隐私计算与公共数据开放白皮书

2022

发布单位

- 数字中国研究院（福建）
- 北京数牍科技有限公司
- 复旦大学数字与移动治理实验室

CONTENTS

一、引言

数据是数字化发展的基本要素。进入数字时代，人类获取、管理和利用数据的能力空前提升，社会各界对数据的价值也愈发重视。2020 年 4 月，中共中央、国务院发布的《关于构建更加完善的要素市场化配置体制机制的意见》明确提出将数据作为一种新型生产要素，与土地、劳动力、资本、技术等传统要素并列，要求“加快培育数据要素市场，推进政府数据开放共享、提升社会数据资源价值、加强数据资源整合和安全保护”。

数字化发展需要整合和利用各种来源的数据，而政府部门在履职过程中生成、获取和保存了大量基础性、关键性的数据资源，是一个国家最重要的数据保有者。在工业时代，政府在大型公共基础设施的建设中发挥了主导作用；在数字时代，公共数据作为一种新的基础设施，将和物理基础设施同等重要。在不涉及国家秘密、商业秘密、个人隐私的前提下，把公共数据开放给社会进行融合利用，将有力促进数字经济和数字社会的发展。因此，开放公共数据，构筑公共数据基础设施，是数字化发展的现实需要，也是政府在数字时代的责任。

近年来，公共数据开放已成为国家政策的重要关切。2020 年 4 月发布的《关于构建更加完善的要素市场化配置体制机制的意见》要求推进政府数据开放共享，研究建立促进公共数据开放和数据资源有效流动的制度规范。2021 年 3 月发布的《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》中提出要“扩大基础公共信息数据安全有序开放，探索将公共数据服务纳入公共服务体系，构建统一的国家公共数据开放平台和开发利用端口，优先推动企业登记监管、卫生、交通、气象等高价值数据集向社会开放”。2021 年 12 月，国务院印发的《“十四五”数字经济发展规划》要求“建立健全国家公共数据资源体系，统筹公共数据资源开发利用，推动基础公共数据安全有序开放，提升公共数据开放水平，释放数据红利”。2022 年 6 月，国务院印发的《关于加强数字政府建设的指导意见》也要求“编制公共数据开放目录及相关责任清单，构建统一规范、互联互通、安全可控的国家公共数据开放平台，分类分级开放公共数据，有序推动公共数据资源开发利用，提升各行业各领域运用公共数据推动经济社会发展的能力”。

然而，公共数据开放的价值与风险并存。一方面，公共数据开放能够释放经济、社会、政治价值，另一方面，公共数据开放也存在诸多潜在风险。一是在数据存储和流通过程中存在数据泄露的风险，可能危及国家秘密、商业机密和个人隐私。二是开放数据面临被不合规利用的风险，数据被误用或滥用会损害公共利益和第三方利益。因此，在数字化发展的背景下，如何安全合规地推动数据的高质量供给与有序利用就成为了重要且紧迫的议题。

为应对公共数据开放与利用中的安全风险，各地政府已从管理体制和技术手段方面展开了诸多探索实践，如建立全生命周期安全管理机制，推进公共数据分级分类开放，以及应用数据加密、数据脱敏、数字水印、数据沙箱等技术手段加强安全保障等。然而，公共数据开放在数据安全保护方面仍存在一些未能完全解决的问题，阻碍了高价值公共数据的开放。

隐私计算作为一种新兴的数据安全技术，有望在保护多个参与主体的数据本身不对外泄露的前提下，实现数据融合分析计算与价值挖掘，通过“原始数据不出域”、“数据可用不可见”以及“数据用法用量可控可计量”等特性，显著降低公共数据开放与利用的风险，成为平衡公共数据开放价值释放与风险管控的助推器，以此推动公共部门开放更多的高质量数据，促进市场和社会的数据利用。

二、 公共数据开放的现状与挑战

（一）现状与问题

我国的公共数据开放起步于地方自主探索。2012年6月，上海市政府数据服务网“datashanghai. gov. cn”（原网址）上线运行，标志着我国内地的公共数据开放实践拉开序幕。其后，全国各地相继上线公共数据开放平台。截至 2021年10月，我国已有193 个省级和城市的地方政府上线了数据开放平台，其中省级平台20个（含省和自治区，不包括直辖市和港澳台），占全部省级地方的71.43% ；城市平台 173 个（含直辖市、副省级与地级行政区），占全部城市的51.33%，如图1所示 。

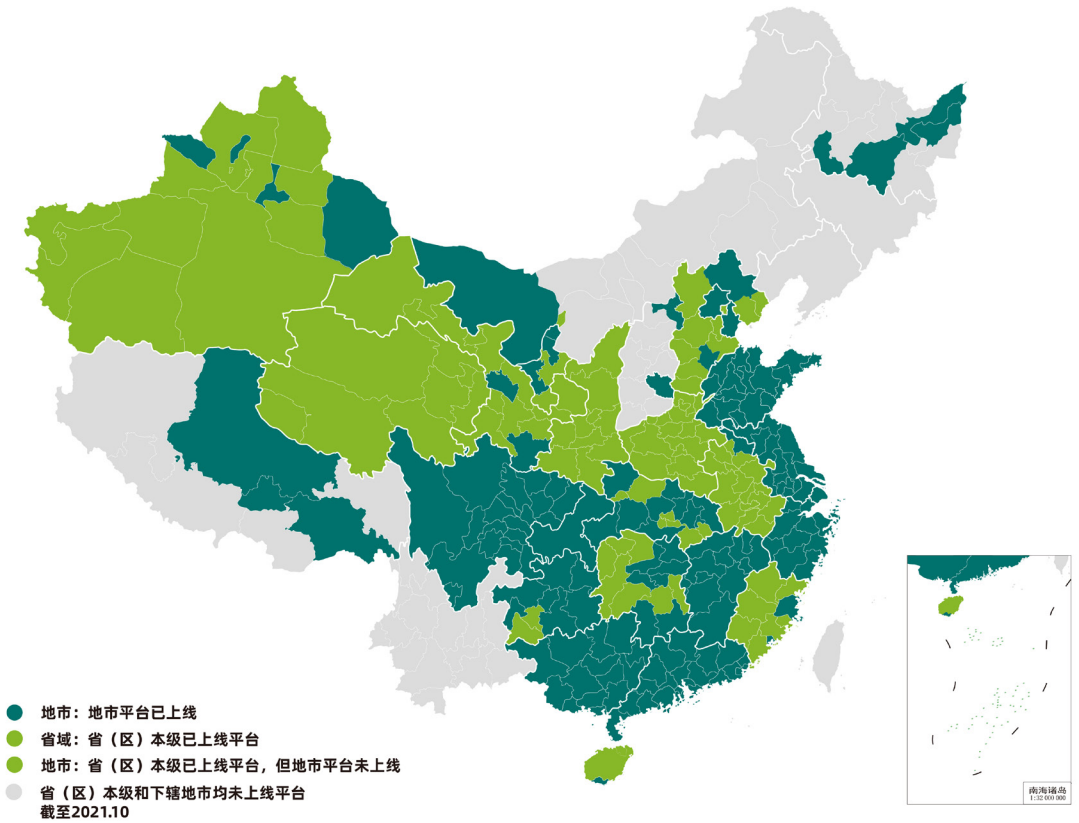

图 1　中国各省域平台整体上线情况地理空间分布

近年来，我国公共数据开放水平正在逐步提升。在制度供给方面，与公共数据开放相关的法律法规、实施细则、标准规范等陆续出台和完善。在平台建设方面，各地公共数据开放平台的功能逐渐扩展和优化，运营维护能力与用户实际体验也在不断提升。在数据供给方面，各地开放数据的数量与质量也在逐步提升。在利用生态方面，各种利用促进活动正在开展和推进，企业、公众等各种社会主体越来越多地参与利用开放数据，也产出了一定数量的利用成果。

然而，推进我国公共数据高质量开放利用，仍存在许多问题和短板。总体上，市场和社会对公共数据的需求尚未得到充分满足，表现为开放数据数量不多、容量较低、质量不高，已开放数据普遍存在字段少、条数少、颗粒度较粗等问题，以 API 接口形式开放的实时、动态、高容量数据尤为稀少。此外，在已开放数据中还存在高缺失、低容量、碎片化等低质量数据，数据利用价值较低；而数据供给端的不足还造成了利用端的成果数量少、质量不高等问题。

（二）困境与挑战

（1）数据开放与数据安全之间的平衡困境

公共数据供给不足的问题在较大程度上受制于数据开放与数据安全之间的平衡困境。国家政策与社会公众对数据安全的重视程度不断提升，保护公共数据安全已成为开放与利用数据的前提。有关数据安全的法律政策体系也逐步建立起来。在顶层设计上，《中华人民共和国网络安全法》、《中华人民共和国数据安全法》以及《中华人民共和国个人信息保护法》从不同程度上提出要在保障数据安全和个人信息的前提下促进公共数据的开放利用；在实践规划上，《“十四五”大数据产业发展规划》等政策文件中均强调了在数据流通过程中保障数据安全的重要性。

虽然数据只有被利用了才能产生价值，但由于数据开放和利用过程中存在的各类安全隐患增加了相关部门所承担的风险，数据泄露、隐私风险等安全隐患的客观存在降低了数据提供部门的开放数据的意愿和动力，导致数据供给部门在数据数量、数据质量、数据开放范围等方面较为保守， 未能充分满足社会对公共数据开放利用的需求。

（2）数据开放面临的各类安全风险和顾虑

《“十四五”国家信息化规划》将数据全生命周期的安全管理划分为数据收集、汇聚、存储、流通、应用等阶段，公共数据在开放利用的各个阶段均存在着不同类型的安全风险与挑战：

第一，数据汇聚与存储中的风险。当前我国公共数据开放主要采用“部门数据供给 - 数据资源汇集 - 数据平台开放”的形式，供给数据过程中的可见性存在数据泄露、数据篡改、数据被重新识别等隐患，也对平台方内部工作人员的素养和技能提出了较高的要求；而数据资源的汇聚使得部分数据管理的主动权从数据提供部门转移到数据开放平台的管理方，而数据平台又存在被恶意攻击、数据泄露等安全隐患，数据平台的安全防护面临较大压力。

第二，数据流通与利用中的风险。这类风险主要体现在两个方面：一是在数据提供部门在将数据提供到数据平台后，对部分数据缺乏监测与控制，无法对数据利用过程及结果进行追踪，难以监管数据应用的合规性与正当性，且现行脱敏技术手段并不能解决所有的隐私问题，在数据利用过程中仍有通过数据拼凑还原个体数据的可能，使个人信息面临泄露风险。二是对于有较高安全级别的数据，我国目前多采用对满足申请条件的用户以有条件的、受限的方式进行开放，而由于缺乏对数据利用场景的监测，存在数据由满足申请条件的用户流通至不满足申请条件的用户从而导致数据泄露和违规利用的风险。目前国家尚没有制定统一的公共数据分类分级标准，各地方部门对数据安全的管控手段主要还是以严格控制数据出域和严格制定数据获取条件为主，也因此间接导致了数据供给不足和数据获取门槛过高等问题。

三、公共数据开放利用中的现有安全管理方式

为了应对上述公共数据开放利用中的安全问题，当前各地方政府在保障公共数据安全开放方面已经开展了诸多实践探索。数据安全保障举措一方面聚焦于管理体制的建立健全上，另一方面则体现为对新兴数据安全技术的积极运用上。

（一）体制机制

数据安全，制度先行，公共数据开放的安全保障需要依赖管理体制机制的不断完善。各地针对数据开放全生命周期的安全风险，围绕事前的数据安全处理、事中的数据安全监控与事后的行为处置等方面都探索出了一些新的管理体制与机制。

一是建立全生命周期安全管理机制。部分地方致力于在数据开放与利用的过程中，建立数据安全事件的风险预判、识别、预警、监测与控制机制，制定应急处置预案与应急响应程序，定期开展应急演练，准备应急队伍与专家资源等。在数据开放准备阶段，对于一些高价值的敏感数据集，通过数据脱敏方式在降低安全风险的前提下对社会公众予以开放。例如，山东省在《公共数据开放》标准中编制了《数据脱敏指南》，对敏感数据的识别、标识、场景确定及脱敏操作等进行了规定，以指导公共数据的脱敏工作。在数据开放后，针对已发生的数据安全事件，多地建立了事后处置追责机制，对违规数据利用主体，会依据情节轻重处以记录信用档案、追究法律责任、行政处罚等不同措施；对违规的公共数据开放主体，则会处以行政处分、限期整顿等处罚措施。

二是实施公共数据的分级分类开放。例如，浙江省出台了《公共数据开放与安全管理暂行办法》，将公共数据分为禁止开放类、受限开放类、无条件开放类三种类型，并将公共数据的秘密级别划分为国家秘密、商业秘密和个人隐私三个类别，对不同类型与层级的公共数据进行分类管理。对于“无条件开放类”公共数据，利用主体可以自由获取使用，而无需进行申请审批；而对于“受限开放类”的公共数据，则要求公共数据开放主体和公共数据利用主体签署“公共数据开放利用协议”，对数据利用情况、数据利用用途、数据利用安全职责、保障措施等作出明确约定。

三是明确数据利用主体的安全义务。各地数据开放平台在利用主体获取数据时，会通过协议签署的方式告知其数据安全义务，以落实数据利用主体的责任。例如，山东省数据开放平台在下载无条件开放数据时，会以弹框提示的方式提醒利用主体阅读协议，让其知悉保护数据安全方面应履行的义务，并以“已详细阅读该协议”作为获取数据的条件；对于有条件开放类数据，数据开放部门则会在通过对利用主体的申请审核后，会与利用主体签订数据开放利用协议，明确数据安全保护义务。

（二）技术手段

公共数据的安全有序开放离不开技术工具的保障。当前，在公共数据开放领域使用较多的主要有数据加密、数据脱敏、数字水印、数据沙箱等数据安全技术。

数据加密技术是指以密码技术为基础对数据进行编码转化，从而让攻击者无法获取有价值的信息，而拥有密钥的一方可从乱码中恢复原始数据。数据加密技术可用于满足数据全生命周期的各个环节的安全需求。在本文中，为了同其他数据安全技术区分，将哈希算法、数字签名等传统密码学技术一并统称为数据加密技术。

数据脱敏技术是指对数据中包含的秘密或隐私信息（如个人身份数据、商业机密数据等）进行数据变形处理，使得恶意攻击者无法从经过处理后的数据中直接获取敏感信息，从而实现对机密及隐私信息的保护。

数字水印技术是指在数据流通时，在提供方原始数据中嵌入具有可鉴别性的数字信号，从而实现溯源追踪的能力。

数据沙箱技术是指利用数据脱敏、加密、权限管理等技术，从网络、数据、业务等多层次建立安全隔离环境，并将数据放置在隔离环境中，并在隔离的沙箱内部进行计算，用户只能从沙箱中获取经过审核确认的计算结果，但无法取走原始数据，从而保证数据的安全。

在数据开放利用的过程中，以上各类技术手段在兼顾数据的安全性和可用性方面各有利弊，其特性比较如表 1 所示。

表 1 现有数据安全技术的特性比较

技术名称	安全性						可用性			
	不可得	不可知	不可还原	不出域	不可篡改	可追溯	可算	可查	可再利用	便捷度
数据加密技术	*****	****	*****	*****	*****	**	**	****	**	***
数据脱敏技术	***	***	****	***	/	/	****	***	***	*****
数字水印技术	/	/	/	/	**	****	****	****	****	***
数据沙箱技术	*****	****	*****	***	*****	*****	****	****	**	**

注：* 越多表示该特性越强

安全性：

不可得：外部无法获得原始数据，可使用加密等技术以保证数据的机密性，也可通过访问控制手段阻止数据获取。

不可知：外部即使在得到数据的情况下，也无法知道数据信息所对应的原始主体或原始数据内容，可通过对密钥的管理或者一定安全强度的数据脱敏技术，以及部分访问控制手段实现。

不可还原：指在经过相应技术处理后使原始数据无法被推导和还原。

不出域：原始数据不出数据方所控制的范围，包括数据被加密后又解密还原的情况，但如果解密操作被限定在特定封闭范围内则不在此情况中。

不可篡改：指数据在流通和使用的过程中，不被非法篡改和破坏，以保证数据的一致性。

可追溯：指数据在流通过程中可以对其来源和去向进行追踪和溯源。

可用性：

可算：处理后的数据能够像原始数据一样进行计算处理（如四则运算、逻辑运算等），以及在计算使用上的友好度。

可查：处理后的数据能够像原始数据一样作为条件进行检索查询，以及在查询使用上的友好度。

可再利用：对于处理后的数据的二次分发利用的友好度。

便捷度：指该技术在数据开发、利用等方面的简易和便捷程度。

总体上，在公共数据开放中，现有保障数据安全的管理机制侧重于开放前的数据审查、数据脱敏，以及开放后的责任追溯和救济补偿。但是受制于监管能力的不足，对数据流通与利用过程中的监测、 追踪与干预还相对薄弱，难以有效遏制数据安全事件的发生。此外，现有的数据安全技术虽然在安全性和可用性等特性上各有所长，却难以做到均衡全面，无法同时为数据供给与利用双方提供足够的便利与支撑。

四、隐私计算作为新型数据安全技术的潜在优势与应用场景

（一）什么是隐私计算

隐私计算是在保护数据本身不会对外泄露的前提下实现对数据价值挖掘和开发利用的信息技术，是一套包含人工智能、密码学、数据科学等众多领域交叉融合的跨学科技术体系。隐私计算技术可在无需改变数据存储位置的情况下支持数据查询、数据建模等多方数据协同利用的场景，进而实现对于数据价值的挖掘。隐私计算技术主要包含多方安全计算、联邦学习、机密计算等关键技术。各种隐私计算技术的利用特性如表 2 所示。

表 2 隐私计算技术利用特性

技术名称	安全性						可用性			
	不可得	不可知	不可还原	不出域	不可篡改	可追溯	可算	可查	可再利用	便捷度
多方安全计算技术	*****	****	****	*****	**	*	*****	****	****	*
联邦学习技术	*****	****	***	*****	**	*	*****	/	****	*
机密计算技术	*****	****	****	****	****	*****	*****	*****	***	***

注：* 越多表示该特性越强

多方安全计算（Secure Multi-Party Computation，简称 MPC）是指互不信任的参与者在不泄露各自隐私数据的情况下，利用隐私数据参与保密计算，共同完成某项计算任务。

联邦学习（Federated Learning，简称 FL）本质是一种分布式机器学习技术。联邦学习过程中各参与方的数据始终保存在其本地服务器，参与方之间交换训练中间结果和模型参数，而不交换数据本身，有效降低了传统中心化机器学习带来的数据泄露风险。

机密计算（Confidential Computing）是指在受信任的硬件执行环境基础上构建安全区域，对使用中的数据进行保护。机密计算的所有参与方将需要参与运算的明文数据加密传输至该安全区域内并完成运算，安全区域外部的任何非授权的用户和代码都无法获取或者篡改安全区域内的任何数据。

报告从安全性和可用性两个维度将隐私计算技术同其他数据安全技术进行了对比，如图 2 所示。

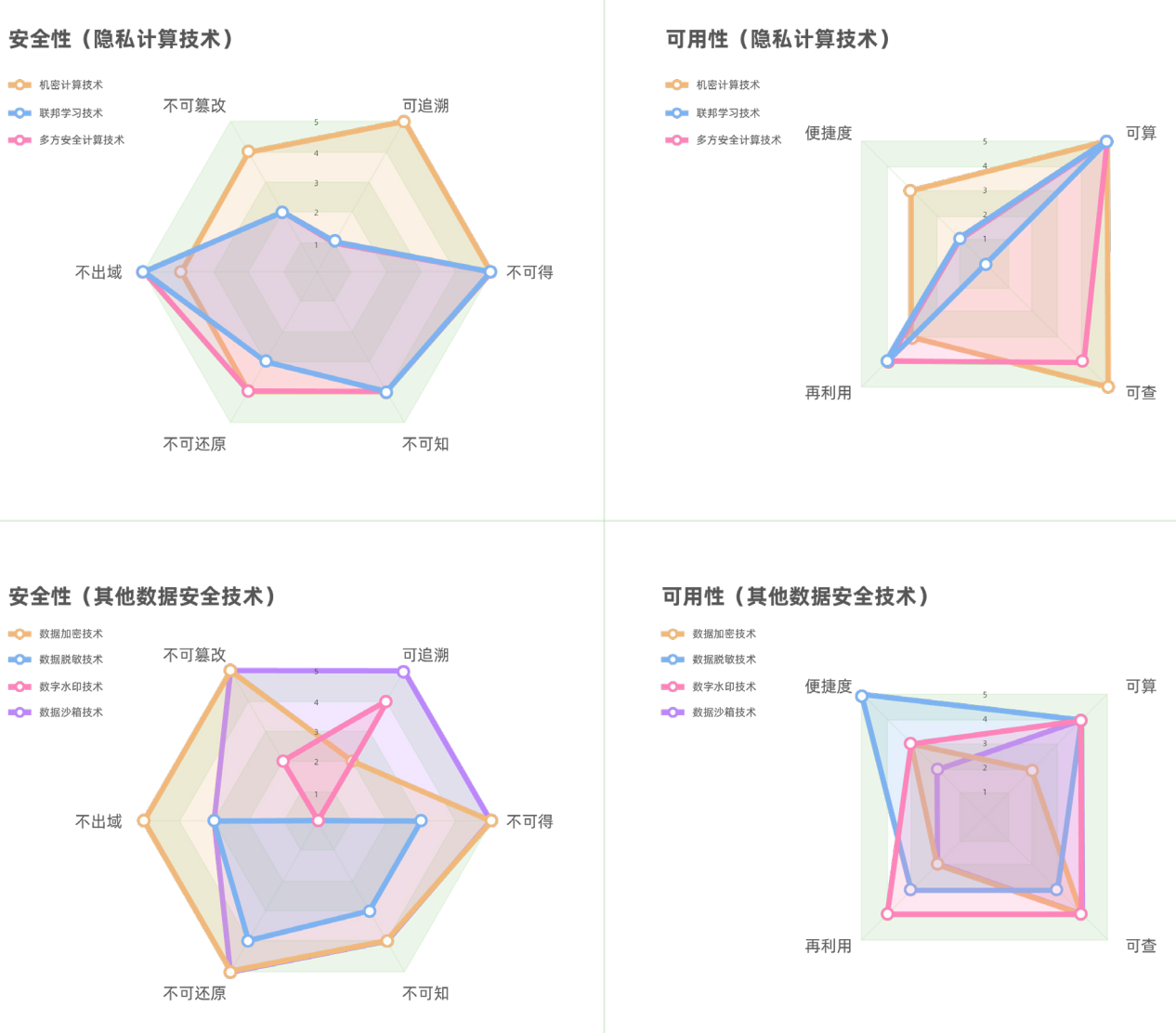

图 2 隐私计算与其他数据安全技术的比较

在安全性上，隐私计算技术同其他数据安全技术各有优势。隐私计算技术在不可得、不可知、不可还原、不可出域等方面均表现良好，不同的隐私计算技术之间在安全性方面也存在差异，如机密计算技术在不可篡改和可追溯方面皆表现良好，而多方安全技术和联邦学习技术在这两方面相对较弱；其他安全技术之间整体上参差不齐，数据加密技术和数据沙箱技术整体上表现较好，数字水印技术表现相对较差。

在可用性上，隐私计算技术总体表现优于其他数据安全技术。除联邦学习技术不涉及“可查”外，隐私计算技术整体上在可算、可查、再利用等方面均表现良好，但在便捷度方面有所欠缺；其他数据安全技术中数据水印技术总体表现良好，在便捷度方面的优势较为明显。

（二）隐私计算的潜在优势

隐私计算平台作为多方安全计算、联邦学习等技术在学术领域成果的直接工程落地实践，满足了大部分场景的商业化落地需求。从技术层面来讲，主要有以下几个方面的优势：

中立性平台。隐私计算平台不存储用户原始数据，隐私计算结果不落盘，数据处理过程可通过区块链等方式进行记录，从而降低了数据安全风险。

最小化查询。隐私安全集合求交（PSI）作为隐私计算平台的核心功能之一，其基于密码学技术实现，允许参与方使用各自的数据集合计算交集，且不会泄露除交集以外的任何数据，从而提供数据的最小化查询。

精细化管控。隐私计算平台可防止未经授权的访问，数据协作方需要事先约定数据的使用用途和使用条件，隐私计算平台可实现对数据用法用量的细粒度管控，在获得数据提供方授权的前提下，数据使用方才可以开展数据协同作业，整个过程中原始数据不出本地数据库。

（三）隐私计算现有应用场景

因其在“数据可用不可见”及提升数据处理行为的合规性等方面的优势，隐私计算目前在金融、医疗、政务等已有较为成熟广泛的应用空间。

金融领域方面，在联合风控、精准营销等场景中借助隐私计算技术“原始数据不出域”、“可算不可得不可导”、“可查不可知”等技术特性，既保障了数据的安全，又实现了数据价值的挖掘。

医疗领域方面，在疫情防控、基因分析、临床医学研究等场景中借助隐私计算技术“原始数据不出域”、“可算不可得”等技术特性，不仅可以实现保护数据隐私保护下的医学数据安全统计分析和医学模拟仿真与预判，还可以实现联防联控、群策群力，方便临床科研成果的产出。

政务领域方面，在办案侦查、房贷违约风险预测等场景中借助隐私计算“原始数据不出域”、“可算不可得不可导”、“可查不可知”等技术特性，不仅规范了数据的查询使用与管理，还通过联合建模提升了预测的精准程度，有利于推动数字政府的建设和数字经济的发展。

随着隐私计算技术的逐渐成熟和商业化落地，中央和地方政府从 2020 年开始陆续发布了相关政策文件（见附录 2）。例如，《“十四五”大数据产业发展规划》提出要加强隐私计算、数据脱敏、密码等数据安全技术与产品的研发应用，提升数据安全产品供给能力，做大做强数据安全产业等。鼓励和支持运用隐私计算等技术构建数据安全防护体系，为数据要素安全流通提供技术支撑。

五、隐私计算在公共数据开放领域的应用展望

如前所述，数据安全风险既限制了可开放数据的范围，也降低了政府部门的开放数据的意愿，造成 “不愿开放、” “不能开放” 、“不敢开放” 等问题，阻碍了高质量的开放数据供给。现有的管理机制与技术手段虽然能够在一定程度上消弥部分安全风险，但仍不能全面均衡地解决数据可用性与安全性之间的困境。而隐私计算技术所具有的“原始数据不出库、数据可用不可见”、“数据使用可控可计量”以及“计算分布式、监管有中心”等特征和优势，有助于在保证数据归属清晰的同时解决数据安全问题，契合了公共数据有序开放的需求。

（一）隐私计算应用于公共数据开放的潜在收益

具体而言，隐私计算作为一种技术手段，通过隐私计算平台的工程化落地实践，满足了一些场景的应用落地需求。隐私计算平台可作为一种底层数据安全能力支撑和上层技术服务工具，其能为公共数据开放工作带来的潜在效益主要体现在以下三个方面：

（1）推动高价值、低风险的公共数据供给

如前文所述，受制于安全风险，当前我国开放的公共数据在数量和质量上都还无法满足社会公众的普遍预期。在公共数据中，有大量高价值、高风险的数据集，由于相关部门缺少数据安全管控能力而不敢开放或不能开放。这些数据对促进数字经济发展与数字社会建设具有重要作用，亟需在数据安全技术的赋能下对社会予以开放。

隐私计算的技术特性使得以较低风险开放较高价值的公共数据成为可能。依托隐私计算技术“原始数据不出域、数据可用不可见”的新型数据流通模式，公共数据资源可以实现“物理集中 + 逻辑集中”双汇聚模式，物理集中是指将公共数据资源物理集中于公共数据开放平台，进行统一集中化存储和管理。逻辑集中是指将公共数据资源目录汇集于公共数据开放平台，而公共数据资源则分散存储于数据提供方本地。数据使用方可以通过公共数据开放平台了解公共数据的全貌，但在使用过程中可通过隐私计算方式实现双方或者多方的数据安全协同计算，整个过程中原始数据不出本地数据库。通过“逻辑集中，物理分散”的新模式，从而降低了开放数据的安全风险。

同时，还需要强调的是，隐私计算技术也不可过度应用，对于列入“无条件开放”属性的低风险数据，仍应尽可能开放原始性的、完整的、可机读的数据集，以最大程度降低数据利用的门槛，释放数据的价值。

（2）兼顾安全性、灵活性的公共数据利用

《数据安全法》、《个人信息保护法》的相继出台，与《网络安全法》一同构成了数据安全合规领域的“三驾马车”。在法规政策的强监管下，公共数据利用也面临着较大的安全合规风险。

隐私计算技术集合了密码学、机器学习等技术，可以通过不同技术方式实现数据的最小化查询。隐私集合求交（Private Set Intersection，PSI）作为隐私计算的关键技术之一，可以允许数据协作方之间使用各自的数据集合计算交集，但不会泄露除交集以外的任何数据。

同时，隐私计算技术支持对开放数据使用进行管控授权。具体表现为：隐私计算平台通过区分用户权限来提供不同颗粒度的数据，并对数据的用法、使用时间、使用次数、并发限制等内容进行设定，从而实现对公共数据的精细化治理。由此，公共数据开放能够在价值与安全之间保持平衡，推动兼顾安全性与灵活性的公共数据利用。

（3）搭建工具化、低成本的安全开发环境

目前各地方政府公共数据开放平台普遍开设了开发者中心板块，为开发者开发数据应用提供便利，但赋能效果不太明显。隐私计算平台作为一种技术能力载体，不仅可作为保障数据安全流通的技术底座，也可成为公共数据开放平台的服务工具，帮助开发者进行数据应用的快速开发。

具体而言，隐私计算平台可将底层复杂的密码学、机器学习等技术和建模过程抽象成算子，支持用户以拖拽式交互，构建可视化建模 pipeline，开发者无需了解底层的技术实现原理，只需关注业务实现，合理的运用不同的算子完成数据应用开发。这种工具化的能力不仅降低了开发者的学习成本，还提高了数据应用的开发效率，有望实现大规模高价值的数据应用开发与流通。

（二）隐私计算对不同主体的价值

隐私计算以隐私信息处理的全生命周期为优化对象，其基础法律关系涉及三类主体，即：数据提供方、数据开放平台管理方、数据使用方。传统公共数据开放平台升级隐私计算能力，将有助于分别满足这三类主体在不同场景的需求。

对于**数据提供方**，通过隐私计算方式供给具有较高价值和较高风险的公共数据时：

- 可只提供公共数据目录，原始数据不出本地数据库，外部“不可见”、“不可识”、“不可知”，保证数据的机密性，保障数据提供方的相关权益。
- 可只经由一个平台对多个数据使用方、多个数据协作场景进行分门别类的管理，有效节省数据维护和管理成本。

对于**数据开放平台管理方**，通过隐私计算方式管理具有较高价值和较高风险的数据时：

- 可将传统数据物理汇集的模式转变成“逻辑集中、物理分散”的模式，保留数据的完整性，提升公共数据的数量和质量。
- 公共数据开放平台由数据资源汇集转变为公共数据目录汇集，一方面可减少平台管理方数据治理的精力和成本耗费，另一方面也可保障公共数据流通和多方主体数据开发协同中的安全合规程度。

对于**数据使用方**，通过隐私计算平台获取具有较高价值和较高风险的公共数据时：

- 可申请可用性较高的公共数据，在安全合规的前提下开发高价值的数据应用。
- 可借助隐私计算平台对数据、模型等再利用的特性进行数据应用开发，降低数据开发利用成本， 提高数据应用开发的敏捷性。

（三）隐私计算面临的局限与挑战

隐私计算并非万能，并不适用于所有场景。由于隐私计算技术种类的多样性，各种技术实施路线之间存在着较大的差异，在保护效果、计算性能、计算精度，以及对于硬件的依赖等方面也有所不同。由于目前还没有完全统一的技术标准和测评标准，隐私计算大规模应用仍存在一定的局限性：

- 多方安全计算包含很多复杂的密码学运算，对于计算的性能会有相对较大的损耗，其性能也会受到通信方面的限制。
- 联邦学习在其安全保障效果方面目前还没有被严格地定义或证明，所以对于有较高安全要求的数据使用场景，需和其他技术配合使用，单一使用该技术会将具有一定的局限性。
- 基于机密计算的隐私计算方式对于硬件有较强依赖，因此会相应地增加使用成本。

此外，由于技术路线的多样性，在异构的技术体系之间，甚至是采用相同技术方案的不同厂商的产品之间，可能在互联互通上存在一定的障碍和壁垒，出现数据“孤岛”变成“群岛”的情况，对于公共数据的流通仍然会形成了一定的阻碍。

六、总结与建议

公共数据开放是推动数字化发展的重要内容。然而，公共数据开放风险与价值并存，公共数据在存储、流通、利用各环节都存在安全隐患。在数据安全日益受到重视的今天，如何在数据开放与安全保护之间达成平衡成为重要命题。虽然各地在体制机制与技术手段两方面，已对公共数据开放中的安全保护做了诸多探索，但仍存在一些不足和短板，制约了高质量公共数据的充分供给，阻碍了社会对开放数据的有效利用。

新兴的各类隐私计算技术能在一定程度上弥补当前数据安全保护能力的不足，有助于在保障数据开放利用的同时有效解决数据安全问题，有望在公共数据开放若干场景中发挥重要价值。

在开放数据的实践中引入隐私计算技术，有利于构建多主体协同、全周期覆盖的数据安全保护能力，具体建议包括：

- 对于高风险数据，可从开放原始数据转向“数据可用不可见”。通过引入隐私计算技术，让数据利用方仅能获得数据利用结果，无法获得与推理原始数据，从而在释放数据价值的同时，降低数据泄露风险。
- 构建全周期的数据管控能力。利用隐私计算“数据流向可追溯”、“数据用法可控可计量”的技术特性，在数据开放利用的全周期监测各主体行为，既实现对安全风险的早预警、早发现、早处置，也为责任追查与事后救济提供依据。
- 从集中式开放走向分布式开放。利用隐私计算“原始数据不出库”的技术特性，数据开放平台可仅提供数据目录与元数据，不再汇集各部门各机构的原始数据，以降低数据泄露的风险。

七、附录

附录 1 有关数据安全的政策法规（部分）

类别	名称	发布单位	发布时间	要求
法律	中华人民共和国网络安全法	全国人大常委会	2016年11月07日	“鼓励开发网络数据安全保护和利用技术，促进公共数据资源开放”
	中华人民共和国数据安全法	全国人大常委会	2021年06月10日	规范数据处理活动，保障数据安全，促进数据开发利用
	中华人民共和国个人信息保护法	全国人大常委会	2021年08月20日	保护个人信息权益，规范个人信息处理活动，促进个人信息合理利用
政策文件	促进大数据发展行动纲要	国务院	2015年09月05日	“健全大数据安全保障体系”，“明确数据采集、传输、存储、使用、开放等各环节保障网络安全的范围边界、责任主体和具体要求”
	“十四五”大数据产业发展规划	工业和信息化部	2021年11月15日	“筑牢数据安全保障防线”，“完善数据安全保障体系”，“加强隐私计算、数据脱敏、密码等数据安全技术与产品的研发应用”
	要素市场化配置综合改革试点总体方案	国务院办公厅	2021年12月21日	“加强数据安全保护”，“探索‘原始数据不出域、数据可用不可见’的交易范式，在保护个人隐私和确保数据安全的前提下，分级分类、分步有序推动部分领域数据流通应用”
	“十四五”推进国家政务信息化规划	国家发展和改革委员会	2021年12月24日	“强化政务数据安全管理”，“统筹推进国家数据共享交换平台与公共数据开放平台的数据安全保障体系建设”
	“十四五”国家信息化规划	中央网络安全和信息化委员会办公室	2021年12月27日	“强化数据安全保障”，加强数据“全生命周期的安全管理，建立健全相关技术保障措施”，“强化平台企业数据安全保护责任”

附录 2 涉及隐私计算的政策法规

政策层级	政策名称	文号	发布单位	发布年份	重点内容
中央层面	国务院办公厅关于印发《公共数据资源开发利用试点方案》的通知	国办函〔2020〕29 号	国务院办公厅	2020年05月	提出以上海市、江苏省、浙江省、福建省、山东省、广东省、海南省、贵州省为试点探索公共数据资源开发利用。要求牵头部门要会同有关政务部门、公共事业单位加强数据脱敏和算法模型审核，探索“数据可用不可见”等不同类型的交互模式，防范差分隐私风险。
	全国一体化大数据中心协同创新体系算力枢纽实施方案	发改高技〔2021〕709号	国家发展改革委、中央网信办、工业和信息化部、国家能源局	2021年05月	提出试验多方安全计算、区块链、隐私计算、数据沙箱等技术模式，构建数据可信流通环境，提高数据流通效率。
	新型数据中心发展三年行动计划（2021-2023年）		工业和信息化部	2021年07月	提出要加强多方安全计算等数据安全关键技术创新突破与推广应用。积极组织做好各类网络数据安全协同处置，及时消减数据安全重大隐患。
	“十四五”大数据产业发展规划		工业和信息化部	2021年11月	提出加强隐私计算、数据脱敏、密码等数据安全技术与产品的研发应用，提升数据安全产品供给能力，做大做强数据安全产业。
	要素市场化配置综合改革试点总体方案	国办发〔2021〕51号	国务院办公厅	2021年12月	提出探索“原始数据不出域、数据可用不可见”的交易范式，在保护个人隐私和确保数据安全的前提下，分级分类、分步有序推动部分领域数据流通应用。探索建立数据用途和用量控制制度，实现数据使用“可控可计量”。
	“十四五”推进国家政务信息化规划	发改高技〔2021〕1898号	国家发展改革委	2021年12月	提出优化完善政务数据资源目录，创新应用区块链、隐私计算等新技术，推进政务数据的算法式安全共享，推进国家数据共享交换平台与国 家公共数据开放平台的协同联动，深化公共资源交易平台数据资源整合共享。
地方层面	广东省数据要素市场化配置改革行动方案	粤府函〔2021〕151号	广东省人民政府	2021年07月	提出要加强数据资源汇聚融合与创新应用，构建数据安全存储、数据授权、数据存证、可信传输、数据验证、数据溯源、隐私计算、联合建模、算法核查、融合分析等数据新型基础设施，支撑数据资源汇聚融合和创新应用。
	山东省“十四五”数字强省建设规划	鲁政字〔2021〕128号	山东省人民政府	2021年07月	提出打造数据应用总门户，搭建集数据建模、隐私计算、数据分析与可视化于一体的若干服务中台，构建跨层级、跨部门、跨业务的“多租户”服务体系。提出促进数据要素市场流通。
	珠海市关于加强隐私计算在城市数字化转型中应用的指导意见		珠海市“数字政府”改革建设工作领导小组	2021年08月	提出要“以隐私计算技术为重要抓手”“建设成为国家级数据要素应用先行示范区，力争到 2023 年我市隐私计算技术应用水平走在全国前列。”。
	海南省公共数据产品开发利用暂行管理办法	琼数组办〔2021〕3号	海南省政府大数据推进工作领导小组办公室	2021年09月	管理办法提出支持利用区块链、隐私计算等新技术充分利用外部平台资源和数据资源，促进数据协同、共享开放、数据融合、产品创新。
	上海市数据条例		上海市第十五届人民代表大会常务委员会第三十七次会议表决通过	2021年11月	提出上海要与长三角区域其他省共同推动区块链、隐私计算等数据安全流通技术的利用，建立跨区域的数据融合开发利用机制，发挥数据在跨区域协同发展中的创新驱动作用。
	广西面向东盟的“数字丝绸之路”发展规划（2021—2025年）	桂政办发〔2021〕113号	广西壮族自治区人民政府办公厅	2021年11月	提出要积极运用隐私计算、多方计算、联邦学习等技术，构建“数据可用不可见、数据可控可计量”的数据流通体系，探索推动中国—东盟大数据交易市场培育。

报告编写人员

数字中国研究院（福建）	
宋志刚	副　院　长
黄　荣	高级工程师
林万枝	高级工程师
杨文琴	工　程　师
石维钗	工　程　师
北京数牍科技有限公司	
宋一民	创　始　人&CEO
蔡超超	联合创始人&CTO
何东杰	高级副总裁
马福忠	高　级　总　监
裴　超	数据安全专家
复旦大学	
郑　磊	教　　授
韩　笑	研究助理
侯铖铖	研究助理

发布单位

数字中国研究院（福建）

北京数牍科技有限公司

复旦大学数字与移动治理实验室

发布单位：

- 数字中国研究院（福建）
- 北京数牍科技有限公司
- 复旦大学数字与移动治理实验室