分域匿名化实施服务--基本原理

GBT-35273-2017(2020)《**个人信息安全规范**》
GBT 37964-2019《**个人信息去标识化指南**》
《**数据安全技术 个人信息匿名化处理指南及评价方法（征求意见稿）**》

T/CAAAD 004-2022 T/CCSA 424-2022
《**互联网广告 匿名化实施指南**》

① **技术**-选择适当的数据去标识化技术模式，构成**数据**的匿名；

② **法律**-结合场景和条件的评估与凭据机制，清晰**行为**的边界；

③ **管理**-配套合约执行过程监控等运营措施，控制**主体**的使用；

$$y = F(x(ID))$$

0 **ID**

个人识别信息：
遵循数据库范式标准，包括标识与标签。
区分“直接标识符/间接标识符”

1 **x ()**

数据去标识化：
静态下，其他人不可识别，也不能复原。
防止“逆向破解/正向彩虹”

2 **F ()**

环境已知可控：
动态中，限定行为边界、控制行为过程。
达成“排除未知/控制关联”

3 **y**

y = 符合匿名化定义的证据集

① **x()，指技术处理过程，遵循数据库范式要求，包括标识与标签两种处理方法：**

- 标识处理，一种IDFV效果的假名技术，不可逆且互不相识；　（标识=直接识别符）
- 标签处理，遵循K匿名原则的泛化规则，多标签组合也不唯一；　（标签=间接识别符）

② **f()，指法律与管理组合的服务过程，梳理既有互联网广告场景，明确可做/不可做：**

- **鼓励非识别应用**，指无需在本方域内识别个体的业务形态；
  （如：本方结合第三方数据进行圈包后委托DSP投放的场景。圈包过程无需识别(非识别)即可开展，识别仅需在DSP侧的投放过程进行，而DSP是有权识别）
- **监管已识别应用**，指既有业务的同意范围已包含从他方收集数据项的场景。
  （如：从第三方补充数据标签的场景。仅需会员注册时，对采集的描述中有 “同意采集或从其他方收集XXXX信息…”相对宽泛笼统的的隐私条款即可开展。）
  （此时匿名化服务应具备匿名态下的“求交/关联/归因”功能，证明所补标签是遵守(未超)同意条款，且经关联而获得(非还原而来!)。
- **禁止可识别应用**，指无清晰授权，通过还原识别获取标识符的业务场景。
  （正常互联网营销场景，非识别和已识别两种模式都能覆盖了，无需可识别模式！！）
  （可识别应用，一般出现在电话外呼等特殊场景。需从数据中找出被叫号码后拨打。找出被叫号码就是“可识别”的过程。）