CAICT 中国信通院　中国广告协会 CHINA ADVERTISING ASSOCIATION　蚂蚁集团 ANT GROUP

集智 蓝皮报告

数字广告数据要素流通保障技术研究报告

（2023 年）

中国信息通信研究院泰尔终端实验室

中国广告协会

蚂蚁科技集团股份有限公司

2023年11月

版权声明

前　言

2022 年 12 月 19 日，中共中央、国务院《关于构建数据基础制度更好发挥数据要素作用的意见》对外发布（简称“数据二十条”）。“数据二十条”指出，以习近平新时代中国特色社会主义思想为指导，深入贯彻党的二十大精神，完整、准确、全面贯彻新发展理念，加快构建新发展格局，坚持改革创新、系统谋划，以维护国家数据安全、保护个人信息和商业秘密为前提，以促进数据合规高效流通使用、赋能实体经济为主线，以数据产权、流通交易、收益分配、安全治理为重点，深入参与国际高标准数字规则制定，构建适应数据特征、符合数字经济发展规律、保障国家数据安全、彰显创新引领的数据基础制度，充分实现数据要素价值、促进全体人民共享数字经济发展红利，为深化创新驱动、推动高质量发展、推进国家治理体系和治理能力现代化提供有力支撑。

互联网广告是密集型数据使用、加工、提供和委托处理的行业领域。无论是广告投放还是程序化交易、广告归因等场景，无不涉及到个人信息相关的数据在不同机构间的快速流转、传输、共享和使用等行为，潜在包含了隐私侵害等数据流通风险。因而，亟需增强产业对互联网广告数据流通的安全管理，在确保数据安全的前提下，激活数据要素的自由流通，促进互联网广告产业的安全、健康、快速发展。

本报告将从数字广告数据要素流通背景、现状、面临的挑战与数字广告数据要素流通保障技术等方面展开分析，在梳理平台设计思路、平台技术架构和分析平台广告方案特点的基础上，探索广告数据流通

平台合规实践方案，最后从强化政策引领、完善标准体系、加强技术研究、加快推广建设、开展评估检测等方面给出规范和发展数字广告数据要素流通的对策建议。

目 录

图 目 录

一、数字广告数据要素流通概述

数据的价值在于充分的流通使用。作为新型生产要素，数据要素具有独特的技术、经济特征，可重新编程性和场景依赖性、用户体验性、广泛赋能性等特征使得数字环境下的数据要素流通能够突破传统时空限制、组织边界束缚与行业壁垒，以多源多模态实时数据的无障碍流通与价值增值为核心，深刻改变人们的学习工作和生活方式，并为数字产业化和产业数字化的加速发展奠定基础。当前，数据已被视为构建现代化产业体系的重要组成部分，有序衔接、高效畅通的数据使用是优化产业结构、发展新质生产力不可或缺的基础要件。麦肯锡全球研究所称，数据流动带来价值已超过全球货物贸易的价值。

发展数字广告是世界性趋势，2021 年，国家统计局在印发的《数字经济及其核心产业统计分类》中，将数字广告列为数字经济及其核心产业的统计范畴，标志着数字广告的产业发展已跨越传统广告业而进入到新的发展阶段。作为全球第二大广告市场，中国数字广告产业市场规模达到万亿元以上，对于促进消费、扩大就业、开发创意空间、激发市场活力从而拉动经济高质量增长具有重要的牵引推动作用。

(一) 数字广告数据要素流通发展背景

1.国家高度重视数据要素发展

数据是数字经济时代的关键生产要素，2020 年中共中央、国务院印发《关于构建更加完善的要素市场化配置体制机制的意见》，将数据与土地、资本、技术、劳动并列为五大生产要素。2021 年，总书记在中共中央政治局第三十四次集体学习的讲话中强调“数据作为新型

生产要素，对传统生产方式变革具有重大影响”，以数据为关键要素推进数字产业化和产业数字化。2022 年“数据二十条”提出要建立“以数据产权、流通交易、收益分配、安全治理为重点”的数据基础制度，数据要素流通在基础制度体系中无疑具有重要意义。

2.数字广告助力数字经济建设

数字技术和数据要素是所有行业数字化转型的基础，催生了全新的数字产业。2021 年 6 月，国家统计局出台《数字经济及其核心产业统计分类（2021）》，明确提出数字广告是数字经济的核心产业之一，属于数字产业化中的数据要素驱动业，对应 GB/T4754-2017《国民经济行业分类》中的“互联网广告服务”。相关统计显示，2018-2022 年，数字广告在数字经济规模中的占比从 1.4%升至 1.8%，增幅超过广告在 GDP 中比重的变化水平，凸显出数据赋能的广告产业，成为数字经济高质量发展的助推器。

3.数字技术推动广告产业迭代更新

我国经济正处于由高增长向高质量发展的重要转型阶段，各行各业数字化转型是经济高质量发展的内在要求。通过数字技术的发展与应用，在数字广告领域，竞价搜索技术、程序化交易技术、展示技术等都在加速产业数字化转型和数字经济建设，为广告行业的数字转型、智能升级、融合创新提供支点，对提高广告投放精准度发挥乘数倍增作用。数字广告实现了从对媒介技术的高度依赖到独立的数字广告技术开发的转变，实则也是以数字技术为基础，通过数字广告数据要素的流动创新，驱动数字经济发展的重要体现。

(二)数字广告数据要素流通概念

1.数据要素流通基本概念

数据要素是指以电子形式存在的、通过计算的方式参与到生产经营活动并发挥重要价值的数据资源。数据要素流通是其价值挖掘与转移过程，主要包括原始数据的产生与收集，数据处理、组织成数据产品，数据产品登记、挂牌上市，数据产品试用与交易，数据产品交付与服务以及数据产品使用等关键环节。“数据二十条”等重要制度明确，建立数据可信流通体系，增强数据的可用、可信、可流通、可追溯水平，是激活数据要素潜能、赋能实体经济的重要途径，而完善的数据要素治理体系则是推动数据要素流通、充分释放数据价值的前提保障。

当前，数据要素流通的权属、合规问题已成为全球高度关注的焦点，尚处在各国的积极探索阶段，并未形成共识性规则。相比传统的生产要素，数据要素的非竞争性、非排他性、非耗竭性、非稀缺性、非均质性、边际效应递减性等特质使其在产权、定价、交易、监管、安全等方面展现出新的形态要求，原有的生产要素市场理论面临全新的挑战。不仅需要系统性探索数据要素的价值实现规律、市场管理框架和风险治理机制，还需要充分运用大数据思维和数字技术，为数据要素流通提供数字化的基础支撑与替代性的技术解决方案。

随着党中央、国务院相关制度文件的密集性出台，2023 年我国数据要素市场建设进入快车道，但数据流通使用实践中长期存在的“获取难、确权难、定价难、互信难、监管难”等难题仍未根治。整体而

言，数据要素流通的理论体系不够完善，方法手段不够成熟，数据权属界定不明晰、估值流通难落地、数据安全风险高、隐私侵害隐患大、数据交易机制不完善等问题严重影响了我国数据要素的流通。

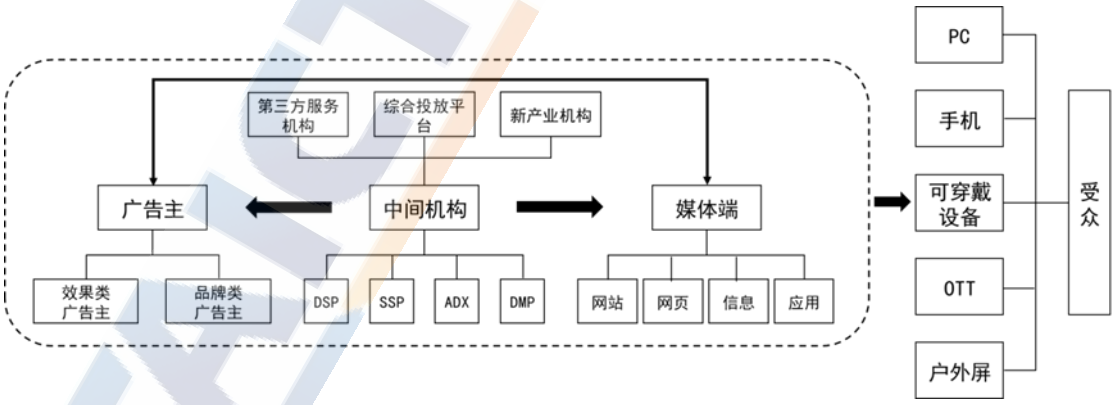
2.数字广告数据要素流通基本概念

数字广告是高度密集型数据使用、加工、提供和委托处理的行业领域，具有数据量庞大且处理速度要求高等特点。目前，数字广告已从内容密集型产业升级为数据密集型产业，所有的广告创作、投放等广告流程以及广告公司经营和转型等都高度依赖数据要素的高效流通与计算使用。例如，用户标签数据、行为数据等个人信息既加速了个性化数字广告精准推送业务的开发，也决定了广告策略、对象群体、可视化风格等传统广告体系的解构与重构。同时，数字广告业务流程的各个阶段，无论是广告投放还是程序化交易、广告归因等，均涉及大量与个人信息相关的海量数据在不同机构间的流转、加工、传输、共享和使用。因此，数字广告产业的发展必须充分考虑数据合规与数据要素的合法利用问题，以充分发挥互联网技术优势，增强数字行业竞争力，释放数据要素经济社会价值。

目前，数字广告已演化为系统性、有强大演进能力的数据处理系统。比如程序化广告，每一次广告展示机会，都可能引发多个广告主参与竞争，他们各自根据所掌握的数据计算该广告的预期收益，如果预期收益较高，广告主就会以较高的价格竞争该次广告，最后出价最高的广告主获得本次展示机会。在这一规则下，每个广告主都在竭力提升自己的评估水平，进而带动整个行业发展水平在竞争中得到快速

提升。另一方面，也折射出广告行业竞争的背后本质上是数据应用能力的比拼。

巨大的数字广告行业，不是由一两种类型机构组成的，而是由代表不同主体利益、拥有不同角色定位、面向不同对象群体的多个机构协同构成的。下图 1 展示了广告行业中的典型角色。其中，媒体端负责提供广告展示，指向网站、内容类手机 APP 等。广告主系指投放广告的一方，比如服装、消费电子产品方。中间机构包括供应方平台（Supply Side Platform，简称 SSP），即代替媒体端对外提供广告位置的平台；需求方平台（Demand Side Platform，简称 DSP），代替广告主执行广告投放策略、购买广告位的平台；广告交易平台（Ad Exchange，简称 ADX），联系广告主和媒体，或者 DSP 和 SSP，组织竞价、撮合交易的市场平台；数据管理平台（Data Management Platform，简称 DMP），为 DSP、SSP 等提供数据挖掘、分析、管理等数据能力支持。



来源：中国信息通信研究院

图 1 数字广告角色

广告行业蓬勃发展的背后是通过大规模数据系统打通不同数字

广告角色的机构系统，并实现数据在不同机构间的流转运行。以某 SSP 公司提供的广告投放相关场景中数据流通为例，SSP 首先会通过 SDK 或者 API 的方式采集相关设备数据，主要包括设备信息、用户信息和应用信息等；然后利用其他数据平台对数据进行赋能操作（例如用户标签补全）；最后 SSP 把上两部分数据整合后传递给 DSP，供其投放广告使用。SSP 从 APP 侧采集的数据主要分为 6 类：**广告位信息、用户信息、设备信息、网站信息、应用信息、地理位置信息。**其中，用户信息、设备信息、网站信息和地址位置信息均有内容涉及个人信息。随着个人信息相关法律法规的颁布，企业对这部分数据的流通持谨慎态度。

（三）数字广告数据要素流通场景

数字广告数据要素流通场景一般包括程序化购买、广告监测、广告效果评估、无效流量排查和反作弊等业务场景。

1.程序化购买

在程序化购买场景中，一般采用开放式实时竞价协议规范（Open Real Time Bidding，简称 OpenRTB），主要涉及广告主、媒体方和 DSP、SSP、ADX、DMP 等广告投放平台。实时竞价过程是指用户在访问媒体网站或媒体应用程序上的广告位置时，该广告位置接入的 SSP 或者 ADX 会根据 OpenRTB 将广告请求数据封装在竞价请求中，并发送给多个对接的 DSP，各 DSP 根据自身的竞价策略决定是否参与竞价，并将结果回复。

在实时竞价过程中，如果广告主需要使用 DMP，DSP 在竞价决

策过程中会向第三方 DMP 发送设备标签查询请求，请求中会携带**设备信息**，第三方 DMP 会回复该设备命中的标签情况。

在实时竞价过程中，如果广告主需要使用第三方无效流量(Invalid Traffic，简称 IVT)过滤服务平台，DSP 在竞价决策过程中会向第三方 IVT 过滤服务平台发送本次流量 IVT 查询请求，请求中会携带本次请求中的**设备信息、IP 信息**等，第三方 IVT 过滤服务平台回复本次流量的 IVT 得分情况。

2.广告监测

广告监测主要涉及广告主、媒体和广告监测公司。广告监测的一般场景为：在广告活动开始前，广告主或者代理公司向媒体、广告投放公司以及广告监测公司提供广告活动排期，广告监测公司提供监测代码，媒体或者投放方在不同广告点位添加对应的监测代码。在广告活动进行过程中，当有曝光或者点击等需要监测的行为发生时，媒体向广告监测公司发送一条监测请求，广告监测公司收到的监测请求，收集广告信息并记录曝光、点击等指标，然后根据收集到的用户行为数据（包含**用户信息、设备信息**）进行处理生成报告和服务，最终提供报告给广告主。

3.广告效果评估

广告效果评估主要涉及广告主、广告监测公司。

一个完整的广告效果评估由**广告效果监测**和**广告归因**共同构建。通过广告效果监测，广告主实时追踪广告投放的数据和指标，了解广告活动的绩效情况。广告归因则帮助广告主理解广告投放中哪些因素

对于转化和销售产生了最大的影响，以及各个广告触点之间的相互作用。

广告效果评估的一般场景是：在广告监测数据的基础上，DMP 通过技术手段获取用户对广告或广告品牌的响应、认知变化等数据（**包含用户信息、设备信息、地理位置信息**）。DMP 对接收到的广告监测数据及广告效果相关数据进行清洗后，按照广告曝光、点击、互动、品牌认知、品牌推荐等进行分类存储。广告监测公司对广告监测数据和广告互动效果数据通过反作弊算法过滤掉无效流量，然后通过用户标识进行匹配，匹配成功即为一个有效转化样本，进而结合收集到的所有有效数据综合评估并产出本次广告效果评估报告。

4.异常流量排查和反作弊

异常流量排查和反作弊主要涉及广告主、媒体方和广告监测公司。

异常流量排查和反作弊的通用场景是：首先由广告监测公司识别异常流量。判定异常流量时，广告主需要向广告监测公司发送本次流量 IVT 查询请求，请求中会携带本次请求中的**设备信息、IP 信息**等。其次对判断为异常的流量，在媒体或广告主需要对判定结果复核时，需要监测机构、媒体、广告主三方在数据安全规范的条件下进行，提取异常数据样本或全部复核。最后，对新型的作弊手段，参与方可通过数据分析，提炼出共性特征，完善现有反作弊规划或丰富黑名单库。

结合数字广告业务场景与数据要素特征，厘清数字广告数据要素流通路径，释放数据要素价值，是推动数字广告业务精准化、助力数字经济发展的必经之路。

二、数字广告数据要素流通现状和挑战

(一) 数据安全相关政策规范逐步健全

我国陆续出台多项数据相关的法律法规和国家标准，明确提出对个人信息的处理要求和保护要求。

2017 年 6 月 1 日施行的《中华人民共和国网络安全法》明确对公民个人信息安全进行保护：第四十四条指出任何个人和组织不得窃取或者以其他非法方式获取个人信息，不得非法出售或者非法向他人提供个人信息。

2020 年 3 月 1 日实施的国家标准 GB/T 37964-2019《信息安全技术 个人信息去标识化指南》，细化了去标识化活动的开展过程，就个人信息去标识化问题给出了具体指导。

2020 年 10 月 1 日实施的国家标准 GB/T 35273-2020《信息安全技术 个人信息安全规范》规定开展收集、存储、使用、共享、转让、公开披露、删除等个人信息处理活动应遵循的原则和安全要求。

2021 年 9 月 1 日正式实施的《中华人民共和国数据安全法》明确提出，开展数据处理活动，应当遵守法律法规，尊重社会公德和伦理，遵守商业道德和职业道德，诚实守信，履行数据安全保护义务，承担社会责任，不得危害国家安全、公共利益，不得损害个人、组织的合法权益。同时第三十二条也明确规定：任何组织、个人收集数据，应当采取合法、正当的方式，不得窃取或者以其他非法方式获取数据。

2021 年 11 月 1 日开始施行的《中华人民共和国个人信息保护法》特别要求，个人信息处理者在处理敏感个人信息、向他人提供或公开

个人信息、跨境转移个人信息等环节应取得个人的单独同意，并要求通过自动化决策方式向个人进行信息推送、商业营销，应当同时提供不针对其个人特征的选项，或者向个人提供便捷的拒绝方式。

2022 年 11 月 1 日正式实施的国家标准 GB/T 41391-2022《移动互联网应用程序(App)收集个人信息基本要求》，明确提出 APP 收集个人信息还应满足定向推送信息和用户画像场景采用唯一设备识别码识别用户时，应使用可交更的唯一设备识别码，且不应将其与用户身份信息或不可变更的唯一设备识别码关联等要求。

2023 年 8 月国家互联网信息办公室就《个人信息保护合规审计管理办法（征求意见稿）》向社会公开征求意见，提高个人信息处理活动合规水平，保护个人信息权益。这意味着广告主必须获得用户明确的、自由的、明示的同意，才能收集、使用、处理和传输个人信息。

2023 年 10 月 24 日发布、2024 年 1 月 1 日起施行的《未成年人网络保护条例》作为我国第一部专门性的未成年人网络保护综合立法，注重保障未成年人在网络空间的健康成长，提出任何组织和个人不得在专门以未成年人为服务对象的网络产品和服务中制作、复制、发布、传播可能影响未成年人身心健康的信息、网络产品和服务提供者不得通过自动化决策方式向未成年人进行商业营销。

2023 年 10 月 1 日实施的国家标准 GB/T 42460-2023《信息安全技术 个人信息去标识化效果评估指南》旨在依据个人信息能多大程度上标识个人身份（即标识度）进行分级，用于评估去标识化活动的效果，以此形成的个人信息标识度分级，该标准尚无法解决不同分级

下法律效果的差异，但是有利于面向未来探讨一种分类分级的治理措施。

（二）广告数据安全相关标准持续制定

为推动数据资源自由流通，促进互联网广告产业的安全、健康、快速发展，中国通信标准化协会（CCSA）与中国广告协会（CAA）联合制定了互联网广告数据安全与个人信息保护系列标准，包括:

- 《互联网广告 数据应用和安全技术要求》
- 《互联网广告 个人信息告知同意分类指南》
- 《互联网广告 匿名化实施指南》
- 《互联网广告 群体标识技术要求》
- 《互联网广告 数据分类分级指南》
- 《互联网广告 数据流通平台技术架构》
- 《互联网广告 隐私计算平台技术要求》

2021 年 12 月 31 日，CCSA 与 CAA 联合发布首个双编号团体标准: T/CCSA329-2021 | T/CAAAD001-2021《互联网广告数据应用和安全技术要求》，旨在规定互联网广告数据的应用原则、应用场景和安全技术要求，并给出了互联网广告数据应用数据交换接口定义和升级要求等，同时提供了技术指引建议。标准适用于广告主、媒体和流量平台、用户、广告代理公司、广告技术公司、广告监测公司、其他第三方组织等，规范互联网广告数据收集、使用、存储、传输和删除等活动。

CCSA/CAA 联合发布的《互联网广告隐私计算平台技术要求》，

通过规定互联网广告场景隐私计算平台的整体框架、技术流程、安全要求，用于指导互联网广告营销企业、科技企业、用户机构、第三方机构等，对广告场景隐私计算平台的设计、开发、测试、使用，保障数据在流通与融合过程中的“可用不可见”、“数据不动模型动”。

2023 年 1 月 3 日，CCSA 与 CAA 联合发布双编号团体标准：T/CCSA 424-2022 | T/CAAAD 004-2022《互联网广告 匿名化实施指南》规定了互联网广告匿名化实施的目标、原则和适用场景，给出了数据匿名预处理技术指引，相应配套了业务法律边界的评估见证方法和运营过程监控的管理控制体系，实现了“技术问题技术解决，法律问题回归法律，管理问题过程控制”的三位一体互信且制衡的实施方法。《互联网广告 匿名化实施指南》形成的“技术保障、评估规制、过程控制”的互信制衡机制，适用于各类互联网广告业务，包括广告投放、程序化交易、广告监测等应用场景下的数据匿名化处理。

（三）数据泄露引发担忧，流通意愿待加强

数据泄露问题已成为全球性难题。据国际数据公司（IDC）预测，2025 年全球数据量将高达 175ZB，其中，中国数据总量增速最为迅猛，正以高于全球平均年增长速度 3%的幅度激增，预计 2025 年增至 48.6ZB，占全球数据圈的 27.8％，成为全球最大的数据圈。

随之而来的是数据泄露、数据滥用等数据安全事件频发。据《中国政企机构数据安全风险分析报告》介绍，2022 年 1 月至 10 月全球政企机构重大数据安全报道 180 起，其中数据泄露相关安全事件 93 起，比如国内某快递公司 10 亿用户信息发生泄露。

国际上也发生多起严重的信息泄露事件，比如 2018 年 3 月曝光的剑桥分析事件中，Facebook 泄露 8700 万用户数据，影响美国大选；Uber 泄露全球 3500 万用户、370 万名司机的个人信息。数据一旦被获取后，用户便失去了对数据的控制权，若个人数据被不法分子获取，很容易对用户的财产、名誉等造成损失，有甚者更是会对国家安全、公众利益和组织利益带来重大损害。

数据资产作为企业重要的价值资产，潜力巨大但也暗藏安全风险，故企业多倾向于在保留源数据的基础上计算使用。而在数字广告数据要素流通实践中，数据持有方多对参与数据要素流通持焦虑态度甚至不愿参与，具体原因可概括为下述五点：

● 隐私担忧：数据持有方担心参与数据要素流通会泄露用户个人信息，引发隐私问题。

● 竞争风险：数据持有方担心会与竞争对手分享数据，使竞争对手利用这些数据优化自己的广告策略或获得市场份额，从而使数据持有方失去竞争优势。

● 价值回报不确定：数据持有方担心无法获得足够的价值回报，对能否获得充分的经济利益持怀疑态度。

● 数据安全风险：参与数据要素流通可能会面临数据泄露、滥用或黑客攻击等风险，尤其像银行、医疗机构等敏感数据持有方会对数据安全风险格外担忧。

● 法律合规要求：由于数据要素相关制度规范的欠缺，以及数据要素流通、安全保障等技术发展的局限，数据持有方担心在共享

数据的过程中，相关工作人员因不熟悉数据要素特征，触碰数据合规底线。

（四）数据“孤岛”向“平台”演进，平台互通待加快

数据孤岛指的是数据在不同部门间独立存储和维护，受制于技术阻碍、企业/行业保护机制以及政策法规等因素而形成的不对称、冗余等封闭或半封闭式现象，普遍存在于所有需要进行数据共享和交换的系统之间，涉及不同部门、企业、产业间数据信息能否共享等问题。

近年来，互联网平台方为加强数据安全和保护用户隐私，纷纷推出广告数据平台。一些大型企业更是建立了具有独立存储空间和严格权限管理机制、能为重要数据加密且保证数据不出库等特点的 Ads Data Hub；广告主和中小媒体可依附于上述数据平台进行数据融合流通。数据平台的建立，可保证域内数据安全流通，实现数据可用不可见，有效保障用户隐私，一定程度上缓解了“数据孤岛”现象。

数据平台一定程度上能够解决数据流通问题，但同时也成为了各个数据控制者为拒绝数据分享而高筑的壁垒，继而形成数据“围墙花园”。数据“围墙花园”的存在，强化了强势媒体的垄断地位，具体表现在：

● 平台的建造需要大量的人力、物力和财力，这些都需要较大的建设成本，不适合中尾部互联网企业建设；

● 越来越多的数据汇入到平台中，广告主和中小媒体为了利用这些数据开展广告业务，只能与众多强势媒体分别合作;

● 由于平台的建造者是强势媒体，这导致合作过程中，双方的地位不对等，其透明度、安全性主要取决于平台主体，对广告主、中小媒体和监测公司降低了透明度。

互联互通是互联网行业高质量发展的必然选择，数据“围墙花园”可以促进数据要素在域内的安全流通，但不利于数据要素跨域流通，“不怕墙内锁春色，墙外亦能传芬芳”，让数据能够在整个互联网安全、高效地流通是互联网行业的努力方向。

（五）广告标识符策略收紧，演进方案待升级

日益收紧的隐私保护法案及针对互联网巨头的反垄断调查都让广告界通用的基于标识符的用户追踪方式愈加受到挑战。从全球来看，通过唯一标识符开展数字广告业务的趋势也在逐渐收紧，国际上诸如 iOS、Android 等主流操作系统平台基于隐私保护方面的合法合规考虑，都在逐渐弱化或取消直接通过唯一标识符来开展广告业务。

2021 年 4 月，随着苹果 iOS 14.5 的上线，苹果公司正式推出 APP 追踪透明（App Tracking Transparency，简称 ATT）功能，该功能可让用户选择是否允许 APP 跟踪用户在其他公司的 APP 和网站上的活动，以便用于广告投放或与数据代理商共享。与此同时，苹果广告标识符（Identifier For Advertising，简称 IDFA）获取方式从 opt-out（手动选择关闭）变为 opt-in（手动选择加入），用户管理颗粒度从设备变为应用，同时更新 APP Store 政策限制任何第三方标识。在 iOS14

发布后，苹果推出了 SKAdNetwork2.0 版本，可以解决 IDFA 缺失带来的安装以及之后在 APP 内注册、购买等转化行为，但仍然存在使用场景局限性、归因回传延迟、结果完全依赖苹果提供等问题。

谷歌为了保护用户隐私，在谷歌应用商店（Google Play）的管理中心页面显示将不会再向开发者提供那些不想看个性化广告的用户的广告标识符。与此同时，为提供有效的个性化广告体验，谷歌为 Android 引入新的解决方案——隐私沙盒。隐私沙盒通过限制与第三方共享用户信息，通过 Topics API 技术能够在没有跨应用标识（包括广告 ID）的情况下运行，以及通过更加安全的方式来完成应用程序与广告 SDK 的集成。但隐私沙盒技术本身存在一定争议，且利用 Topics API 技术进行广告定向投放、广告归因等业务实践效果仍需相关数据予以支持和证明。

国内也在积极探索新的广告标识符方案，主要分为两条技术路线：

一是探索建立统一高效、安全合规的广告标识符。中国信息通信研究院联合华米 OV 制定了《移动智能终端补充设备标识体系规范》、实现了开放匿名设备标识符（Open Anonymous Device Identifier，简称 OAID）等标识的生成，OAID 可作为广告标识使用，可以被用户主动关闭和重置，但在统一管理和用户对标识符的控制机制上有待进一步提高。

二是群体 ID 技术。群体 ID 代表一群具有某种共性的用户的标识符，对该群共性用户，标识符一致。目前，群体 ID 的生成是基于用户在平台使用产生的数据，单一平台内部能做到通过群体 ID 实现

广告投放、广告推荐，在不感知具体用户及其敏感隐私信息的前提下，为用户提供一定个性化程度的广告推荐服务。但由于各平台对用户群体划分，以及数据表达不一致等多方面因素，群体 ID 尚未完全验证跨平台的应用，基于多方数据的群体 ID 实现方案还需要进一步讨论。

（六）隐私计算助力数据流通，落地应用待加深

为做到数据被无障碍使用且能确保数据安全和个人信息不被泄露，实现“数据可用而不见”，以隐私计算技术为基础的流通方案成为行业首选探索方向。近年来，媒体、监测公司等企业纷纷入局，投入资金和人力进行深入探索，市场也陆续推出相关数据流通平台产品。但由于缺乏统一的标准指导、技术水平良莠不齐，导致平台所提供的数据流通能力也是参差不齐，体现在业务场景支持、技术方案选择和流通过程中数据保护能力等方面。

行业普遍认可的隐私计算技术，涉及多方计算和通信，需对敏感数据进行加密和解密操作，数据量庞大且处理速度要求高，伴随的是计算资源和网络带宽的极大消耗，这一点在大规模广告数据处理中尤为明显。而在大规模广告数据处理中，为保护数据隐私，需使用特定算法，如同态加密、安全多方计算等，这相对于传统计算方法而言更复杂，易降低算法的运行效率，且需要额外的计算开销，增加了计算时间和资源的消耗。据调研，隐私计算技术目前虽初步形成技术体系，但在数字广告行业的落地应用中还需进一步强化与发展。

（七）匿名化技术受关注，实施方案待完善

“数据二十条”明确提出：“创新技术手段，推动个人信息匿名

化处理，保障使用个人信息数据时的信息安全和个人隐私”。尽管我国已将数据上升为“生产要素”，然而在实践中出于数据安全的担忧和隐私合规要求，各个主体掌握的数据是分散而碎片化的，数据往往难以在规模化基础上实现价值利用。在此背景下，匿名化技术作为一种行之有效的解决路径越来越引起更多关注。个人信息匿名化处理有利于实现个人信息向数据要素的转化。个人信息匿名化处理是平衡数据要素流动和数据合规安全利用的一种关键路径。高效地对个人信息进行匿名化处理，能够确保数据要素的合规流通，极大地提高数据利用效率。

但在广告领域的业务实践中，匿名化技术方案没有大规模推广应用，主要挑战在于：匿名化、个人信息的定义标准不明晰，在实践中扩大解读个人信息范畴；同时一些追求绝对安全和零风险的思路导致匿名化合规门槛高、难以落地，在数据使用和共享具体场景中匿名化无法被有效证明；产业实践中的匿名化技术方案还缺乏标准方面的合规认定支撑，难以进行大规模推广。

（八）“告知同意”难实施，机制实现待突破

数字广告涉及曝光、点击、归因分析等多个环节，数据需要在多个机构间进行流转。在法律法规的要求下，用户个人信息不能再随意流转，必须取得用户的同意并采取合理的安全保护措施。如果每个数据处理环节都在用户同意的前提下进行，将对广告投放效率效果和用户体验带来影响。告知同意机制偏重于数据收集环节，难以满足后续数据处理的发展要求。在广告行业，数据持有方既难以在数据收集时

要求用户给予广泛的授权，也难以在数据流通过程中获得用户的二次授权。此时，不仅告知同意的难度加大、成本上升，强调告知同意还可能会影响数字广告产业的发展。

透明和同意框架（Transparency and Consent Framework，TCF）是行业为遵从 GDPR 而制定的传达程序化同意信号的机制，由欧洲互联网广告局（IAB Europe）和互动广告局技术实验室（IAB Tech Lab）在 2018 年 8 月联合推出。2019 年 8 月 21 日推出 TCF 2.0 版本，持续深化 TCF 的总体推动力，涵盖消费者授予或拒绝同意的权利，以及行使反对处理其数据的权利，消费者还可以更好地控制广告技术供应商是否以及如何使用精确地理定位等数据处理的某些功能。IAB 的会员可使用这一框架允许合作伙伴为了广告和其他目的合法收集和处理数据。但是，为满足 GDPR 的要求而设计的这一技术框架，在 2022 年还是被欧盟监管方按照 GDPR 进行了处罚。

目前在我国，仅有少数有海外广告业务的国内厂商加入 TCF 框架，但大部分仍处于研究和观望阶段，类似 TCF 同意管理框架是否符合我国法律法规还有待商榷，国内尚未形成统一的框架体系和同意管理平台。

三、数字广告数据要素流通保障技术

保障技术主要分为两类：匿名化技术，例如假名化、泛化、加噪、受控匿名化；隐私计算技术，例如可信执行环境、多方安全计算、联邦学习、可信密态计算。

匿名化技术着重强调去除数据中的个人身份信息、但保留其他信

息，进而达到“既能保护个人隐私、又能让数据用于计算”的目的，但是这一方向的技术存在数据精度损失的问题。近期提出的受控匿名化技术则更好地兼顾了保护个人隐私和保证数据精度。

隐私计算技术能够在保证数据不对外泄露的前提下完成计算，所以，它可以很好地保证数据方的数据利益，这一点是匿名化技术不具备的。此外，在（高安全的）隐私计算过程中，数据的暴露范围并没有扩大。也就是说，“个人隐私信息原来由哪一方持有，还是哪一方持有”，个人隐私信息暴露的范围并没有扩大，因此，隐私计算技术也是个人隐私信息保护的有效手段。实际应用中，隐私计算技术还可以与差分隐私等技术联合，避免从结果中泄露个人隐私。

除了匿名化、隐私计算之外，一些简单的数据裁剪也能达到很好的个人隐私保护效果，比如只传递身份 ID 或者只传递属性信息。前者只泄露了某个人参与了某项事情，在该项事情不具备敏感性的时候，对个人隐私的侵犯很小；后者再结合前述的匿名化方法后，反推出个人身份的难度较大。但这些方式具有一定的局限性，只适用于部分场景。

“跨域管控”指的是“数据方管控自己运维域之外的数据”，前面的几种技术都具备一定的跨域管控能力，但大多集中在计算环节。而大型数据流通中心需要支持丰富的功能、超高且可扩展的性能、灵活的微服务体系等。任一环节的设计失误，都会导致数据方的数据面临威胁。此时，需要有一个体系化的技术理论，来指导实践。本章介绍的跨域管控技术从框架抽象、数据生命周期等入手，全方位地保障

了数据方对域外数据流通的数据的管控。

（一）基础安全技术

1. 典型匿名化技术

匿名化是指个人信息经过处理后无法识别特定自然人且不能复原的过程。匿名化技术通过对数据进行模糊化处理，改变数据颗粒度，进而降低数据的可识别风险。常用技术包括假名化、泛化、加噪等。

（1）假名化

假名化是使用假名替换真值的技术，通常用来处理直接标识符。假名化技术一般通过随机假名分配、散列函数、加密算法来实现，使用过程中需要对假名分配表、散列函数、密钥等辅助信息采取合理的安全保护措施。

在广告行业中，假名化技术能够很好地隐藏用户 ID、身份证号、邮箱等直接标识用户身份的信息。但是假名化技术一般是基于密钥、映射表构建假名和真名之间的连接，数据流通的部分参与者知晓这些因素，因此它们能够从假名中恢复身份信息。假名化技术主要是保护身份类信息，无法保护需要参与计算的属性类信息，对于外部攻击者、部分数据流通参与者，是无法从假名化信息中获得身份信息的。因此，假名化技术更多的是保护用户的个人隐私，对企业的数据利益保护效果较小。

（2）泛化（Generalization）

泛化是将属性值抽象为较一般化、不易区分值的过程。泛化技术

的目标是减少独特记录的个数，使得泛化后的属性值在多个用户中都会出现，从而降低从属性反推用户信息的可能性。

除了对单一属性进行泛化外，目前泛化技术研究还考虑了两种扩展情况：①从多维度属性反推用户信息；②针对流式数据，如何保障用户信息。前者的典型处理技术包括 k-匿名算法等；后者主要是基于扰动、树状结构、伪造值和聚类等构建方案。

在广告行业中，泛化能够减少对外传播的信息的量，进而减少个人隐私暴露以及企业信息泄露的程度。但是泛化也会损失数据精度，导致部分计算任务无法完成。实际使用中，泛化程度的选择常常令人左右为难。

（3）加噪

在原始数据上添加扰动噪声，能够降低攻击者识别出数据主体的可能性。常用的加噪技术为差分隐私，该技术为隐私保护提供了严格可量化的数学定义。差分隐私算法一般通过在特定分布中生成不可预测的随机数的方法实现，同时能够保证加噪结果在真实值附近，保留了数据的统计特征。

按照差分隐私实施的位置，又可以分为本地差分隐私（Local Differential Privacy，简称 LDP）和集中式差分隐私（Central Differential Privacy，简称 CDP）。LDP 一般作用在数据开始计算以及对外传播之前，比如机器模型训练之前；CDP 一般作用在数据完成计算之后、对外传播之前，比如数据库返回聚合信息。

加噪与泛化的作用类似，能一定程度上保护用户信息和企业数据

利益，但同时也面临“加噪程度如何选择的难题”。

2.可信执行环境及可信计算技术

可信执行环境（Trusted Execution Environment，简称 TEE）提供一个基于硬件的隔离运行环境，其隔离性不受任何外部软硬件和运营人员的影响。所以多个参与方可以把数据都放到一个 TEE 中，而不担心 TEE 的物理持有者可以窥探其中的内容。TEE 还提供远程验证能力，通过该技术，远程客户端可以确认 TEE 内执行的代码逻辑，进而可以判断该代码是否会恶意输出数据。与典型的匿名化技术相对，TEE 不需要损失数据精度，可以全面保护各种类型的数据；与 MPC/FL 相比，TEE 在计算过程中不需要密码学计算和网络交互，性能可接近明文。

可信计算（Trusted Computing，简称 TC）指的是计算机系统的行为如预期的计算技术。广义上也包括 TEE，一般情况下特指基于可信平台模块（Trusted Platform Module，简称 TPM）、可信平台控制模块（Trusted Platform Control Module，简称 TPCM）的计算系统度量和验证技术。在恰当的软硬件配合下，比如去除软件的特权账号、防止物理攻击内存等，上述技术可以实现抵御物理持有者的攻击。

在广告行业中，TEE 和 TPM 技术可以用来保护各个场景下的用户数据，对场景和性能的约束很小。它主要缺点是依赖特殊硬件，会引起一些额外的成本，从现实情况来看，部分广告场景数据量大、数据价值密度低，额外的成本市场接受度低。TEE 和 TPM 技术一般需要部署在数据方自己的管理域外，一旦有系统漏洞，数据可能会面临

失窃风险。

3.多方安全计算与联邦学习

多方安全计算（Secure Multi-Party Computation，简称 MPC）通过密码学技术让多方共同计算一个目标且不需要将自己的数据泄露给其他方。MPC 一般会针对每个基础运算设计不同的协议，再通过基础运算的组合实现复杂运算。每个基础运算的协议一般都会伴随着密码学运算和网络交互，所以 MPC 协议一般需要大量的密码学运算和网络交互。

为了探索技术理论，一些弱化的 MPC 算法会在“参与方不会篡改本地逻辑”的假设下进行研究，但这种假设一般与现实不符，通常仅能提供一定的安全性，但无法充分满足现实要求。

联邦学习（Federated Learning，简称 FL）指多个参与方在不交换原始数据的情况下，仅通过交换模型参数和中间结果，完成机器学习训练和预测。与 MPC 相比，FL 存在中间变量泄露、被恶意参与方窃取有价值信息的风险。目前一些 FL 在探索使用同态加密、差分隐私、TEE 等技术缓解上述风险。

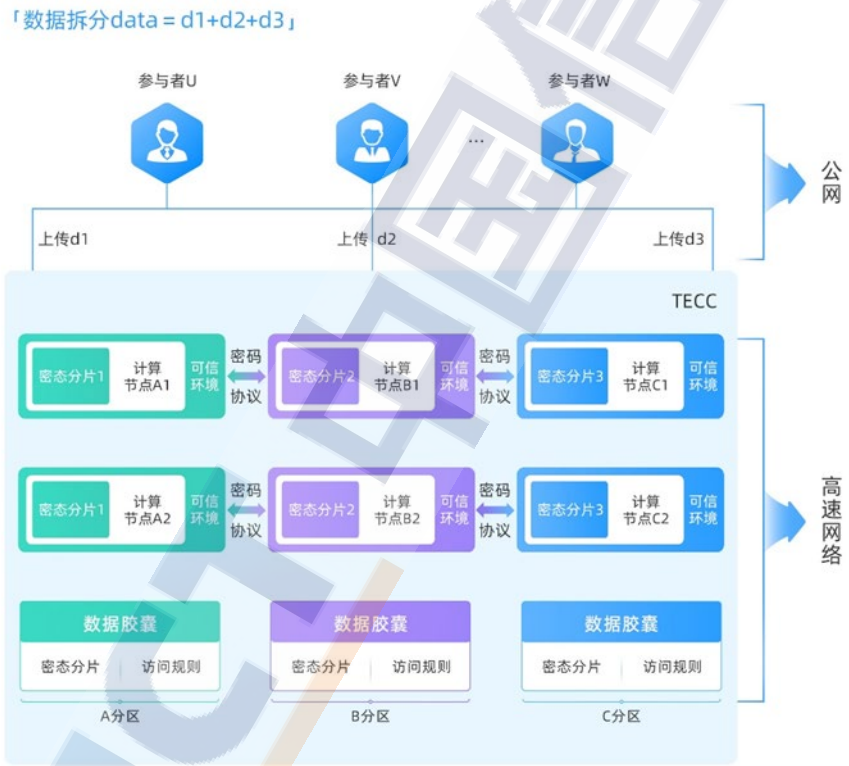
在广告场景中，MPC 和 FL 可用于数据价值大但规模较小的场景，因为 MPC 和 FL 不需要特殊硬件，部署相对灵活。但与此同时，MPC 和 FL 同样也有着“安全性”和“性能”处于跷跷板关系的困境，实际使用中需根据情况甄别。

(二)可信密态计算

可信密态计算（Trusted-Environment-based Cryptographic

Computing，简称 TECC)是指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、存储和流转的一种可信隐私计算技术。可信密态计算既能够抵御常见的安全隐患又能够快速处理大规模数据。具体表现在，它能够缓解供应链攻击、系统漏洞等常见的硬件安全隐患，以及抵御合谋攻击、恶意敌手攻击等常见的密码协议攻击。同时，它不受公网传输瓶颈和复杂的密码计算拖累，没有显著的性能瓶颈。

图 2 是 TECC 典型的原理示例图:



来源：蚂蚁科技集团股份有限公司

图 2 TECC 原理示意图

● 数据提供者在本地将数据拆分成多个密态分片数据，并将每个密态分片数据传递给不同分区的可信计算节点。这里单个密态分片数据不会泄露原始数据的任何信息。

● 每个分区的可信计算节点只接触一份密态分片数据，不接触任何明文数据。多个分区的可信计算节点通过密码协议（MPC、安全联邦学习等）完成目标计算，单一分区被攻破不会产生数据泄露风险。

● 可信计算节点使用可信计算技术（TEE/TPM/全栈可信等），保证运营者无法进行窥探。

● 密码协议的同一个角色由一个可信计算分区集群承担，计算资源可以进行动态水平伸缩。

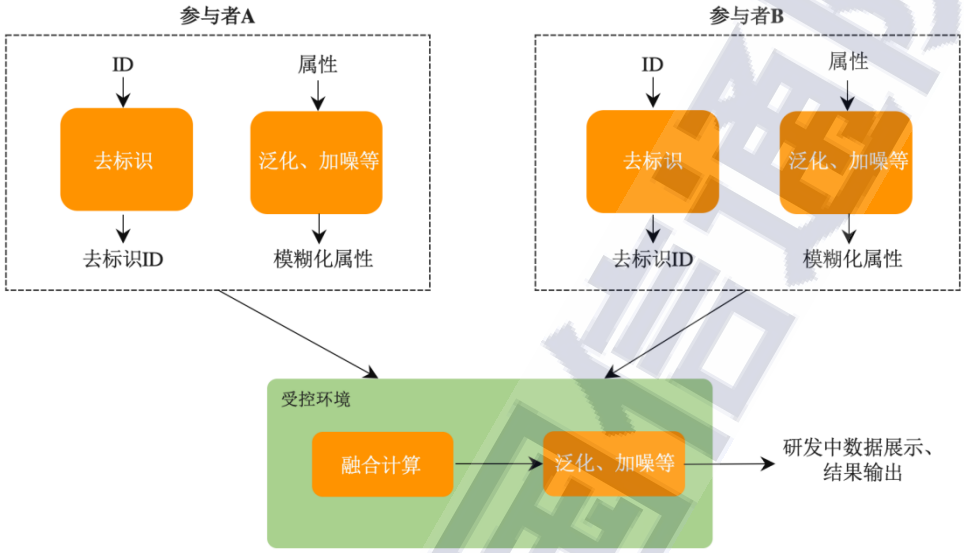
● 数据采用密态胶囊形式进行存储，包括密态分片数据以及与其绑定的访问规则，运营者无法滥用密态数据。

在广告行业中，TECC 可以用于构建面向广大机构的数据流通的平台，一方面让中小企业能够享受到高安全的数据流通服务，另一方面通过多场景的共享，平摊 TECC 的建设成本。

（三）受控匿名化

绝对匿名化是指任何情况下都无法识别特定自然人且不能复原，但科学无法证明未知，因此绝对匿名化难以被有效证明。在具体实践中，相对匿名化会是更加切实可行的方案。相对匿名化是指个人信息经过处理，在不结合额外信息、在经典算力和合理时间范围内，无法识别特定自然人且不能复原的技术。受控匿名化是指将相对匿名化的数据限制在受控环境中使用，以确保在受控环境中，达到无法识别特定自然人且不能复原的匿名化效果。通过严格管控受控环境与外界的交互，进而满足了相对匿名化的限制条件。

受控匿名化技术中，各参与方首先在本地对数据进行去标识和模糊化处理，并且数据的后续处理和使用也受到严格的管控，通过受控环境来限制其与外部的交互。图 3 所示为典型的受控匿名化流程。



来源：公开资料整理

图 3 受控匿名化流程

在参与方本地，需要分别对标识符（Identifier，简称 ID）和属性进行处理。首先，参与方在本地对 ID 进行去标识。为保证 ID 能够支持融合碰撞计算，一般通过 HMAC 或确定性加密算法实现去标识。其次，参与方对属性信息进行模糊化处理，可采用泛化、加噪等方式，通过对属性的模糊处理能够进一步降低数据的可识别性。最后，所有参与方完成数据处理后，将数据传输到受控环境中进行融合计算，相对匿名化的数据在离开参与方后仅出现在受控环境中，在研发中的数据展示和结果输出中，需要对数据进行泛化、加噪等模糊化处理，以避免通过结果推断原始数据。

受控环境通过可信计算等安全保障技术，可实现受控环境与外部

通道交互的严格管控，降低了数据泄露的风险。此外，受控匿名化极大程度地保留了 ID 和属性的数据价值，且计算量小性能高，能够适用于较多数据碰撞、融合计算等场景中。

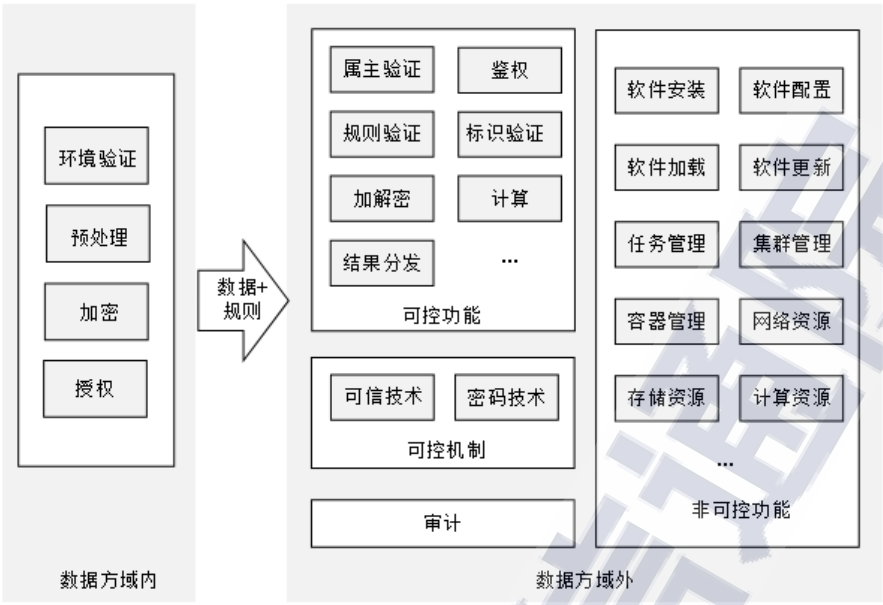
在广告产业中，受控匿名化可以保障数据流通过程中用户的隐私，尤其是同时需要流通身份、属性信息的场景，且由于受控匿名化方案性能损耗小，能够适用大规模或实时运算场景。例如在广告程序化交易流程中，供应方将其用户的 User ID、IP 地址和广告位信息发送给广告主，广告主结合其已有的用户信息，给出更为精准的报价，提升资金利用效率。

对受控环境的审核是合规的重要环节。由于广告行业数据传输链路非常多，全部自建受控环境审核压力较大，因此，部分自建、部分适用公共服务，也是平衡上述问题的一种选择。

（四）跨域管控技术

在数据流通场景中，跨域管控是指数据离开持有者（也叫数据方、数据持有方）的运维域后，数据方仍然能够有效地控制数据的流转过程，避免其被窃取或者非预期地使用。

图 4 是跨域管控技术通用逻辑的抽象示意图。一部分工作需要数据方亲自进行（在数据方域内进行），包括：1）验证数据方域外的环境，以确认该环境是否安全；2）对数据做预处理，以满足后续处理的格式要求，或者减少对外传递的信息量；3）对数据进行加密，并且保证只有前述验证过的环境才能解密；4）当有其他方请求数据时，要对数据进行授权。



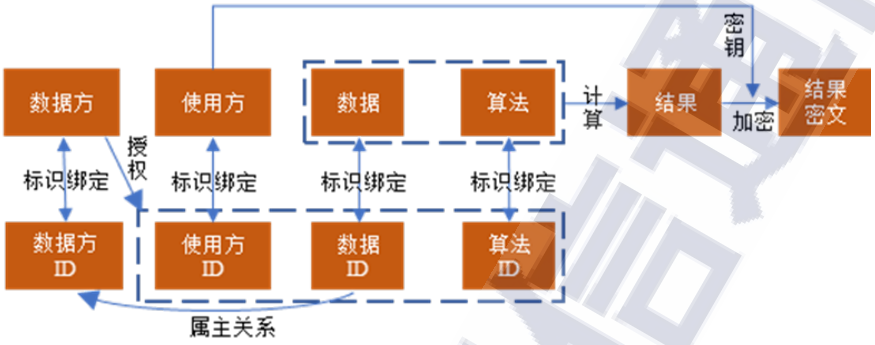
来源：公开资料整理

图 4 数据跨域管控示意图

在数据方域外，要有相应的机制，提供数据跨域管控的底层基础，如图中的“可控机制”；在此基础上，设计支持数据生命周期管理的相应技术功能，如图中的“可控功能”，这里需要指出的是，跨域可控技术里面的数据生命周期比一般的数据生命周期要更细致，因为任何一个细小的生命周期设计不当，都有可能导致数据泄露，除此以外，数据生命周期的相互转换，也需要进行周密的设计，否则也有可能被攻击者利用。部分域外的功能可以由域外运维者自主控制，叫做非可控功能，包括各种资源的管理、任务的管理、任务与资源的映射、软件环境的维护等。这一灵活度可以很好地提升资源利用率和保障系统稳定性。

图 5 是一个“如何将参与方、ID、数据、计算、结果严密地绑定在一起”的示例，避免因设计不周密导致数据受威胁。假设整个数据生命周期在可信环境中进行，但是由于采用微服务体系，不同

的生命周期可能被分散到不同的可信环境中，数据也是通过可信环境的密钥加密后存储在可信环境之外。其中的标识、属主关系等在“非数据流通场景”通常不需要施加安全保护措施，也是容易在数据流通方案设计中被忽略的地方。主要原理如下:



来源：公开资料整理

图 5 各实体及其间的绑定关系

● 授权：数据方声明“谁能够对我的哪份数据使用什么算法进行计算”，声明的方式一般采用数字签名。

● 标识：声明消息中的三个实体一般采用 ID 进行标识。所以必须要有可靠的机制确保 ID 和其背后的实体关联起来，并且关联程度要能够防止域外的恶意者进行破坏。一种可行的实现方法是采用身份公钥的哈希值作为身份 ID、采用数据的哈希值作为数据 ID、采用算法的哈希值作为算法 ID。

● 属主关系：在采纳数据方声明的授权之前，还需要确认“该数据”是否属于“该数据方”，并且这一关联关系不能被域外的恶意者破坏。一种可行的实现方法为，数据方通过密码学手段证实自己拥有解密该数据的密钥。

在广告行业中，跨域管控技术可以助力构建大规模的数据流转中

心，使得大规模数据中心既可以具备丰富的功能、高的性能和灵活性，同时也避免了因为这些能力的增加，导致数据方失去对自己数据的管控。从而极大地提升数据方参与数据流通的意愿。

四、数字广告数据要素流通保障平台实践探索

(一)基于可信密态技术的广告数据流通平台

1.平台设计思路

（1）流通模型

基于可信密态技术的广告数据流通平台数据流通链路中的主要角色分为数据提供方、数据使用方、数据经营方、平台管理方:

● 数据提供方/数据持有方，简称“数据方”，是为数据流通提供数据的一方。

● 数据使用方，简称“使用方”，是使用数据的一方。

● 平台管理方，简称“平台方”，是运营数据流通平台的一方。

● 数据经营方，简称“经营方”，是代替数据方进行数据经营的一方。

与之相对应，数据流通平台也设计了三种权限: 数据资源持有权、数据加工使用权和数据产品经营权。

● 数据资源持有权，简称“持有权”，数据在数据流通平台中的最高权限，可以决定其他方如何使用该数据。

● 数据加工使用权，简称“使用权”，加工使用数据的权力，能够使用的范围是持有方决定的。

● 数据产品经营权，简称“经营权”，可以代数据方向其他方授予使用权，但自身不包含持有权和使用权。

● 平台管理方不拥有上述数据权限，只能运行维护平台。

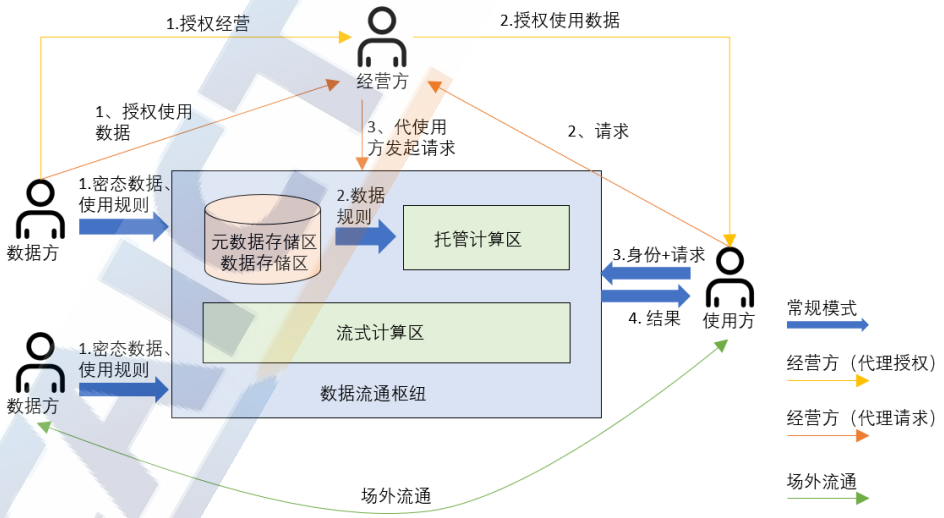
图 6 展示了在上述角色和权限的设计下，可以实施的几种流通模式。

● 常规模式下，数据方直接向使用方授权，获得授权后的使用方向平台发起请求。

● 代理授权模式下，数据方向经营方授予经营权，经营方向使用方授予使用权。

● 代理请求模式下，数据方向经营方授予使用权，使用方将请求发给经营方，由经营方代为发起请求。

● 场外流通模式下，经营方在场外撮合数据方和使用方，之后按照常规模式运行。



来源：中国信息通信研究院

图 6 数据流通平台示意图

（2）安全模型

由于数据具有独特性、可复制性、高价值等特征，在实际的数据流通过程中的各参与方都有窃取他人数据的动机，在系统设计中应按照其他方存在窃取的动机去设计，以适应更多的情况，一定程度上规避风险。

情况稍微特殊的是平台的管理方，平台的管理方可以选择信誉度高的机构承担。但即便如此，也可能因为被系统入侵、少量员工作恶等，导致平台的管理方做出恶意行为。为此，在系统设计时，我们假设平台管理方也可能是恶意的，这样可以在上述突发情况下依旧保障数据方数据的安全。

为了支持丰富的功能，数据流通平台中包含了多个数据生命周期，这些生命周期及生命周期之间的衔接都必须进行充分的安全保障。也就是说，即使平台管理方是恶意的，以下安全性质也能够得到保障：

- 数据存储：存储数据的机密性和完整性不会被破坏。
- 数据属主：数据和属主的对应关系不会被破坏。
- 标识：标识（ID）与“其背后实体、实体关联的资源”的对应关系不会被破坏。
- 鉴权和规则验证：无法绕过鉴权和规则验证，发起计算、结果分发。
- 计算：无法通过计算过程窥探数据。
- 结果分发：不会将结果分发给非预期方。
- 上述功能要能够无缝地衔接在一起：无法通过干扰功能切换间隙，进行攻击。

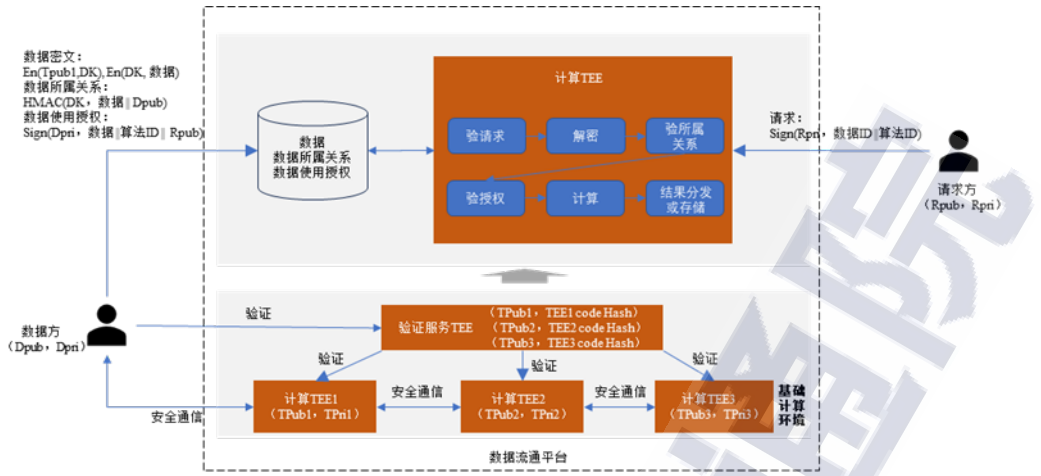
（3）计算引擎选型

在数据广告数据流通实际业务场景中，数据流通如果采用频繁跨越公网的一些密码协议（如 MPC 方案），则可能在性能（如延时性、QPS 等）和成本上无法满足诉求，因此，数据流通平台着重选择 TEE、TECC 等集中式的数据流通技术。TEE 在隔离环境中进行明文运算，隔离环境会防止来自外部的攻击，明文运算使得 TEE 能够提供更高的性能；TECC 在隔离环境中运行密文运算，由于其内在性质，能够抵御常见的密码协议攻击、典型硬件漏洞隐患，密文运算使得 TECC 安全性更高。TEE 和 TECC 相互配合能为用户提供更强的安全性和更多的性能选择。

（4）跨域管控落地实践

数据流通平台的目标是成为大型的数据流通中心，既要确保数据持有方对数据拥有管控力，也要具备数据开发行为的多样性能，挖掘数据要素潜在价值。为此数据流通平台实践了前述的跨域管控技术。

图 7 是数据流通平台原理示意图：数据流通平台具有大量 TEE 组成的集群，便于数据方能够对整个集群进行验证以及集群不同节点之间相互验证，平台中包含了一种专门提供验证服务的 TEE，称为“验证服务 TEE”。每个 TEE 在启动时，会向“验证服务 TEE”注册自己的代码哈希值和公钥，“验证服务 TEE”使用远程认证机制验证过该消息真实性后，接受该消息；之后，数据方、其他 TEE 节点就可以利用“验证服务 TEE”获得其他 TEE 的准确信息，验证服务本身也在 TEE 之中以防止外部篡改其内部信息。



来源：中国信息通信研究院

图 7 数据流通平台原理示意图

数据方提供的信息包括三部分：数据密文、数据所属关系和数据使用授权。

● 数据密文一般采用数据方认可的 TEE 的公钥加密，确保只有该 TEE 能够解密。图 7 中采用信封加密，即公钥加密数据密钥，数据密钥再加密数据。

● 数据所属关系用于证实“数据是谁的”，以免攻击者篡改存储在平台中的数据和其所有者之间的关系。实践中，可以使用数据密钥对数据和所有者公钥签名，因为“知道数据密钥的人，一定是拥有数据的人”，所以数据密钥的签名可以代表数据所有者。

● 数据使用授权，即使用数据方私钥对“谁可以使用数据进行什么运算”进行签名。需要注意的是，数据所属关系正是证实为什么该私钥的签名是有效力的基础。

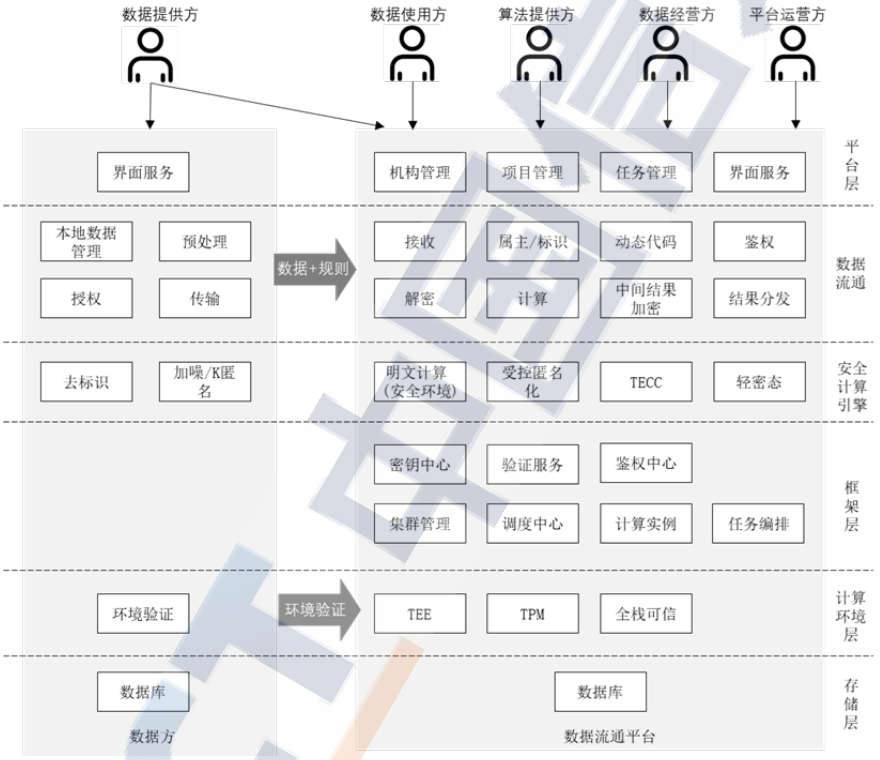
请求方对其要使用的数据 ID、运算进行签名。

真正发生计算的 TEE，首先验证请求方的请求的签名是否正确，

然后解密数据，验证数据与其所有者之间的对应关系，再验证数据所有者是否对上述请求进行授权，如果全部验证通过，则进行计算。计算完成后，再将计算结果使用请求方公钥进行加密，返回；或者使用 TEE 专属的密钥，保存在数据流通平台。

2.平台技术架构

数据流通平台分为以下几个部分，具体见图 8。



来源：中国信息通信研究院

图 8 数据流通平台技术架构图

存储层：数据流通平台的核心能力在“流通”上，存储是其可选的功能，所以数据方本地建议保留存储能力。数据流通平台的存储能力，是为了避免数据方反复上传同样的数据，浪费时间和资源。

计算环境层：数据流通平台使用 TEE、TPM 和全栈可信，为数

据方提供一个能够抵御恶意人员攻击的域外运行环境。数据方在使用该环境前，需要从远程验证该环境，以免将数据传入到虚假的安全环境中。

框架层：其底层为基础的计算框架，包括集群管理、调度中心、单个计算实例的基础环境，以及大数据计算的一些优化能力，比如任务编排等。除此之外，还包括一些基础服务，比如密钥管理服务、协助环境验证的服务、数据访问权限鉴别服务等。

安全计算引擎：安全计算引擎能够提供隔离的计算环境，在提供计算能力的同时防止攻击者在计算环节获得敏感信息。安全计算引擎提供多种计算引擎，使得用户可以根据不同场景选择合适的计算引擎，包括受控匿名化、TECC、TEE、轻量级密态计算（适度放弃安全性换取高性能的密态计算算法）。数据方本地也具有一定程度的密态化能力，包括去标识、加噪、K 匿名等，与服务端的能力相互辅助。

数据流通层：数据流通层包含数据流通的生命周期过程。1）在数据方本地，首先对数据进行管理（包括增加、删除等），然后对要流通的数据按规定进行预处理，并进行加密传输；2）流通平台在接收数据后，进行加密存储；3）流通平台在收到请求后，依次验证数据与数据方的所属关系、数据方对使用方的授权、ID 与实体的绑定关系等，验证完成后进行相应的计算；4）计算的结果可以加密留存在数据流通平台，或者返回给数据使用方。因为数据的全生命周期流动都使用跨域管控技术保障数据方对数据的全程可控，所以即使平台方是恶意的也不破坏该可控性。

平台层：平台层包括机构管理（机构的注册、审核等）、项目管理（项目的建立、成员准入等）以及任务管理（任务的配置、发起等）。数据流通平台提供用户可操作的界面，为达到“平台管理方恶意的情况下也能保证数据安全”，数据方不能完全信赖数据流通平台（服务端）提供的界面服务，其本地也必须拥有界面访问服务。本地界面服务作为可信的媒介，将能保证数据方的操作安全传导到其数据和密钥上。

3.平台方案分析

数据流通平台基于数据生命周期，涵盖各类数据要素流通角色、支撑多种安全性和具有性能差异的计算引擎、具有可弹性扩展的分布式计算能力，为海量数据流通提供公共平台，非常适合大规模数据流通，包括为业界大量机构提供数据流通的公共平台。

数据流通平台使用了多种技术，突破现有的安全性、性能等瓶颈，使得数据流通适用场景大规模拓展：

● 受控匿名化技术突破了隐私保护和数据价值难以兼得的瓶颈，使得用户个人隐私保护、数据价值无损失成为可能。

● 可信密态计算突破了安全性和性能难以兼顾的困境，使得高安全、高性能的数据流通成为可能。

● 跨域管控技术使得数据方可以基于技术手段对域外的数据进行管控，支持丰富的数据生命周期、灵活的性能扩展，有力地支撑了大规模数据流通系统。

(二)数据匿名化实施服务平台

1.平台设计思路

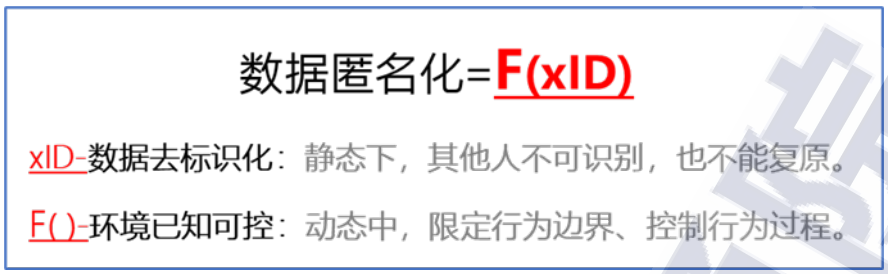
《中华人民共和国个人信息保护法》第四条规定：“个人信息是以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息，不包括匿名化处理后的信息。”

从平衡保护与利用角度，由于数据处理环节多向多样，且其中潜在的数据价值的发现和实现是在“共享”中达成的，但缺少用户交互能力、机构与机构间的“共享”处理行为。因此，可利用匿名机制，既保持数据的匿名化状态，也保证即使泄露也无法识别具体个人甚至影响到个人。

国家标准 GB/T 35273-2020《信息安全技术 个人信息安全规范》和国家标准 GB/T 37964-2019《信息安全技术 个人信息去标识化指南》中，去标识化作为一种技术，是以实现匿名化为目的的多种技术或技术组合（如假名、加密、泛化等）；在无其他影响因子的情况下，以上适当的去标识化技术可以证明其所处理形成的静态的数据符合匿名化要求。

但当数据进入流动利用的活动中，“活动中的数据”不仅有控制人的变化、数据形态的变化（如隐私计算等加密形态），也会有数据内容的变化（如添加随机标识、泛化标签）。对于活动(变化)中的数据，技术已无法独立证明是否符合匿名化要求，需要配套相应的业务法律边界（评估见证）和运营过程控制（分域监控）的服务控制体系，通过对各种“数据活动”所构成的影响因子进行约束和控制，才能证

明活动中的动态的数据是否继续符合匿名化要求。



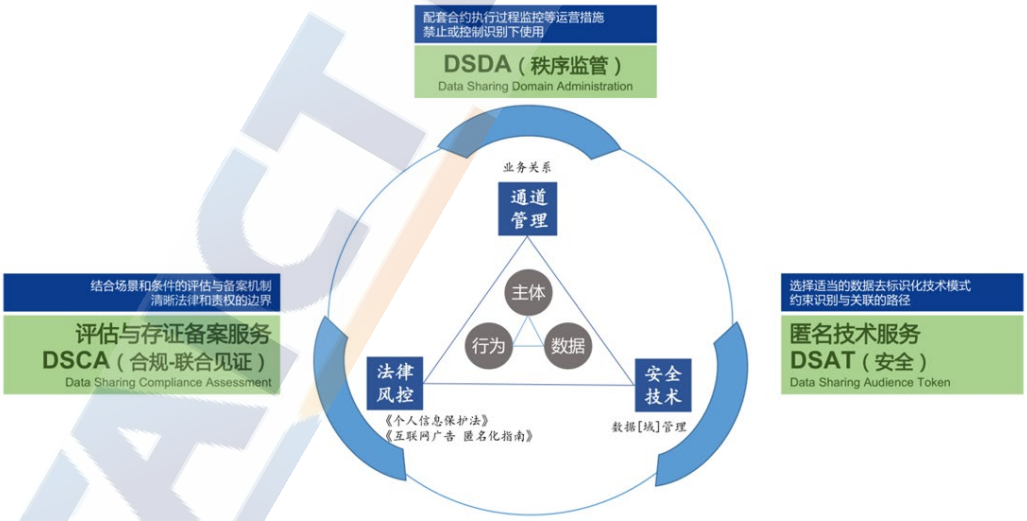
来源：中国广告协会

图 9 数据匿名化实施思路

如图 9 所示，数据匿名化实施基于有效的数据去标识化 （匿名预处理 xID）技术，配套相应的业务法律边界评估见证的方法和运营过程监控的管理控制体系 F( )，最终能够使数据遵循控制者的意愿，通过合法有序的方式实现社会化的充分利用。

2.平台技术架构

数据匿名化实施服务平台主要包含三个服务，具体见图 10。



来源：中国广告协会

图 10 数据匿名化实施服务平台架构图

匿名技术服务：选择适当的数据去标识化技术模式，构成数据的

匿名。数据控制者对信息主体标识(符)处理得到 Token 标记，相同信息主体标识(符)在不同数据控制者分域内生成的 Token 标记不可逆且互不相识。并由本服务平台作为独立第三方，依据数据交换合约的合规评估意见，管理分域密钥，开关控制分域 Token 标记间的关联。采用的分域去标识化技术，可达到如下效果:

● 不可逆推：保证预处理的数据(集)无法直接还原出个体标识(符)等识别信息。

● 互不相识: 保证预处理的数据(集)无法被他人间接还原出个体标识(符)等识别信息。

● 关联受控：可遵循业务法律边界，禁止或控制所处理的数据与特定个人的关联，具备防范滥用匿名化和无序使用数据的能力。

● 有效证明：技术本身应通过国密认证，技术执行可留存清晰的记录以供评估监控审计等形成对应法律证据。

评估与存证备案服务：结合场景和条件的评估与备案机制，清晰行为的边界，遵循 T/CCSA 424-2022 | T/CAAAD 004-2022《互联网广告 匿名化实施指南》要求，通过合规自测、合约评估和备案公示，提供以行业内数据交换合约为对象的合规性评估与存证备案服务。合规自测通过对 200 多项的评估因子自动检测，帮助客户快速找到业务场景中的合规风险点以及合规路径; 合约评估由具有数据合规和安全评估能力的独立第三方服务机构以及职业律师提供服务实施评估，形成具有真实性、有效性、相关性的评估结论与法律意见书样本; 备案公示通过提供备案证明，证明已符合最佳实践。

秩序监管服务：配套合约执行过程监控等运营措施，控制主体的使用。

3.平台方案分析

数据匿名化实施服务平台具有轻量级、无需部署的优势，同时还有专业服务团队作为支撑，能够便捷使用。该服务平台融合权威第三方 xID 技术、知名律师事务所法律评估见证，以及行业组织的监督管理，能够进行符合 T/CCSA 424-2022 | T/CAAAD 004-2022《互联网广告 匿名化实施指南》要求的数据匿名化实施并有效存证。

- 打破传统的单一驱动模式，打通企业法务、业务和技术的诉求，将企业内部的单向妥协（零和博弈）扭转为多赢局面。

- 提供“安全+合规+有序”的三位一体解决方案，增强互联网广告行业竞争力，发挥数据要素的商业价值。

五、数字广告数据要素流通发展建议

(一) 加强政策引领，推动社会共建共治

以政策引领数字广告数据要素流通建设方向：**一是**在国家数据局统筹指导下，以“数据二十条”为基础，进一步完善数字广告领域数据要素的顶层设计，制定数字广告数据要素指导文件，奠定数字广告数据流通的基础；**二是**加强对《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律在互联网广告领域数据要素流通的司法解释工作，加强相关法律法规宣传教育，推动树立广告行业法治意识；**三是**在总体国家安全观指引下，不断推进数据安全保护工作，构筑起维护数据安全和促进数字经济健康持续发展的法律制度保障；

四是加强数字广告领域行业组织的协调引领作用，加强行业自律和业务指导，形成政府监管、行业自律、企业自治、社会监督的社会共治模式，促进数字广告产业的高质量发展。

（二）强化标准指导，完善数据流通体制机制

以标准指导数字广告数据要素流通方案落地：**一是**实现统筹规划，建立和完善数字广告数据要素流通相关标准体系，以标准化手段指导数字广告数据要素流通工作的体系化建设和业务推进；**二是**制定数据要素流通平台相关技术标准，从设计、开发、测试、部署、监测全流程推动数字广告数据要素流通平台范化运营、精细化管理、有序化发展；**三是**制定数字广告数据要素流通测评方法和应用指南，制定科学评估体系和应用建议，建立健全评估与监督机制，严格落实政策要求。

（三）坚持守正创新，加强技术研究与应用

以技术夯实数字广告数据要素流通应用基础：**一是**加强推进兼顾数据流通效率和数据安全技术的研发和应用，以技术手段推动数字广告数据合规且高效地流通，如对主体明确的数据，在流转过程中应采用技术手段明确用户授权并对过程记录留痕，保证数据来源的合规性和数据流转链路的完整性；**二是**对于数字广告业务开展过程中沉淀的海量泛众数据，应采用相关安全技术手段使数据以匿名化形式进行流转，使数据加工计算方无法识别特定主体，实现数据可用不可见和可算不可识等目标；**三是**探索不同模式的数据共享机制和模式，通过应用匿名化等技术，结合具体场景，综合运用合规加技术措施，实现技术的可理解性和透明性，加强数字广告相关方的信任感。

（四）完善基础设施，推进平台建设与推广

以生态共建数字广告数据要素流通基础设施：**一是**推进广告主、平台、监测公司等多方参与者资源共享、纵横延伸，以数据安全为基础，建立开放互补、融合发展的数据要素流通平台；**二是**行业组织发挥组织协调能力，建立公共的数据要素流通平台，释放中小企业持有数据的资产价值，帮助中小企业加强数据流通意愿，促进中小企业参与数据要素流通的能力；**三是**行业龙头企业率先发挥引领作用，形成新老帮扶、差异结合的合作模式，引导不同类型、不同规模的数字广告企业资源共享数据、维护数据流通秩序、筑建数据安全防护网。

（五）开展测试评估，加强平台审核与管理

以测评提升数字广告数据要素流通服务质量：**一是**对数字广告业务链上下游不同企业各平台进行能力测评，如需求侧平台、供给侧平台、数据管理平台、第三方技术服务商等，保障平台的业务能力；**二是**根据不同场景下的数据使用需求，对数字广告生态相关方安全能力、数据流通解决方案进行测试评估，提升安全水位；**三是**推动第三方评估机制建立，对现有市面发布的数据流通平台进行安全能力审核，达成以评促建、以评促管、以评促提升的目标。

编制说明

本报告编写参与单位：中国信息通信研究院泰尔终端实验室、蚂蚁科技集团股份有限公司、南京大学数据管理创新研究中心、国家广告研究院、北京师范大学法学院、北京沃东天骏信息技术有限公司、北京腾云天下科技有限公司、北京市环球律师事务所、深圳市腾讯计算机系统有限公司、秒针信息技术有限公司、上海优比客思科技有限公司。

中国信息通信研究院 泰尔终端实验室

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62309656

传真：010-62304364

网址：www.caict.ac.cn