数据安全治理的几个基本问题

杜跃进

2018-12-04

Several Basic Questions about Data Security Governance

（根据 2018 年 10 月 18 日在中国国际大数据大会上的演讲整理）

数据安全如今在全球都受到极大的关注，大家都在讨论这个话题。但什么是数据安全、该如果做数据安全等基本问题，还存在很多混乱甚至误解。本文围绕当前数据安全相关的一些基本概念进行探讨。

1 为何强调“数据安全”，而非“大数据安全”

很多人在讨论大数据安全这个话题的时候，会纠结于“这是不是大数据”或者“这是不是大数据系统”。这会忽略今天我们面临的真正问题：数据无处不在、无处不用的情况下如何保证数据安全。如果用大数据的定义来限定所谓大数据安全的范围，而忽略非大数据系统更加容易成为坏人得手的重灾区这一现实，那么就不会让数据安全的现状有任何改善。例如让全社会开始高度重视数据安全的“徐玉玉”案件，显然和大数据以及大数据系统都无关系。如果大家搞了半天所谓“大数据安全”，却不能减少下一个“徐玉玉”事件的发生概率，那么这样的“大数据安全”并没有什么意义。所以当前全社会需要关注的其实是“大数据时代下的数据安全”。因此，当我们今天讨论数据安全问题的时候，不能用是否是“大数据”或“大数据系统”来限定范围。而且，由于在大数据时代下数据的存在形式、使用方式、流转共享模式等都和原来极为不同，因此大数据时代下的数据安全是个新问题，而不是过去传统的数据安全概念。

2 “数据安全”不仅仅是防窃取

现在很多人说的数据安全问题实际仅仅是实际要解决的问题的一小部分。从用户的角度来看，当前数据安全至少包括以下三方面问题：

一是数据被窃取,即担心自己的数据被坏人偷走。实际上坏人可能从外面偷数据，也可能从内部偷数据。按照以往安全行业的基本共识，来自外界的安全威胁只占三分之一左右，三分之二的安全威胁是从组织内部发起的。根据我们的调查，在一些特定行业中内鬼窃取数据的比例还要更高。然而迄今依然有很多人把防范外部数据窃取作为数据防窃取的全部内容，而忽视对来自内部的攻击应对。仅仅强调用户侧应用软件的安全、数据通信过程加密、防火墙设个大门等，以为门外管好就行了，这是远远不够的。

二是数据被滥用，即用户对拥有数据的服务方不放心，担心他们会滥用自己的数据。用户的数据在服务提供方那里,他们的员工会不会滥用权限、随便访问用户数据？他们会不会因为好奇或者朋友要求去看某个用户的个人信息？他们会不会把用户的个人信息倒卖出去？和通过内部网络攻击窃取数据的行为不同,这里说的是服务方工作人员在授权范围内从事了不符合业务场景的数据访问。例如，用户要求客服帮忙，客服人员在这种场景下访问该用户的数据解决其问题，这是正常的业务场景。但是如果没有用户请求，客服人员擅自访问用户数据，就属于滥用行为。在大数据时代，数据和业务都在高频产生和变化,若对每一次数据访问行为都重新进行权限申请与审核，将直接扼杀业务，用户也无法接受这样的效率。因此数据安全的工作需要包括对数据滥用行为的识别、报警甚至阻止，并且有制度保障对实施滥用行为的员工进行严厉制裁。从目前曝光的数据黑灰产案件中看到，大量案件都是通过买通内部人员滥用数据

权限进行数据倒卖的，而这些情况都是案发之后才被调查发现，说明这方面能力的欠缺。

三是数据被误用，即在大数据加工使用的过程中会不会侵犯用户利益，也就是很多人谈之色变的“用户画像”“精准营销”等。在各种描绘大数据给人类带来美好机会的故事中，实际上都离不开精准的个性化服务。而这些故事已经在金融、健康、医疗、教育等很多领域中开始切实发生，甚至在制造领域也在快马加鞭地发展，未来所有的领域都会走到这一天。这种精准的个性化服务的背后其实就是“用户画像”“精准营销”这些大数据加工技术。技术本身是中立的，关键在于技术被如何应用。“精准营销”是被用来“杀熟”还是用来更好地满足用户的个性化需求？“用户画像”是用在支持个性化服务或者保护用户安全，还是用在满足其他不良动机？在大数据分析加工的过程中，有没有人能够从中窥探到某个特定人的个人信息或者隐私？这些是防误用的内容。当前的技术和管理总体上能够控制大数据加工过程中的误用，使人们在享受大数据带来的好处的同时，把危险关在笼子里。但是现实中，还有很多企业和组织在发展的过程中忽略了这个问题，让用户感到大数据是个恐怖的恶魔。数据防误用的问题，现在被关注得更少。

3 数据安全为何必须“以组织为单位”

目前有一个比较常见的误区是把手机上的移动应用软件（APP）安全等同于数据安全，例如通过评估手机上的移动应用软件安全来判断个人数据是否得到保护。手机APP 一直都是安全的重灾区之一，很多 APP 确实存在不经用户许可秘密采集用户数据、植入广告、秘密产生费用或者自身安全做得不好导致其数据被窃取等情况。但是APP 就算不存在这些问题，也不等于用户的数据就是安全的。

在如今的时代，用户的数据并不都是沉淀在移动应用软件内部、存在用户终端上的。移动应用软件背后连接的是云端，而且可能连接不同业务。因此即便是在“知情同意”“最小够用”等原则下取得了用户的授权，用户数据还是会存在后端，并且在今天的社会化大协作的模式下进行共享使用。因此，大数据时代下数据的边界是在产品、设备、业务、人员还是系统？显然都不是。数据至少会在提供服务的组织内部的不同产品、业务、设备、系统、人员中流动，甚至为了完成用户的一个服务需求，数据还必须在产业链上的不同组织之间流动。例如，为了完成用户的一个购物需求，数据就必须在商家、平台、物流、独立软件供应商、金融等多个环节中流动。

因此，更合理的方式是以一个组织为单位来衡量数据安全的情况。这里“组织”指的是拥有数据、提供服务的企业或者机构，其具有相对独立和完整的管理，也能够对业务和安全负责。数据在一个组织内的不同产品业务中形成流转闭环，组织是数据流动的最小边界。组织与组织之间通过可控的制度程序或者接口实现数据的跨组织流动、共享、交易等，这时候也可以以单个组织的数据安全能力为基础，进行责任的划分或者数据流动风险的控制。

如今，政府在大力推动多部门数据打通和共享利用，很多企业也都在进行组织变革，建立更强大的中台能力，其原因是看到数据只有打通和流动起来，才能更好地发挥价值。在这个过程中，单个产品或者技术平台的安全都不代表数据本身就是安全的。只有以组织为单位，才可以跳出频繁变化的产品、业务、人员等带来的困惑，寻找到支撑今天数据安全需求的方法。

4 为何传统安全方法不适用于大数据时代下的数据安全

大数据时代下的数据安全是一个全新的问题，无法简单地用原来的安全方法来解决。这主要体现在以下两个层面。

一是不能用“以系统为中心的安全”思路解决问题。以系统为中心的安全是大家熟悉的安全方法，例如看某个软件、某个服务器或某个手机终端安全与否。这主要是看这些系统在各种人为干预下是否会出现与预期设计不符合的功能，从而导致运行状态失控。如今，数据要在不同的系统之间流动，若某个系统出了问题，可能影响到当时在这个系统中的数据（包括被窃取），但这些数据也可能在别的系统中出问题。数据本身并不存在运行状态，数据出问题的概念和系统出问题的概念也不同。这两者的关系，有点像医院里“心血管科”和“血液科”一样，前者解决的是血液循环系统本身的安全（运转正常），后者则是要保障血液自己的安全。两者显然有关系，但又有很大不同。单个系统的安全并不等价于数据的安全，系统被入侵也不等于数据一定会被偷走，每个系统都固若金汤也不等于数据就不会被滥用或误用。解决数据自身的安全问题，需要切换到“以数据为中心”的安全思路上来。

二是不能用传统的“数据安全”方法解决问题。数据安全是最古老的安全概念。从古代战场上就产生了数据安全的需求，并推动了相关技术的不断发展。对一个文件、一个数据库的记录的保护等都是数据安全的概念。但是，如今的数据安全的概念和方法已经和过去完全不一样了，数据的存在形式、使用方式和共享模式与过去有了极大的变化，数据的权属也不都是数据处理者的。数据可能以文件、记录、字段等方式在不同的环节中被快速打散、重组、流动，在这个过程中还会源源不断地产生新的数据。在一个业务里，数据可能涉及很多设备、服务器、产品、用户和不同部门的人的信息，然而真正需要回答的是数据在这么复杂的全过程中，从用户的角度来说安全不安全？显然，这和传统的“数据安全”概念有很大区别。

5 数据安全为什么需要强调治理而非管理

这里说的管理，指的是根据事务的规律制定一整套规则，然后自上而下进行规则的执行落地，控制系统达到预期状态。而治理指的是找到若干关键抓手和基本逻辑，调动多方力量和资源形成某种协同或者生态，通过社会协同引导整个系统达到预期状态。大数据时代下的数据安全治理无法使用传统的管理模式达到目标，必须走协同治理的道路。

全社会所有的行业都在进行数字化转型，一切都将步入数据驱动，数据将成为所有领域的基本生产资料。因此数据安全问题也将涉及所有行业，并且涉及产品、业务、人员、共享机制等，并不是某个垂直领域的知识或者某个层面的单一方法就可以解决的。如果按照过去管理某个垂直特定行业的方式，设立若干部门自上而下进行管理，不但成本无法承担，效率也无法适应今天的实际情况。因此，政府、行业、企业、安全、第三方机构等需要发挥各自的优势形成有效的配合，才能建立适应当今数字时代的协同治理模式，共同提升全社会的数据安全水平。

6 数据安全治理的关键目标是让安全成为竞争力

我们不希望数据安全工作堵住所有行业的血管，阻碍大数据时代的创新和发展。如果简单化地看如何“消除数据安全问题”的话，那么不要有数据、不要有数据流动和应用是最安全的。数据安全治理的最终目标是实现数据安全和数字经济发展之间的平衡，甚至是二者的相互促进。我们通过法律法规、政策标准、技术产品、产业发展、测评培训、监督机制等进行综合数据安全治理模式设计的时候，如果能够瞄准以下两个关键目标，就可以推动数据安全治理最终目标的实现。

一是让数据安全成为组织的竞争力而不是成本，实现“能者多得”，即数据安全做得好意味着有资格得到更多的业务机会。安全在过去一直是成本，所以几乎毫无例外，每一个创新业务、每一个创业公司、每一个创新产品一开始都不愿意做安全，因为大家首先考虑的都是活下来的问题，没有精力和资源管活得好不好的问题，而且若对安全进行投入，可能在竞争中由于开发速度更慢、开发成本更高而失败。甚至很多大型项目系统的建设也不愿意做安全。于是等产品逐渐成熟、业务逐渐做大、工程项目投入运行之后，慢慢发现很多安全的坑已经难以填补了。

谁不做安全或者少做安全，谁更可能赢得竞争，从安全的角度来说，这就是“劣币驱逐良币”。在数据安全领域通过建立科学的治理模式可以改变这个现象，要点是让一个组织能处理数据的类型和规模，与其数据安全能力水平挂钩。例如健康医疗行业迫切需要利用大数据技术大幅提升技术和业务水平，造福百姓，但是这类数据又非常敏感，那么行业主管部门可以规定：处理哪些类型或多大数量的数据的组织必须证明其达到相应的数据安全能力级别要求。这样，当一个组织想要使用健康医疗数据开展研究或业务之前，就需要先具备足够的数据安全能力。于是数据安全能力越高的企业，意味着有权处理更多类型和数量的数据。这样他们才会积极而且认真地去提升自己的数据安全能力，实现业务竞争力上与安全的正向挂钩。这样也才会带动整个数据安全产业发展和水平逐渐提高