ICS 33.030
CCS M21

团 体 标 准

T/CCSA 548—2024 T/CAAAD 001-2024

互联网广告 数据流通平台技术架构

Internet advertising——Technical framework for data circulation platform

2024－07－03 发布 2024－10－01 实施

中国广告协会 中国通信标准化协会 发 布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国广告协会和中国通信标准化协会共同提出，并分别归口。

本文件起草单位：中国信息通信研究院、蚂蚁科技集团股份有限公司、国家广告研究院、北京勾正数据科技有限公司、北京风行在线技术有限公司、四川长虹电子控股集团有限公司、郑州信大捷安信息技术股份有限公司、百胜中国控股有限公司。

本文件主要起草人：杨正军、朱岩、杨阳、白晓媛、潘无穷、李婷婷、彭晋、李宏宇、顾明毅、马磊、潘冲、黄德俊、刘献伦、刘为华、邹文佐。

引　言

为适应信息通信发展对标准文件的需求，由中国通信标准化协会和中国广告协会共同组织制定本文件，推荐有关方面采用。有关对本文件的建议和意见，向中国通信标准化协会和中国广告协会反映。

数据作为生产要素，其市场化配置可以促进数据的社会化共享与流通利用，让更多的数据使用者能更好地获取数据和利用数据，以支撑科学研究、社会治理和商业决策。互联网广告业务是数据使用、加工、提供和委托处理密集的行业领域，在广告投放、程序化交易、广告归因等场景，均涉及到各类数据在不同机构间的提供、加工、传输、使用等处理行为。

互联网广告领域数据流通的情形多样，有必要制定标准，定义广告数据流通共性技术架构，明确数据流通的流程与实施过程，指导具体的生产实践。增强对互联网广告数据流通的安全管控能力，在确保数据安全的前提下，促进数据资源自由流通，促进互联网广告产业的安全、健康、快速发展。

互联网广告 数据流通平台技术架构

1 范围

本文件规定了互联网广告数据流通平台的基本原则、基本模型和技术架构。

本文件适用于互联网广告相关方在设计、开发和使用数据流通平台时参考。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 37964—2019 信息安全技术 个人信息去标识化指南

3 术语和定义

GB/T 37964-2019界定的术语和定义适用于本文件。

3.1

数据流通 data circulation

将数据供方的数据提供给数据需方使用的过程。

注：包括直接将数据供方的原始数据提供给数据需方使用，也包括对原始数据进行加工处理后提供给数据需方使用。

3.2

数据流通平台 data circulation platform

为数据流通提供各种数据服务的平台。

3.3

数据供方 data provider

数据流通活动中提供数据的主体。

3.4

数据需方 data consumer

数据流通活动中使用数据的主体。

3.5

元信息 meta-information

描述数据自身的信息。

注：包括数据ID、数据方ID、数据名字、数据描述信息、其他（数据类型、数据格式）、数据供方的签名。

4 缩略语

下列缩略语适用于本文件。

DMP：数据管理平台（Data Management Platform）

DSP：需求方平台（Demand Side Platform）

FL：联邦学习（Federated Learning）

MPC：多方安全计算（Secure Multi-party Computation）

SSP：供应方平台（Supply Side Platform）

TECC：可信密态计算（Trusted Environment based Cryptographic Computing）

TEE：可信执行环境（Trusted Execution Environment）

5 互联网广告数据流通基本原则

互联网广告数据流通应遵循如下原则：

a) 合法正当：数据流通应遵守我国关于数据安全的相关法律法规，数据流通全流程不得危害国家安全、公共利益，不得损害组织和个人的合法权益；
b) 目的明确：应具有明确、清晰、具体的数据使用目的和场景；
c) 最小必要：只提供满足特定数据使用目的所需的最少数据类型和数量；
d) 可算不可识：对数据的加工处理兼顾数据的有用性和在可信计算区中个人信息的不可识别性和不可复原性；
e) 专数专用：根据数据供方的授权范围限定数据的使用范围，不应不经数据供方授权随意跨主体跨场景使用数据；
f) 权责一致：数据流通各相关方根据所处的不同环节，采取必要的措施保障数据的安全，对数据流通过程中对组织和个人的合法权益造成的损害承担相应的责任；
g) 安全可控：采取必要的措施保证数据流通全流程数据的机密性、完整性和可用性，确保全流程数据处理活动的可追溯，对数据安全事件具备数据安全应急响应和处置能力。

6　互联网广告数据流通平台基本模型

6.1　概述

互联网广告数据流通框架如图1所示，通过控制面和流通面两个层面的结合，共同完成数据的安全流通，保障数据流通各方的权益。

控制面是数据流通的管控层，通过一系列管控机制，使数据流通各个环节对数据的加工处理和使用都有授权和依据，实现专数专用，防止数据滥用。

流通面是数据流通的实现层，通过提供可信计算环境，对参与流通的数据做密态处理，开展安全计算，实现数据全密态化流通。

整个平台中数据以密态流转，一方面，除数据供方外，其他各方拿到的都是密文数据，保障数据供方的持有权，流通面通过与控制面的管控机制结合，实现数据使用权跨域管控；另一方面，实现数据可算不可识，保障数据流通过程中的机密性和完整性。

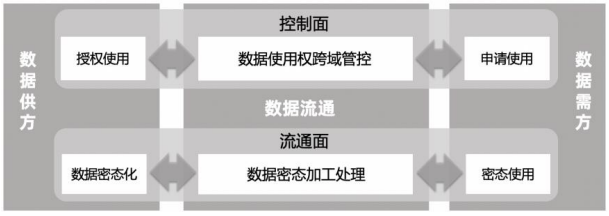

图 1　互联网广告数据流通框架

6.2　参与主体

互联网广告数据流通参与主体包括广告主、广告经营者（如DSP、SSP、广告代理等）、广告发布者（如媒体）、及其他服务提供者（如DMP，广告监测公司等）。根据不同的场景，担任数据流通中的不同角色，参与主体的角色包括：

a) 数据供方：原始数据的拥有者。数据供方将数据进行预处理后提供给数据流通平台。根据安全要求、计算方式等的不同，数据供方对数据的预处理包括但不限于去标识化、数据加密、加噪等；
b) 平台管理方：数据流通平台的运营者。平台管理方提供数据撮合、数据计算任务管理、数据权限管理等服务；
c) 数据需方：数据分析计算结果的需求者。数据需方向数据流通平台方发起请求，并接受来自数据流通平台提供的数据计算服务或算法服务；
d) 算法提供方：提供计算脚本、计算方法者。数据需方可直接使用算法，对数据进行分析加工。

6.3　模型设计

互联网广告数据流通处理数据在多个主体间流动，涉及数据存储、使用、加工、传输、提供多个环节，需要提供安全的数据流通方案，保证数据流通过程中的安全可管控。如图2所示为互联网广告数据

流通平台模型设计逻辑示意图，主要借助密码学和可信区等手段，提供密态数据出域管控能力。密码学提供端到端的安全能力，屏蔽中间环节的攻击；可信区能够抵御实际管理者的攻击，为多个数据供方提供一块公共的、安全的运行环境。模型应包括以下能力：

a) 数据存储：数据流通平台可提供枢纽模式和管道模式。枢纽模式下，数据流通平台存储数据供方提供的数据，并应对数据进行加密存储；管道模式下，数据流通平台不存储数据供方提供的数据；
b) 信道建立：数据流通平台应提供技术手段，如：运行在可信执行环境之中，数据方能够通过可信执行环境的远程认证机制确认数据流通平台的代码等方式，让数据供方确信数据只能被合适的可信区接收；
c) 属主验证：即数据及其属主关系关联。应关联数据和它的所有者，并保证数据及其所有者的关联关系不会被篡改或伪造。对于存储的中间结果，如包含多个属主，都应进行关联；
d) 标识验证：可信区可使用 ID 标识实体，应确保 ID 与实体的对应关系不会被篡改。ID 与实体的对应关系可包括参与方 ID 与参与方密钥之间、数据 ID 与数据之间、算法 ID 与算法之间的对应关系；
e) 授权验证：应实现数据供方对其他方在进行数据使用前的授权，并确保该授权信息不会被篡改或伪造。应实现数据供方对数据发布和使用授权的撤销功能；
f) 身份验证：应具备验证数据需方身份和与其交互的参与方身份的能力；
g) 数据解密：应保证只有在可信计算区对数据进行解密；
h) 安全计算：可信区应能够承载计算功能，并确保攻击者不会通过窥探、干扰计算过程，获得敏感信息；
i) 结果分发：应确保仅有被授权的一方才能接收计算结果；
j) 监测功能：应实现数据全流程可审计、可追溯。

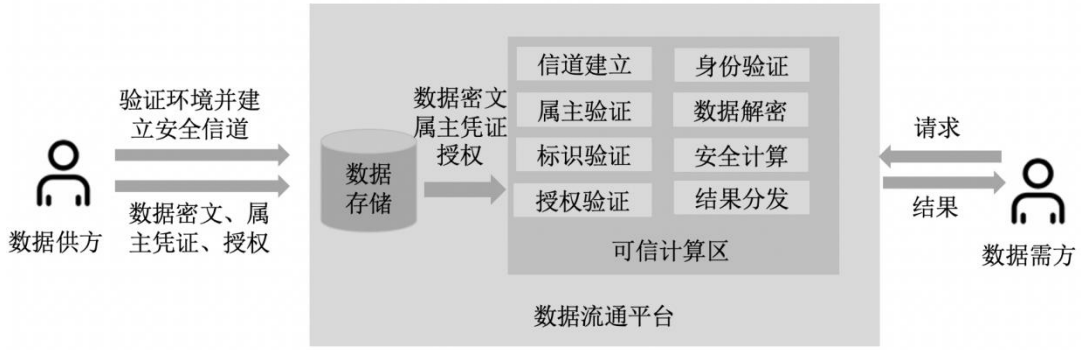

图 2　互联网广告数据流通平台模型设计逻辑示意图

6.4　适用场景

互联网广告数据流通平台应能满足广告投放、广告监测、效果归因、程序化购买、反作弊等典型场景的数据流通需求。附录A为基于数据流通平台的广告营销-媒体曝光率统计示例。

7　互联网广告数据流通平台技术架构

7.1　架构概述

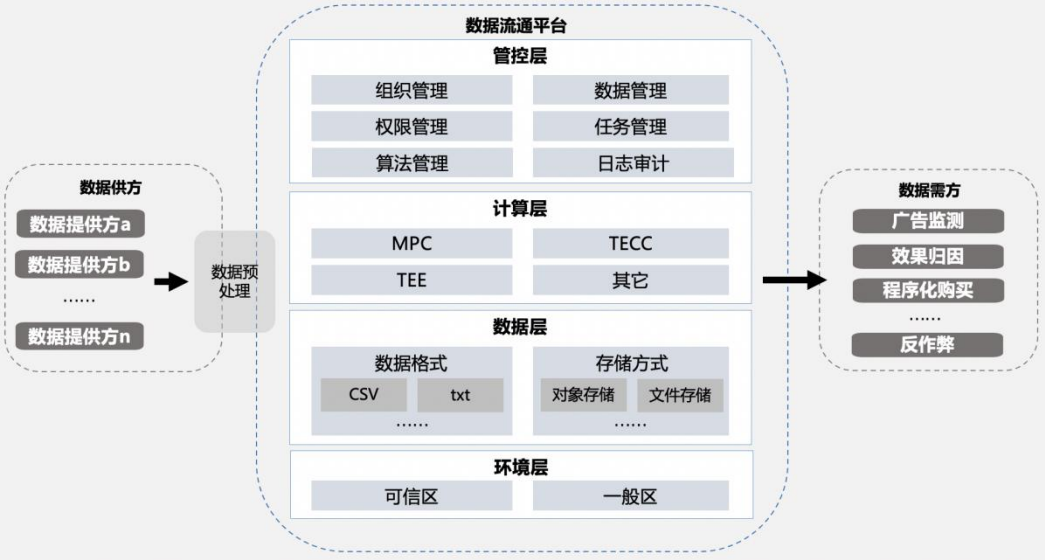

图 3　互联网广告数据流通平台技术架构

如图3所示为互联网广告数据流通平台技术架构，包括环境层、数据层、计算层和管控层。

7.2　架构解析

7.2.1　环境层

环境层应包括可信区和一般区。具体要求如下：

a)　可信区：应提供代码级验证和强隔离能力，为互联网广告数据的计算、存储、密钥生成和存储等提供安全环境；

b)　一般区：数据流通平台管理、计算集群管理、任务分发等功能可在一般区中实现。

7.2.2　数据层

数据层应提供多种数据接入格式和存储方式，支持公有云、混合云、本地数据接入。具体要求如下：

a)　数据格式：支持 CSV、txt 等文件格式；

b)　存储方式：支持对象存储、文件存储等方式。

7.2.3　计算层

计算层应提供多种密态数据计算引擎，可支持管道模式、枢纽模式等异构计算引擎，完成多方隐私求交、相关分析、安全数据查询以及联合建模等功能。计算层可通过预置大量函数库来满足不同场景下的计算需求；同时可支持自定义函数。具体要求如下：

a)　管道模式：应支持 MPC、FL 等至少一种隐私计算技术的计算引擎。数据供方和数据需方可利用计算引擎直接进行数据加工处理；

b)　枢纽模式：支持 Spark、TEE、TECC 等计算引擎。枢纽模式对数据的加工处理应在数据流通平台中进行。

7.2.4　管控层

管控层应结合数字广告实际业务场景和数据流通需求，提供组织管理、数据管理、权限管理、任务管理等功能。具体要求如下：

a)　组织管理：应提供参与方注册、注销等功能：

1)　数据流通平台应提供参与方身份注册功能，参与方要向数据流通平台注册自己的证书或公钥，同时平台应分配给参与方身份唯一 ID；

2)　数据流通平台应提供参与方身份注销功能，身份注销后平台应销毁参与方相关的证书或公钥记录信息；

b)　数据管理：应提供数据元信息的增、删、查等功能：

1) 数据供方应在数据流通平台上管理自己的数据元信息资源，数据需方可在数据流通平台上查询自己需要的数据资源信息；
2) 数据供方可对数据做预处理后将数据上传到流通平台；
3) 流通平台应对数据供方上传的数据验证数据所属关系，验证通过后，平台应记录所属关系，并签署数据持有者凭证给数据供方；
4) 数据供方可通过平台删除已发布的数据，同时平台应注销签署的数据持有者凭证；
5) 流通平台应在可信区中存储加密后的计算结果并通过平台发布计算结果。数据需方在流通平台上可查看到某个任务的计算结果文件，也可通过申请授权获取下载计算结果对应的文件；

c) 权限管理：提供数据使用审批、数据授权管理、计算结果分发审批等功能，为各类计算引擎提供通用的鉴权引擎：
1) 数据需方可在流通平台上发起申请数据使用的请求。流通平台应将这个请求转发给数据供方进行审批，数据供方在审批通过后，应签名访问控制规则并上传到平台；
2) 流通平台收到数据需方发起的任务请求后，应先验证该任务请求相应的数据权限是否合理，在确认合理后才可提交到计算引擎执行该任务；
3) 计算任务结束后，若数据需方需要获取计算结果，可在流通平台上发起申请获取计算结果的请求，流通平台将请求转发给数据供方进行审批，后者在审批通过后才可获取计算结果。
4) 数据供方可在流通平台上查看自己的数据授予给了哪些机构和场景，可向平台发送撤销权限的请求；
5) 计算引擎在执行计算任务前，需要验证数据与数据供方的所属关系以及对数据需方相应的授权，验证通过后方可执行计算任务；

d) 任务管理：提供计算任务管理、任务编排等功能：
1) 数据需方在流通平台上可配置计算任务信息，任务信息包含要执行的算法类型，选择使用哪些数据，流通平台会记录数据需方、数据供方的 ID 和当前的计算场景；
2) 数据需方可在流通平台上查看自己创建的任务的状态和详情；

e) 算法管理：提供算法的管理功能。算法提供方可上传算法脚本，平台审核通过后可发布供数据需方使用；

f) 日志审计：对数据流通全流程进行日志记录与存证，保证数据流通全流程可审计、可追溯。

7.3 业务流程

数据流通平台业务流程如图4所示，涉及数据供方、数据流通平台（及平台管理方）、数据需方、算法提供方等角色，涉及数据注册、数据发布、数据申请、数据授权、数据加工、取消授权、数据撤销、数据结果获取等环节。

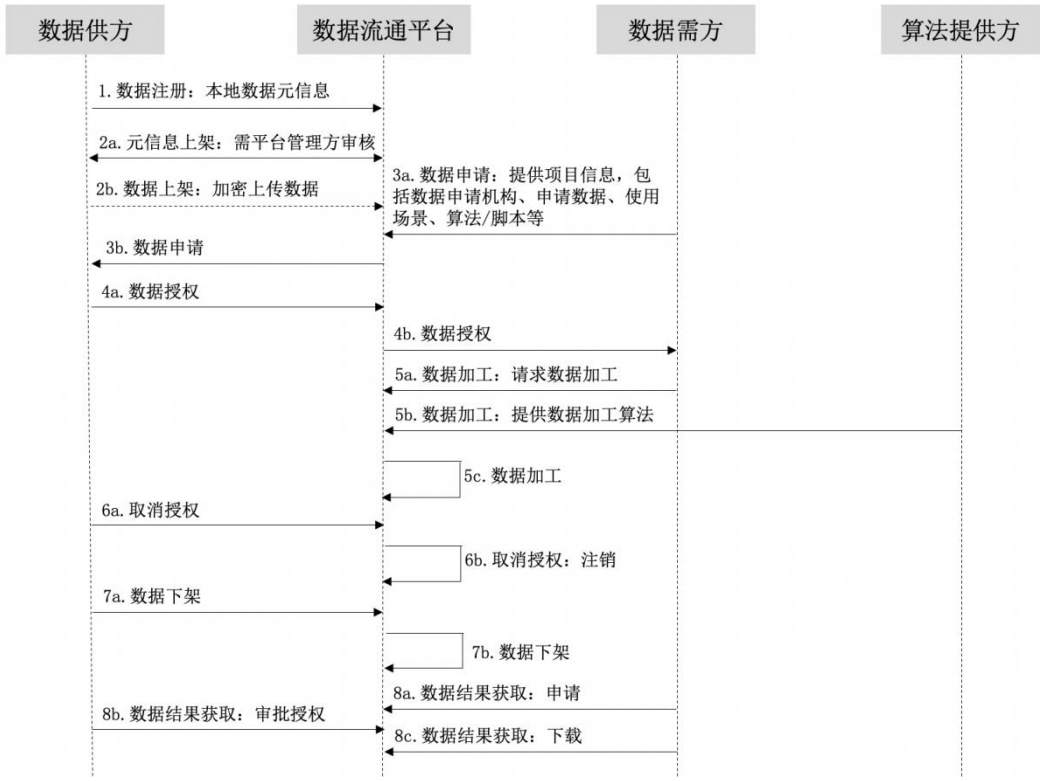

图 4　数据流通平台业务流程

数据流通平台业务流程具体包括：

a) 数据注册：数据供方将本地数据元信息同步至数据流通平台；
b) 数据发布：平台管理方对数据供方提供的元信息进行审核，审核通过后，将元信息发布，并给数据供方颁发数据所属关系凭证；数据供方可选择将本地数据预处理后传输至数据流通平台，也可选择仅在本地存储。根据不同的需求，数据所属关系凭证可仅用于审计和展示或权限鉴别；
c) 数据申请：数据需方以项目形式在数据流通平台发起数据使用申请；
d) 数据授权：数据供方将数据授权至数据流通平台具体项目，数据流通平台生成数据资源持有凭证和数据加工使用凭证。根据不同的需求，数据资源持有凭证和数据加工使用凭证可仅用于审计和展示，或用于权限鉴别；
e) 数据加工：数据需方发起数据加工申请，平台管理方审核通过后进行数据加工处理。其中数据加工的算法/脚本可由数据流通平台、数据需方、算法提供方中的任意一方提供。如采用枢纽模式，数据加工处理应在数据流通平台中进行；如采用管道模式，数据加工处理可由数据供方和数据需方直接交互完成；
f) 取消授权：数据供方可撤销已授权项目的数据，授权撤销后项目不可使用该数据，数据流通平台注销数据加工使用凭证；
g) 数据撤回：数据供方有权将已发布数据进行撤回，撤回后的数据在数据流通平台上不可见，并在数据流通平台中删除此数据及其元信息；
h) 数据结果获取：数据需方在流通平台发起数据加工结果获取申请，数据供方在数据流通平台进行审核授权，通过后数据需方下载数据加工结果。

7.4　关键技术

在数据流通中可用到的关键技术包括：

a) 数据加密：在数据存储和传输过程中，可使用数据加密技术保障存储和传输的安全性，降低明文存储及传输导致的数据泄露风险；

b) 去标识化：对个人信息的进行处理的技术集合， 使其在不借助额外信息的情况下， 无法识别个人信息主体。常用的数据脱敏技术包括：泛化技术、抑制技术、假名化技术、随机化技术、密码技术等；
c) 数字签名：在数据融合计算过程中，可使用数字签名技术验证操作人身份及数据完整性，以确保融合计算及合约的不可抵赖性。数字签名基于非对称密码学算法实现；
d) 差分隐私：在加工或输出数据计算结果时，可使用差分隐私技术对差分攻击进行防御，以实现个人信息和主体权益的保护；
e) 联邦学习 FL：当多方进行联合建模数据处理服务时，可考虑使用联邦学习实现多方的数据保护。基于数据特点，联邦学习可分为横向联邦学习、纵向联邦学习和联邦迁移学习。在联邦学习的技术上，还可使用差分隐私、同态加密等技术对数据进行保护。适用于管道模式；
f) 安全多方计算：多方原始数据不出域进行联合计算，并且数据计算量不大的场景可考虑使用安全多方计算。安全多方计算通常基于秘密共享、不经意传输、混淆电路等基础技术实现。应用较多的安全多方计算技术包括：隐私集合求交、隐私集合求并以及隐匿信息查询等。适用于管道模式；
g) 可信执行环境：在数据融合计算场景中，可使用可信执行环境搭建安全可信的数据加工和使用环境，以增强硬件安全能力。适用于枢纽模式；
h) 可信密态计算：将数据以密态形式在高速互联的可信节点集群中进行计算、存储、流转的一种可信隐私计算技术，实现数据持有权有效保障、使用权出域可控，支撑任意多方大规模数据安全、可靠、高效地融合与流转。TECC 具有可信节点内进行密态计算、数据持有方与计算方的解耦、域外可控的数据密态封装等基本特征，可通过安全编程语言、形式化验证、多级别可信节点等进一步提升安全性和适用性。适用于枢纽模式。

附　录　A
（资料性）
广告营销-媒体曝光率统计示例

A.1　业务场景描述

在数据化的大时代，曝光次数的可精准计算成为媒体、广告营销商的主要优势之一。以广告投放效果评估-投放媒体曝光重复率为例，通过不同的工具收集诸如电梯广告、屏幕广告、移动端广告的单次曝光重复率，以及总曝光次数，并且确保每次曝光精准覆盖广告主所需场景，触达目标受众，实现广告的“聚焦攻击”。

在投放媒体曝光重复率场景中广告营销商拥有各自的广告营销平台和广告营销渠道（如移动端渠道、大屏渠道、电视渠道等），广告营销商均为广告主提供广告营销解决方案。当前，各广告营销商通过自建DMP（数据管理平台）获取用户投后相关数据，并基于自有数据，统计广告投放后的效果情况（如曝光率、转化率等），由于自身数据覆盖度、精准度等均存在一定程度的问题，导致曝光次数统计结果精准度不高。需要结合更多机构的数据，提升曝光次数统计结果的精准度。然而广告曝光数据涉及用户、机构等多方，需要在使用数据的同时，保障数据相关方的权益。

A.2　基于互联网广告数据流通平台的实践

广告数据流通平台的建设，能够提供不互信机构间数据合作的安全环境，在保证各方数据安全和各方权益的情况下，密态计算机构间的投后曝光率次数统计结果，基于该结果优化广告营销方案，提升广告投放效果效果。具体方案如下图A.1所示：

图 A.1 基于广告数据流通平台的方案

广告数据流通平台依托数据密态计算技术，结合数据权属模块的精细化管控数据使用，避免数据的越权滥用。在该模式下，广告营销商、广告主之间能够保证数据安全、业务合规的情况下实现多方共赢，打消数据方对数据流转失控的顾虑，提升数据协作意愿，充分释放数据价值，提升营销业务效果。

参 考 文 献

[1]GB/T 25069-2022 信息安全技术 术语
[2]GB/T 37964—2019 信息安全技术—个人信息去标识化指南 附录A
[3]T/CAAAD 004-2022 T/CCSA 424-2022 互联网广告 匿名化实施指南