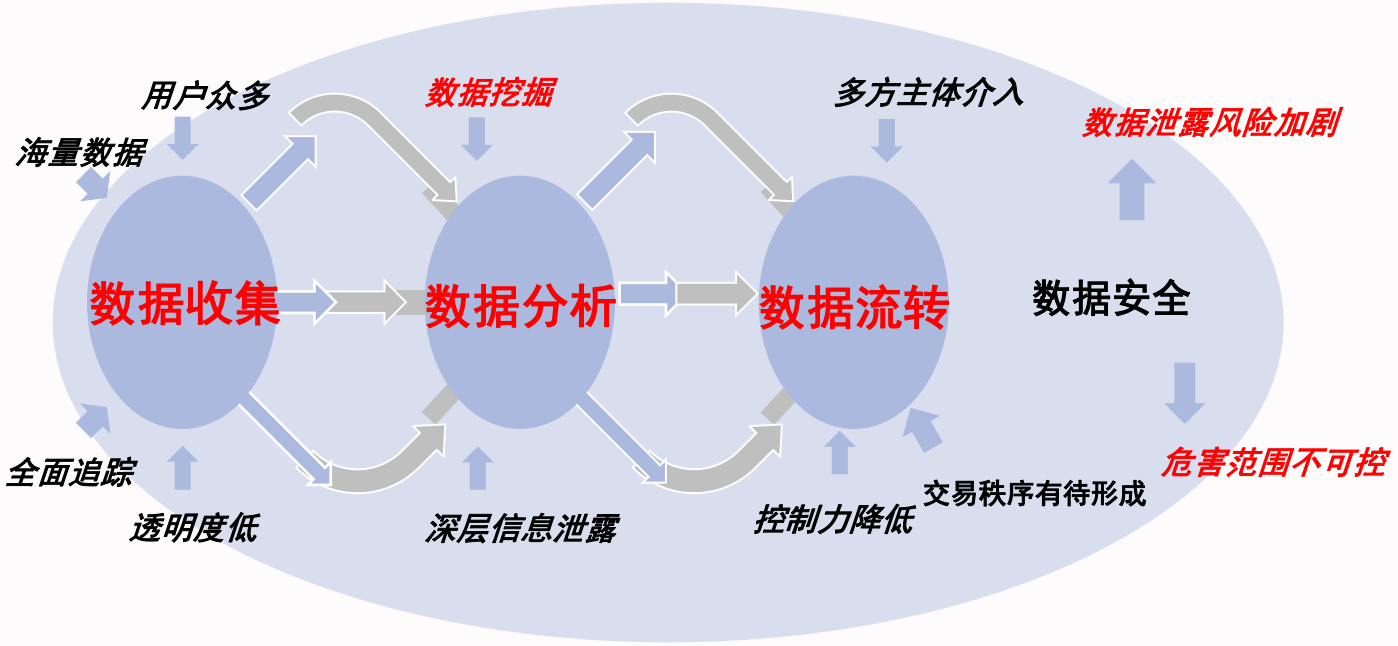
BDIC 2017 BIG DATA INDUSTRY CONFERENCE 2017

电信大数据标准及合规性思考

中国信息通信研究院高级工程师
大数据发展促进委员会办公室主任
韩涵

大数据环境下个人数据保护面临新的挑战

大数据技术的广泛应用为个人数据保护带来**新**的挑战，集中体现在**数据收集**、**数据分析**、**数据流转**等环节，及增大**数据安全**及隐私侵害**风险**等方面。



匿名化——隐私保护的关键问题之一

1、《网络安全法》：网络运营者应当建立健全用户信息保密制度，对其收集的用户信息必须严格保密。网络运营者不得泄露、篡改、毁损其收集的公民个人信息；未经被采集者同意，不得向他人提供公民个人信息。但是，经过处理无法识别特定个人且不能复原的除外。

2、《中国互联网定向广告用户信息保护行业框架标准》
实现用户身份关联信息的去身份化，即使得该信息无法用于识别、确认或关联至某个特定用户。

3、《WP29》：如果企业对其拥有的个人数据进行匿名化处理，以至于通过该数据已无法识别出个人身份。

匿名化的常用处理方法

删除：• URLs清洁化•任何精确地理位置信息；•终端用户特有的日期信息•用户年龄•直接联系信息•身份证号或其他政府授予的身份识别信息• 生物识别信息

映射：将身份关联信息替换成非身份信息后再进行常规化数据存储和计算分析

隐藏：修改从统计角度而言范围狭窄的准身份识别数据值，或将数据归入某个更宽的范畴内。

合并：将可能有助于识别个人身份的小数值范畴重新分类。例如，将18-24岁的值与24-30岁的值合并为一个18-30岁的范畴。

衍生：综合数据是从原始数据中生成的，然后替代原始数据，但同时保留了原始数据的价值

数据分类——隐私保护的关键问题之二

长期困扰着数据流通行业从业人员和监管人员的困惑：

是否需要对所有信息设定相同的隐私保护边界？

你身份证号码这个信息受到的保护程度，是否可以与你喜不喜欢打棒球这个信息受到的保护程度不一样？

CAICT 中国信通院

数据分类——隐私保护的关键问题之二

用户使用互联网公司提供的网络信息服务所产生的数据信息，进行匿名化处理后产生的新的数据信息，其权属如何界定？

个人数据兼具人格权与财产权双重属性，同时具有价值和使用价值。数据权属是数据流通和数据产业化的逻辑起点。数据资产所有权的归属决定这数据价值利益的分配及对数据质量、安全责任的划分。

两种不同的观点

个人对数据拥有绝对的财产权，个人产生的数据理所当然归个人所有。无论数据如何被流转或利用产生的新数据，其所有权都归属个人。

用户使用互联网企业提供的服务所产生的数据属于企业所有。谁投资，谁所有。谁记录，谁所有。

数据分类——隐私保护的关键问题之二

新思路：将数据区分为基础数据和增值数据，二者对应不同权属。
建议采取合理识别范围。能够识别出特定身份的数据构成基础数据，反之不构成。

能识别！基础数据！

不能识别！增值数据！

知情同意——隐私保护的关键问题之三

知情同意原则是指信息管理者在收集个人信息之时，应当充分告知信息主体有关个人信息被收集、处理和利用的情况，并征得信息主体明确的同意。

知情同意原则已经成为我国个人信息保护的一项重要原则。

知情同意是个人信息权的核心，是最能够体现个人价值的原则，信息人本人的知情同意是对信息进行收集、处理和使用的基础，没有当事人的知情同意，除非法律强制规定的情况以外，任何的收集行为都是没有合法性基础的。

知情同意原则的拆解

本人同意　书面同意　明示同意

内容明确　合理及必要

欧盟最新制定的《统一数据保护条例》（GDPR）规定：

- 基于数据主体自由意志作出；
- 书面声明中的同意应当与其他事项明显区分；
- 书面同意必须明确；
- 同意可以被撤销，撤销不具有溯及力；
- 可能包含默示同意；
- 对于不满16周岁的儿童的数据处理行为须获得有监护权的人的同意。

大数据时代对知情同意原则的冲击

- 个人信息范围扩大
- 企业经营中信息收集行为日常化
- 从注重信息间的因果关系到关联关系的转变
- 信息匿名化难度加大
- 信息滥用可能性加大

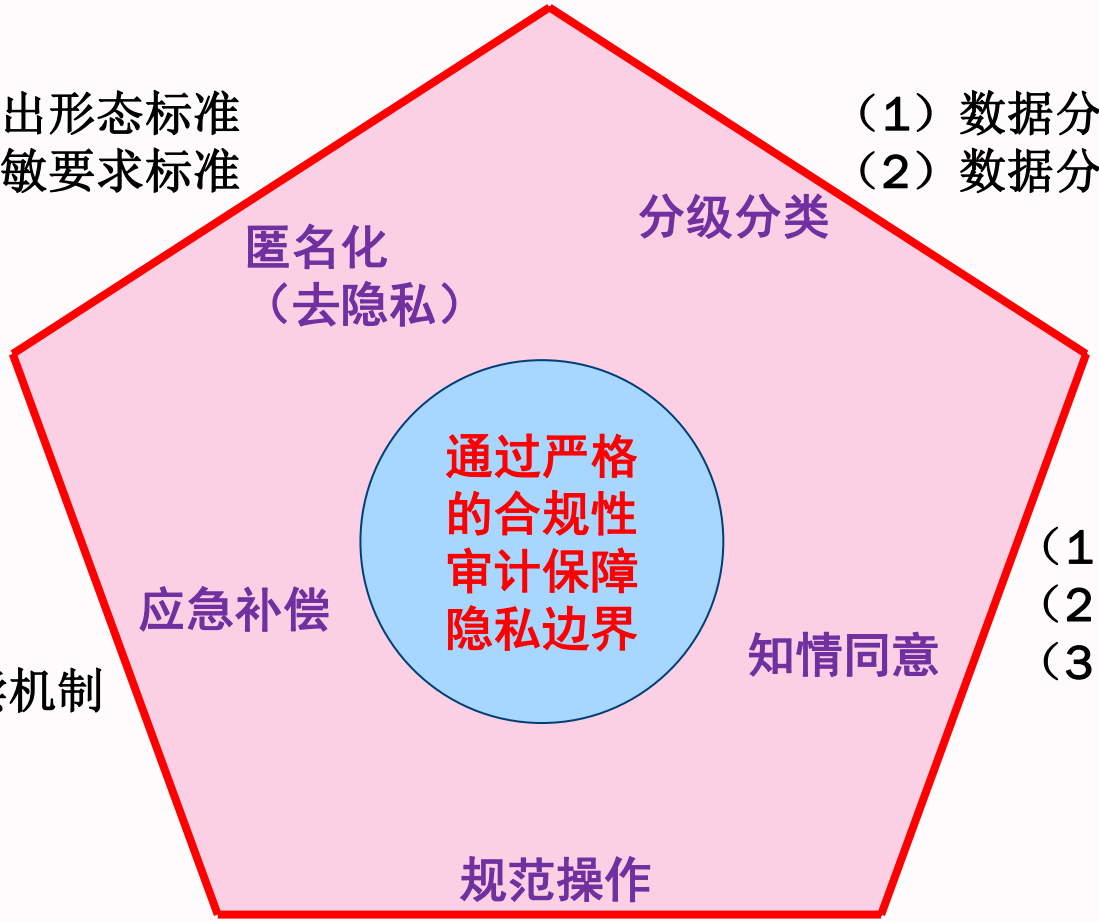
解决思路

- **格式合同？**
- **区分敏感信息与一般信息**
- **由事前预防转变为事中、事后的监管**
- **完善风险评估机制，健全标准认证**

保护数据安全和个人隐私的5条边界

（1）数据输出形态标准
（2）数据脱敏要求标准

（1）数据分级标准
（2）数据分级管理要求



（1）应急机制
（2）风险损失赔偿机制

（1）授权分类
（2）授权流程
（3）标准合同文本

（1）安全管理办法
（2）安全组织机制
（3）安全操作规范

数据中心联盟大数据委员会标准体系

大数据产品能力认证标准包括：

- 《Hadoop平台基础能力测试方法1.0》
- 《MPP数据库基础能力测试方法1.0》
- 《Hadoop平台性能测试方法1.0》

数据流通应用合规性标准包括：

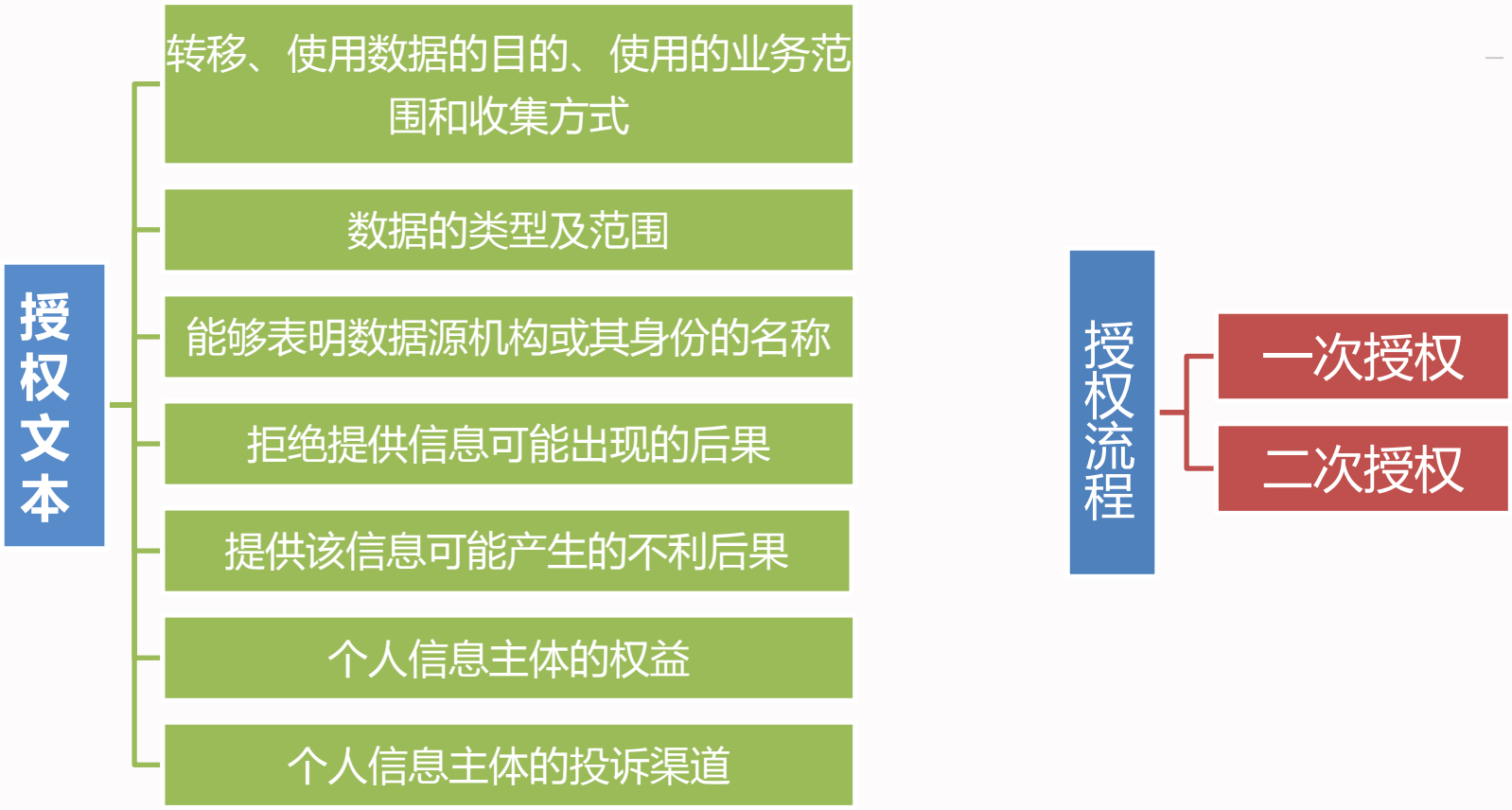
- 《数据流通标准：通用标准》
- 《数据流通标准：征信类产品及服务标准》
- 《数据流通标准：金融风控类产品及服务标准》
- 《数据流通标准：数据流通中心数据处理及业务管理标准》
- 《数据流通标准：精准营销类产品及服务标准》
- 《数据流通标准：区块链应用场景标准》
- 《数据流通标准：位置信息类产品及服务标准》

数据流通合规性评估主要指标——安全管理

安全管理机制	安全组织机制	安全操作机制
• 完整的数据安全管理文件 • 周期性的安全评审管理机制 • 明确的安全相关部门的管理职责 • 修订数据安全管理办法的机制	• 人员背景审查机制 • 人员访问权限控制 • 保密机制	• 敏感数据的加密机制 • 数据系统访问控制机制 • 操作监控机制 • 安全事件预警及处理机制

审查方式：文档相关 内容审查、制度和流程审查、操作机制说 明及系统查看

数据流通合规性评估主要指标——用户授权



审查方式：授权文档 查看、授权流程说明 和审查

数据流通合规性评估主要指标——数据使用

禁止对外转移的数据：法律法规禁止转移数据列表

安全方式下可对外转移的数据

数据源机构可以用安全的软件接口，在取得用户授权的情况下，将经过严格去 隐私化的数据，直接向数据服务机构、第三方支付机构等合格的数据接受者转移， 直接转移的数据包含且只包含以下几种形式：

以“是否”等二元方式返回的查询类数据；

以“分级”等方式返回的查询类数据，且分级数量不大于10；

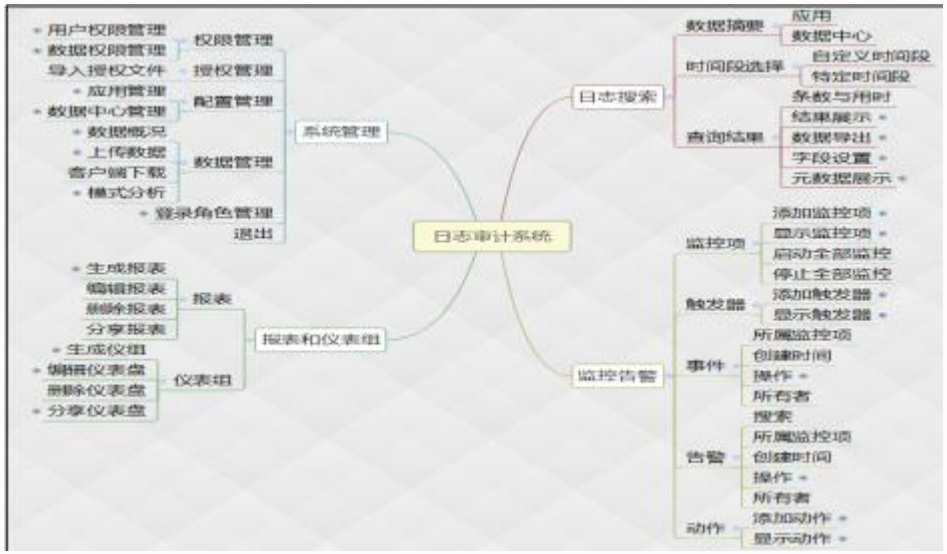
以“分数”等方式返回的经过加工计算后的信息，且确保无法还原出原始信息。

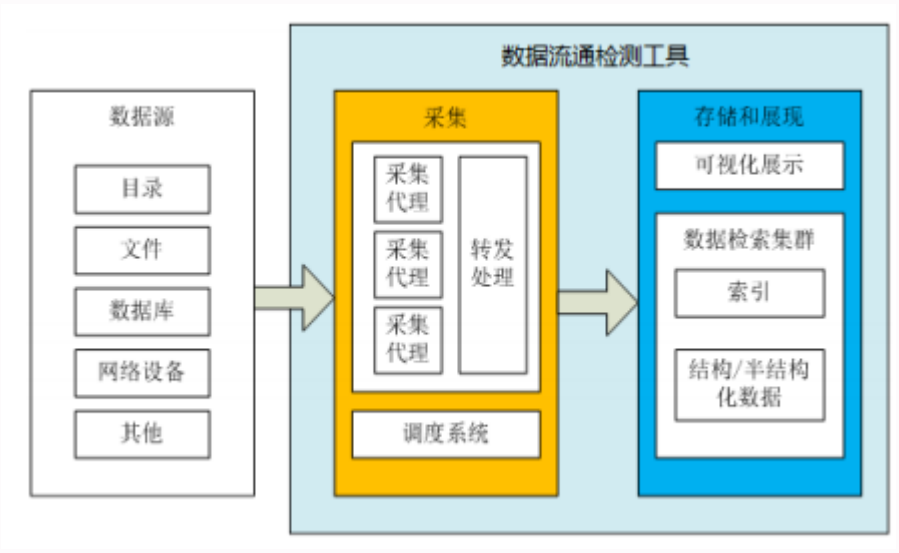
安全监管环境中可以使用的数据

数据源机构所管理的用户数据，除第一类数据外，并且不以第二类数据转移方式输出的数据，则必须在安全、可控、被监管的环境中才可以输出，大量在电信业务中产生的原始数据以及粗加工后的数据属于这一范畴。

审查方式：测试工具审查

测试工具描述





（1）数据通过前端部署的探针采集到的数据上传到服务器，也可以通过HTTP POST上传数据文件；

（2）数据接收 数据上传到系统的中央接收器，进入缓冲队列，等候处理。

（3）字段抽取 系统通过系统自带或用户配置的规则引擎解析数据，抽取关键字段，将非结构化的数据转化为结构化数据

（4）索引 对关键字段及数据全文做分布式索引，索引文件方便用户对关键字段或全文进行检索

（5）搜索分析 数据进入索引文件之后，用户可以像使用搜索引擎一样进行数据搜索，查找满足特定条件的数据。

（6）监控告警 根据用户预设的搜索告警条件自动检索数据，查看搜索结果是否满足预设告警条件，如果触发告警，则通过邮件或短信通知 用户。（7）统计分析、数据可视化 对搜索结果进行统计分析和可视化，展示时间折线图、条形图、饼状图等，让数据分析更直观。

谢谢大家！

韩涵
hanhan@ritt.cn
13811249349