

电信大数据应用创新与合规性

魏凯 中国信息通信研究院

提纲

- 电信大数据资源具有独特价值
- 电信大数据应用的合规性要点

大数据生态体系



电信大数据具有独特的应用价值

数据类别	数据字段	与互联网公司对比			变现价值	典型应用	应用分析
		独有性	完整性	准确度			
人口统计数据	姓名、性别、号码、身份证号、IMEI等	高	高	高	低	统计报告类	基础数据，单独变现价值很低
消费行为数据	套餐选择数据、资费数据、缴费记录等	均等	高	均等	中	征信类	数据可用性高，但价值有限
位置信息数据	地理位置、基站经纬度、停留时间等	低	高	低	高	客流分析等	数据价值高，但精度低，易被替代，
上网行为数据	连接方式、访问记录、关键词等	高	低	低	高	精准广告投放等	数据价值极高，但采集、清洗难度大

电信大数据应用场景

- ρ 内部大数据应用场景设计：客户挽留、客户迁移、精确营销、客户服务提升等。
- ρ 外部大数据应用场景设计考虑因素：现有基础、实现难度、发展前景等。

近期应用场景设计

- ρ 现有场景优化
 - 智能出行提醒（位置信息数据）
 - 交通人流分析（位置信息数据）
- ρ 新场景设计
 - 银行信贷分析（消费行为数据）
 - 商业辅助选址（位置信息数据）
 - 数据分析报告（综合数据）
 - 景区客流分析（位置信息数据）

中期应用场景设计

- 广告精准投放（上网行为数据）
- 公共安全管理（上网行为数据、地理位置数据等）
- 辅助廉政建设（上网行为数据）
- 整体解决方案

远期应用场景设计

- 能力开放平台
- 数据交易平台
- 广告撮合平台
- 产业生态系统构建
 - ✓ 智慧医疗
 - ✓ 智慧城管
 - ✓ 平安城市
 - ✓ 智慧教育
 - ✓

交通与人口迁移应用

重点利用位置数据，提供交通拥堵实施监控、人流实施监控等辅助管理服务。

交通实时监控

客户需求：通过信令数据为广州交委提供广州火车站、火车东站等特定目标区域内的实时人流量统计、辅助进行应急指挥。

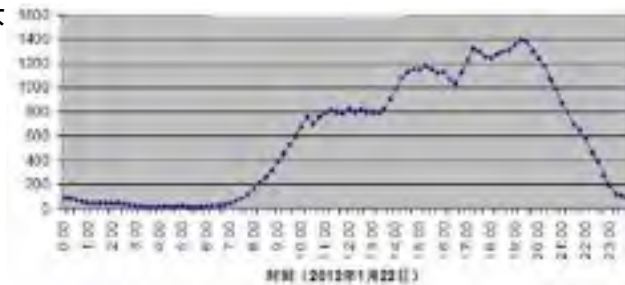
热点区域交通拥堵分析



人流实时分析

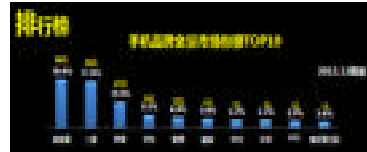
客户需求：深圳市发改委区域经济处牵头组建“全口径人口信息集成应用系统”，基于信令数据，计算全市人口数量、分布、流动等。

XX社区人口变化曲线图



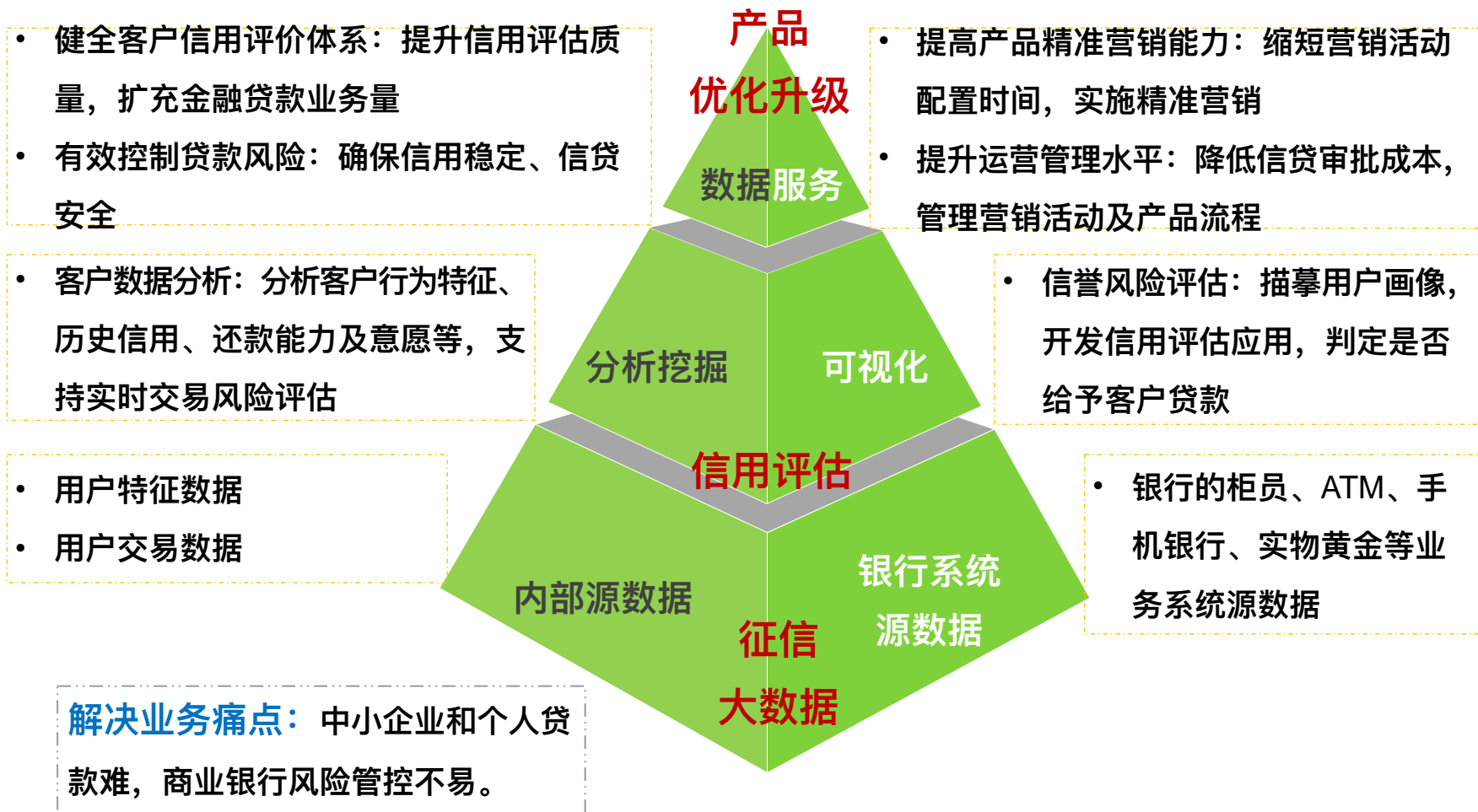
数据报告

- 《人口流动研究报告》
- 《人口分布研究报告》
- 《公共交通研究报告》
- 《人口特征研究报告》



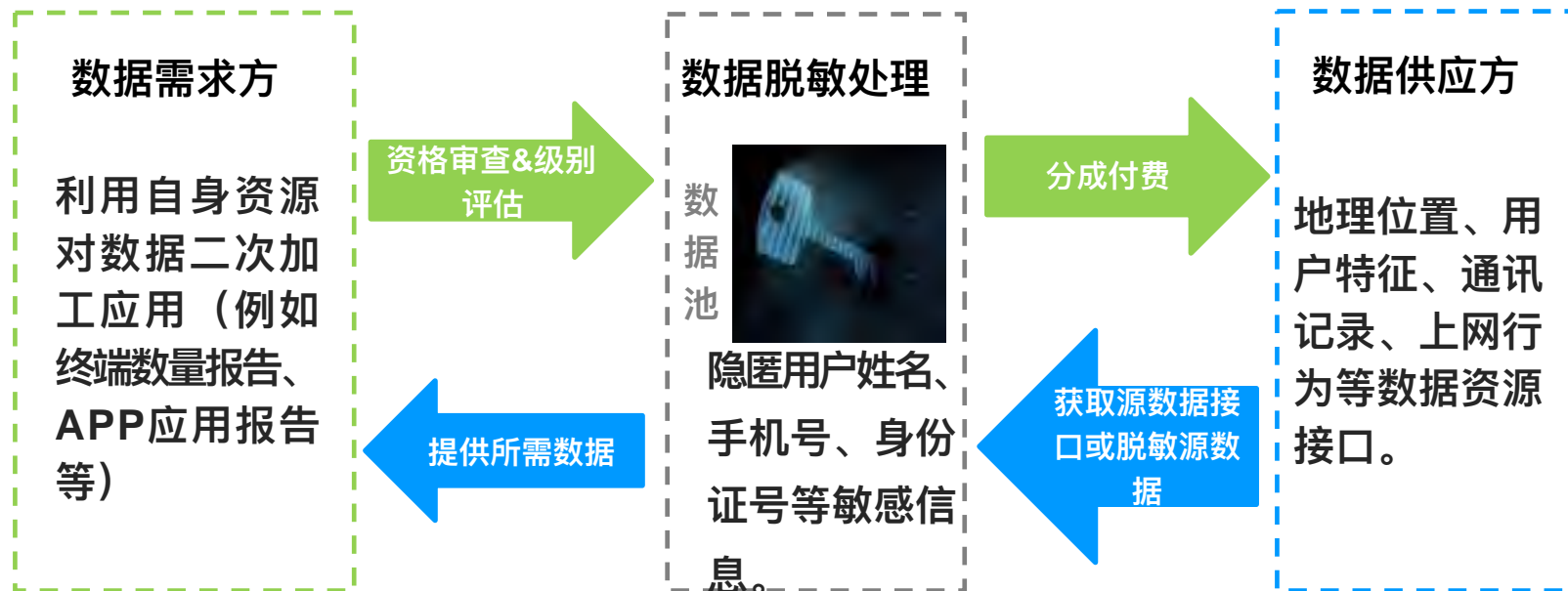
征信与金融风控应用

综合分析、挖掘电信和金融双方数据，及时把握客户行为并进行信用评估。



统计数据分析报告应用

在严格执行数据保护的前提下，充分挖掘数据资产价值，以数据报告的方式，向第三方提供源数据供应服务。



功能模块

APP使用分析

网站使用分析

区域分析

移动设备分析

搜索引擎分析

定制报告

分析指标

用户数

通讯次数

通讯时长

用户性别比例

年龄段分布比例

.....

城市规划与商圈分析应用

抓取、分析、挖掘内部及各类互联网监控数据，掌握目标地区及周边用户特征，准确、高效地实现预期利益最大化的商业选址及服务内容决策。



周边用户特征数据

- 周边用户特征分类数据
- 人流量、交通运行数据
-



分析挖掘与可视化

- **基于GIS技术**：城市区域栅格化处理，分析每个栅格特征数据，挖掘店铺选址决策依据；
- **结合可视化技术**：实时展示店铺周边人群关注内容等情况。



支撑商业选址决策

- **选址分析与推荐**：选址分析、优劣对比及最优化推荐；
- **其他数据服务**：可提供商圈分类、位置管理、客户挖掘、竞品分析等支持。

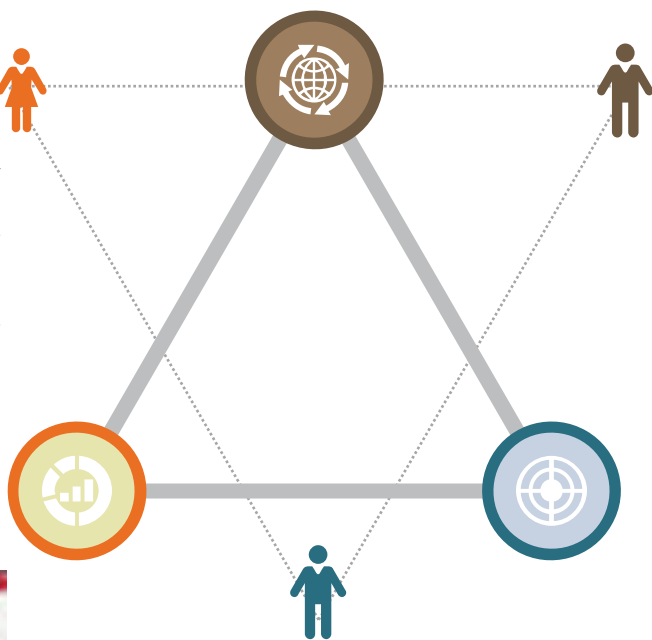


智慧旅游景区分析应用

通过景区客流信息实时监控分析以及对游客客源的分析统计，加强景区服务和管控，为旅游信息化建设及市场推广提供有效的数据支撑。

景区实时客流量监控GIS

- 显示景区内各区域实时游客人数，人员密度根据游客舒适度进行颜色标注，在系统顶部以消息滚动条形式实时播报景区内游客总量和拥挤区域信息。同时手机客户端将推送拥堵信息给景区管理员，提醒景区管理人员做好游客疏导工作。



游客特征及趋势分析

- 利用手机网络信令数据和用户数据结合，为旅游景区提供游客特征分析、游客消费分析、客流趋势历史对比、游客精确营销等功能，满足景区的游客洞察与营销需求，改善景区的营销与管理水平。



广告精准投放应用

通过综合分析集团及各类互联网在线数据，实现跨平台、跨终端广告投放。

用户行为特征数据

- 自然属性：性别、年龄、收入、职业等；
- 通信记录：通信时间、IMEI、GPS定位、社会关系网等；
- 行为特征：到访页、访问次数、浏览主题词、搜索关键词等；
- 消费记录：购物历史、交易额、计费代码、消费频次等；
- 偏好特征：终端偏好、APP偏好、长期兴趣爱好等。

大数据处理

- 数据分析：基于数据分析技术，形成消费者画像，建立客户挖掘模型，实时追踪处理；
- 广告推送：依托统一账号等身份识别方式实现跨平台、跨终端广告推送。
 - ✓ 跨平台：门户网站、搜索引擎、垂直电商、视频网站、社交网络等；
 - ✓ 跨终端：PC、平板、智能手机、互联网电视、穿戴式设备等。

实时广告投放

- 实现机制：互联网实时广告平台、APP推荐优惠券、多方广告合作机制等；
- 营销效果：降低成本，提升转化率，实现广告精准投放等；
- 预期收益：有效提升碎片化流量变现能力，挖掘潜在用户群产生实际购买，实现以数据换流量、换收入，按照收看效果收取费用。

公共安全管理应用

基于手机定位技术和用户通信行为，进行针对性数据挖掘和分析，实时监控公共空间各类活动，及时识别犯罪行为和重大安全隐患。

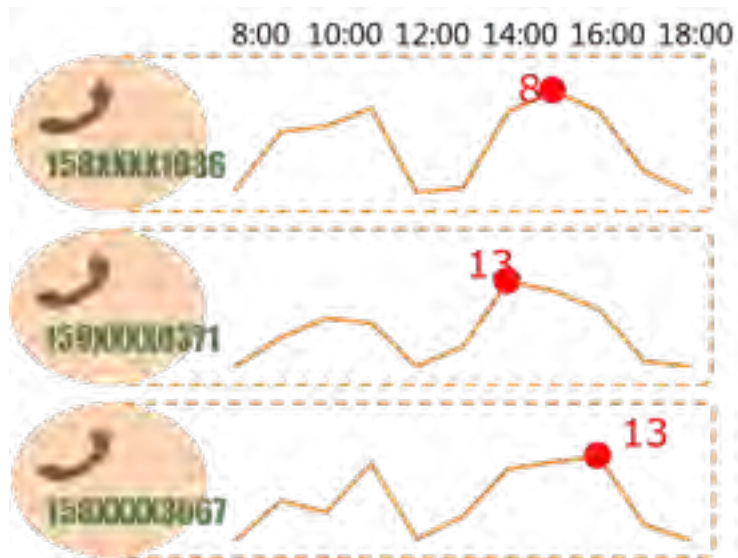
犯罪行为识别

- 通过与公安部门合作，挖掘团伙传销、诈骗电话、群体性上访、涉暴安防等特征，建立相关识别模型及其监控应用产品。



公共安全监控

- 通过对城市热点、交通枢纽的实时人流量分析，发现区域人群活动规律，便于突发事件的快速处理；在重大活动、重大事件发生前可以为预案准备提供参考；可以了解大型集会活动人员的聚集、疏散情况，便于突发事件处理。

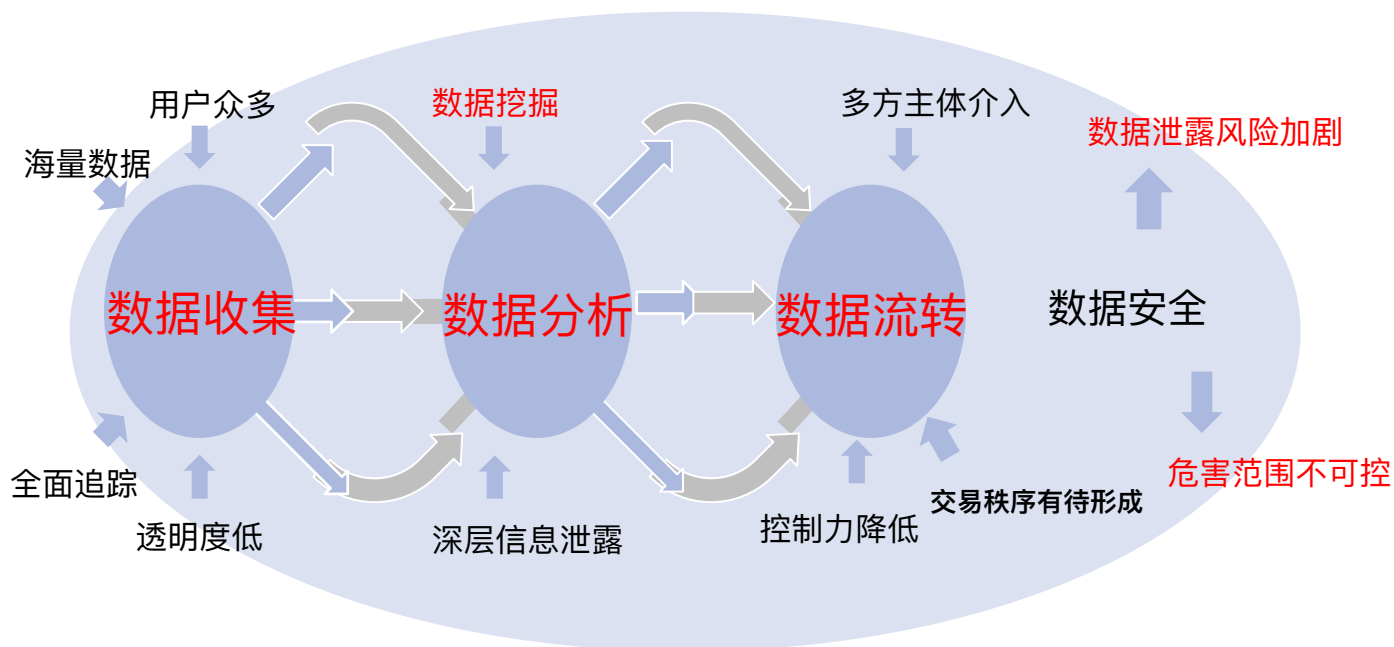


提纲

- 电信大数据资源具有独特价值
- 电信大数据应用的合规性要点

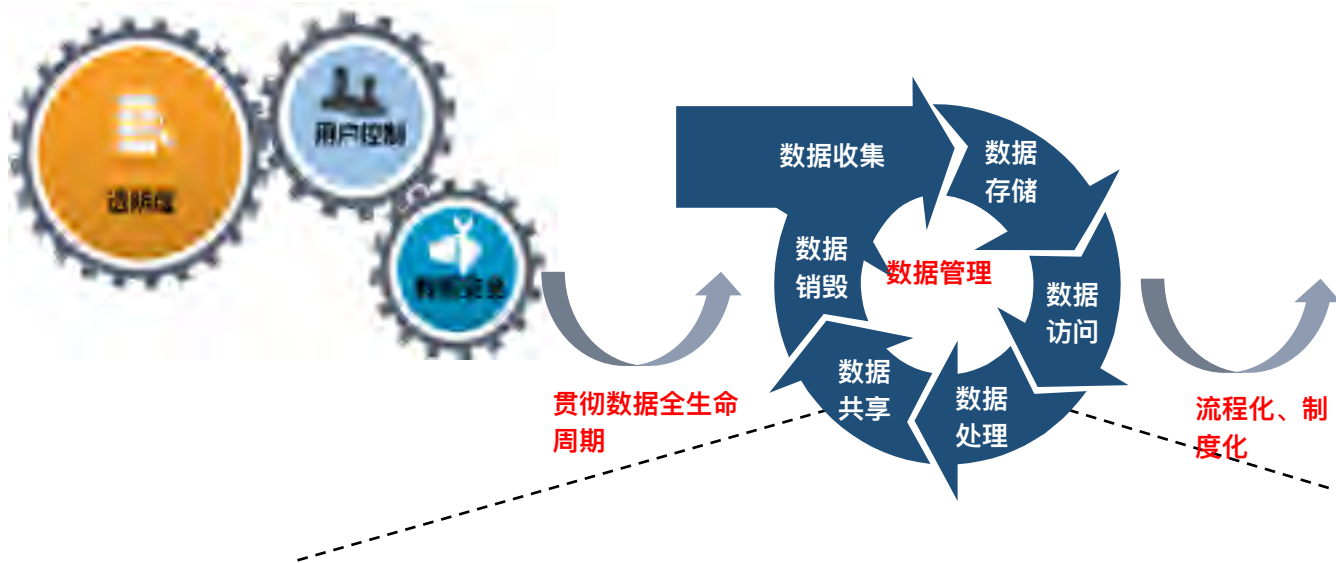
大数据环境下个人数据保护面临新的挑战

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。



保护好个人信息的安全是电信大数据应用的最大前提！！

个人信息保护的全生命周期管理

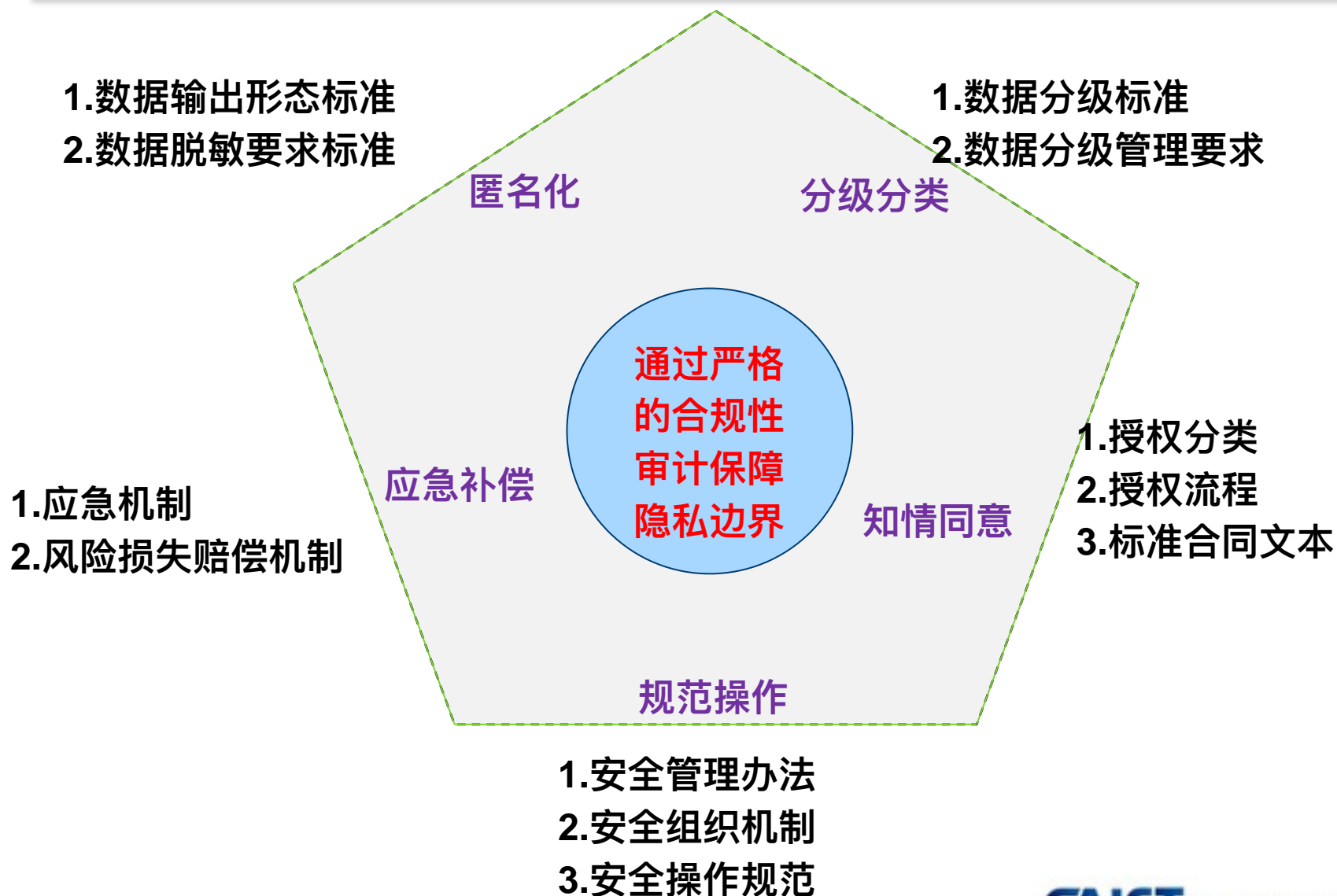


- 制度化的数据保护策略
- 规范化的数据操作流程
- 成熟的安全事件应急机制
- 严格的合规和审计制度
- 由点到面的培训课程开发

数据收集	数据存储	数据访问	数据处理	数据共享	数据销毁
<ul style="list-style-type: none"> • 公示数据收集、处理情况，以使用户知情并作出明智的选择。 • 传输保护 	<ul style="list-style-type: none"> • 加密安全技术 • 防泄漏等技术保障+应急预案 • 权限管理 • 备份、存储隔离 • 介质管理 	<ul style="list-style-type: none"> • 访问控制 • 身份验证 • 权限分配 • 审批管理 	<ul style="list-style-type: none"> • 权限分配 • 审批管理 • 分类分级使用和保护 • 隐私增强技术保护 • 匿名化技术保护 	<ul style="list-style-type: none"> • 数据责任 • 第三方保护水平审查 • 第三方用途限制 	<ul style="list-style-type: none"> • 达到目的 • 遵守法律规定

隐私影响评估、合规性审核、审计、培训、安全事故应急响应

电信大数据应用合规性框架



关键点1：数据需求合规性

数据安全四大原则：敏感数据不出门、数据服务有时限、避免手工实操作、建立系统级平台

- 数据服务不提供含敏感数据的产品
- 按照服务需求分批分次提供数据，定期销毁数据和相关载体。
- 数据加工、提供尽可能在系统上实现
- 建立数据安全保障系统架构，建立涵盖需求审核、数据采集、存储加工、数据服务各个环节的系统保障平台

数据需求应遵循“最小集合”原则

每种应用场景对于数据的采集和使用不应超出合理的范围。应事前组织合规性审查。

关键点2：用户授权合规性

用户授权是数据服务的前提，多种授权方式具有同等效力

为保障用户的知情权、选择权，在收集、转移和使用用户个人信息时须征得用户同意。获取用户授权可以采用签署纸质授权协议的方式，也可以采用短信、网页、APP等电子授权方式以及借助技术授权实现远程人像授权等新方式。

授权信息应进行长期保存及企业间传递

考虑到信息安全和法律责任要求，每次数据请求均需要授权并将授权记录进行保存，确保授权信息可追溯，建议存储期限不少于1年。建议授权信息传递给合作企业以备争议时追溯。

尽量保持授权协议内容完善性

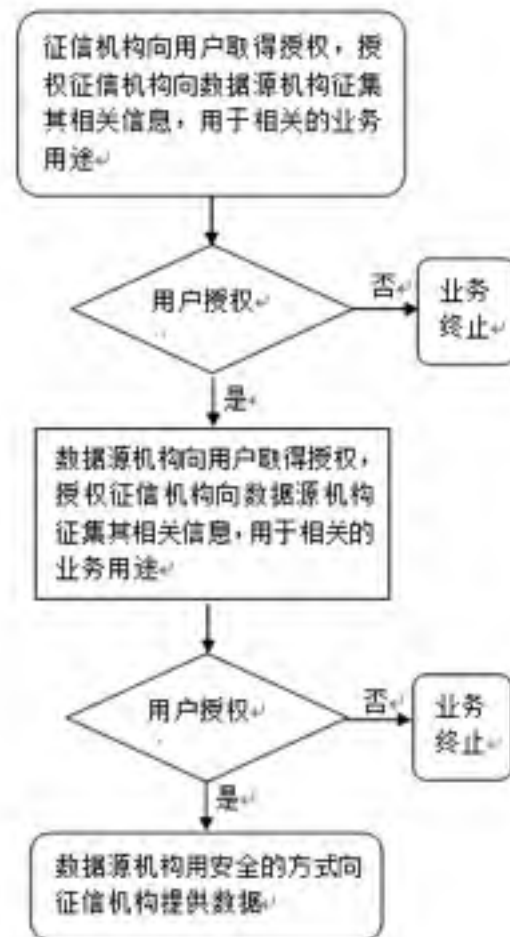
转移、使用数据的目的、使用的业务范围和收集方式；数据的类型及范围；能够表明数据源机构或其身份的名称；拒绝提供信息可能出现的后果；提供该信息可能产生的不利后果；个人信息主体的权利；个人信息主体的投诉渠道。

关键点2：用户授权合规性

特别敏感信息建议电信企业取得直接授权

例如，以下信息类型的原始数据

- 用户姓名、身份证件号码、工作单位、家庭住址等用户基本资料的原始数据；
- 用户电话号码、服务账号、IP地址、邮箱地址、服务涉及的密码、口令等非交易类用户身份标识和鉴权信息的原始数据；
- 通讯录、好友列表、群列表、联系人的通话记录等涉及其他个体信息的原始数据；（涉及其他个体授权）
- 用户所在经纬度、地区代码、小区代码、基站号等用户位置信息的原始数据。



关键点3：数据形态及转移方式合规性

法律法规、公开标准禁止转移的数据

例如，以下类型的数据集

- 揭示个人种族、宗教信仰有关的信息；
- 揭示个人疾病和病史等与个人健康状况有关的信息以及与个人性生活有关的信息；
- 基因、指纹、血型等揭示个人生理标识的信息；
- 交易类服务中的账号、密码、密码保护答案等交易类服务身份标识和鉴权信息；

建议只有严格脱敏的数据集可以直接对外提供

严格脱敏使得指向用户个体的信息变成统计类信息，例如以下类型的数据集

- 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5；
- 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。

关键点3：数据形态及转移方式合规性

多方可信计算环境可用于解决企业间数据流通

多方可信的计算环境即经数据输入、输出及相关审查机构等多方确认及认可，且在数据的收集、存储、计算、传输等各方面不会出现因人为、设备等因素导致的数据丢失、修改、泄露等情况的可以信任及信赖的计算环境。

- 多方可信计算环境应由数据输入、输出各方共同监管，取得多方的共同认可；
- 征信机构等数据使用方从多方可信计算环境中读取数据时，应采用严格的鉴权方式，且经过权限审计；
- 数据传输、处理等过程的操作日志、数据表等应进行存储以备审计，建议期限不小于6个月；
- 多方可信计算环境的输出数据只以“是否”、“分级”或“分数”的形式对外提供；
- 多方可信计算环境的输入/输出数据将进行人工和机器审核；
- 多方可信计算环境的运行/管理规则将进行人工审核；

电信大数据在征信中的应用规范

数据中心联盟/中国通信标准化协会，中国信息通信研究院联合三大运营商+征信企业
电信大数据在征信中的应用场景



2 授权方式及授权流程

- 授权协议要求
- 需明示授权的个人信息范围及要求
- 授权流程

3 数据形态及转移方式

- 禁止对外转移的数据
- 安全方式下可直接对外转移的数据
- 安全监管环境中可以使用的数据

4 数据使用及审计方法

- 总则
- 多方可信计算环境
- 审计方法

5 数据交换接口

6 数据服务定价方法

7 安全保障机制

通过行业自律建立大数据信任体系

中央网信办、工业和信息化部指导，数据中心联盟（DCA）牵头，中国信息通信研究院发起，各数据交易所、互联网企业和电信运营商等54家企业共同签署

2016大数据产业峰会（4月27日）



1. 数据流通的定义

2. 企业对数据的权益

3. 流通环节风险评估

4. 企业增强社会责任

5. 个人、企业利益协调

6. 第三方平台责任义务

7. 技术标准与审计认证

8. 监督与惩戒

9. 持续完善更新

最佳实践——行业自律——法律法规

下一步：与签署单位一起，建立信息披露和合规性审计机制

大数据发展促进委员会

China Big Data Council (CBDC)



联合政、产、学、用各环节，解决产业发展共
性问题，鼎力支撑国家大数据战略落地

打破数
据孤岛

推动技
术成熟

探索应
用模式

(包括但不限于)

- 建立数据开放和流通行业规则，推动行业自律；
- 制定大数据产品标准，建立大数据产品能力认证体系；
- 建立大数据人才认证体系，提供大数据人才认证服务；
- 推动金融、医疗、政府等行业大数据应用发展；
- 开展大数据发展关键问题研究，提供相关政策建议。

谢谢

BDTC 2016中国大数据技术大会
Big Data Technology Conference 2016