

依照数据用途界定公共数据

On public data



朱扬勇(1963-),男,博士,复旦大学教授,《大数据》期刊编委会副主任。国际数据科学倡导者,提出数据界、数据学、数据身、数据自治、数据财政等概念和体系。发表学术论文200多篇,出版《数据学》《旃旒数据》《特异群组挖掘》《数据自治》等专著,并任《大数据技术与应用丛书》(22册)主编、《大数据资源》主编。主要研究方向为数据科学和数字经济,近期研究重点为数字化转型、数据财政、数据资产、数据自治与数据跨境等。

中图分类号: TP399

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2024037

当前,许多地方政府在地方立法中将公共数据界定为:国家机关、事业单位,经依法授权具有管理公共事务职能的组织和法人在履行公共管理和公共服务过程中收集和产生的数据。这类过于宽泛的公共数据界定导致公共数据的开放和开发利用与国家数据安全、个人信息保护形成冲撞。例如,公用事业部门的缴费数据涉及每家每户的个人信息,如何保护个人隐私?即使进行了匿名化处理,但是如果集中全国的缴费数据用于不当目的,将严重危及国家安全;又如,医疗机构中的电子病历数据是个人重要的隐私数据,即使进行匿名化处理,但是如果集中全国电子病历数据用于不当目的,将民众的人口质量公开化,同样将严重危及国家安全。因此,不应简单地将公共部门履职产生的数据界定为公共数据,更不应将公共部门持有的数据界定为公共数据。从公共品的视角来看,公共部门生产或者持有的数据不一定是公共数据。公共数据不一定是公共的数据,而一定是公众必须要用又无法自给的数据,所谓公共品数据。因此,依照数据生产者的公共属性来界定公共数据是不合适的,建议依照数据的用途来界定公共数据。

1 数据的公共性问题

数据自身带有的非竞争性和非排他性,使数据具备了公共品的必要条件,但显然不是充分条件,否则所有的数据都是公共品了。那么,公共数据的充分条件是什么呢?与公共性相关的有公共资源^①、公共品^②和公共部门^③,它们和数据的公共性的相关程度如何?

问题1: 数据可以是公共资源吗?

按照公共资源的定义,公共资源是指自然生成或自然存在的资源。由于数据非

天然和自然生成,因此,数据(或数据资源)不是公共资源。

问题2: 数据可以是公共品吗?

数据具备了公共品的必要条件,在满足另外一些条件的情况下,一些数据可以是公共品。什么类型的数据是公共品呢?如果按照公共品的定义来看待公共品数据,那么公共品数据是指个体无力独自提供(无论是否自愿),而必须由公共部门提供的社会正常存在与发展所必需的、具有非竞争性和非排斥性的数据产品和服务。有这样的数据吗?显然是有的,例如天气预报数据,天气预报数据是个人无力提供而必须由公共部门(气象局)提供的数据,可以是公共数据;又如列车班次、航班班次、导航地图等数据,也具有公共品的特征,可以是公共数据。但大部分数据不是公共品,也不应该是公共品。

问题3: 公共部门产生的数据是公共品吗?

公共部门产生的数据不一定是公共品。例如,公共部门的人事数据、管理数据、财务数据等显然且肯定不是公共品。公共部门在履职过程中产生的数据就是公共数据吗?也不一定,事实上,水电气提供部门产生的缴费数据都不是公共品。因为缴费数据不是社会正常存在与发展所必需的,即不是公众必须要用的数据,因此缴费数据不是公共品。

问题4: 私人部门生产的数据就是私人的吗?

私人部门生产的数据也可能是公共品。例如:卫星定位数据就应该是公共品,各类导航数据也应该是公共品。随着社会的发展和数字化转型,“公众所必需的”概念也会发展,例如,酒店报价数据、点评数据等将会变成公众所必需的数据,就可能成为公共品。也可能会有越来越多的私人部门提供公共数据服务。

① 公共资源: 公共资源是自然生成或自然存在的资源,它能为人类提供生存、发展、享受的自然物质与自然条件,这些资源的所有权由全体社会成员共同享有,是人类社会经济发展共同所有的基础条件。

② 公共品: 公共品是个体无力独自提供(无论是否自愿),而必须由公共部门提供的社会正常存在与发展所必需的、具有非竞争性和非排斥性的产品和服务。

③ 公共部门: 公共部门是被国家授予公共权力,并以社会的公共利益为组织目标,管理各项社会公共事务,向全体社会成员提供法定服务的政府组织,包括政府、公共企业、非盈利组织、国际组织等。

2 界定公共数据

通过上述讨论得知：从公共资源的视角来看，数据不是自然生成或自然存在的资源，因此数据（数据资源）不能自然成为公共数据资源；从公共品的视角看，不能以数据的来源界定数据是不是公共数据，而应该以数据的用途来界定数据。那么，数据的公共用途有哪些呢？数据的公共用途主要可以分为以下3类。

第一类，公众要知道的数据：是指公众知情权数据，在公众知情权范畴内，必须对公众开放的、由公共部门生产并控制的数据。这类数据就是政府数据开放目录中的数据。

第二类，公众要用的数据：是指公共用途数据，公众必须使用又难以（公众个体无能力）生产的数据，例如天气预报数据、列车班次、航班班次等数据。这类数据需要由政府专门的部门生产、管理，并向公众提供。这是公共数据最核心的部分，即公众必须要用的数据。

第三类，公共部门要用的数据：是指公共部门履职必须要用的数据。如果没有这些数据，公共部门将无法履职某种职能。因此，这些数据也间接地是公众所必需的数据。当前，公共部门持有的大部分数据就属于这类数据。

这样，把第一类、第二类数据作为公共数据是合适的，而第三类数据是否可以作为公共数据则需要进一步讨论。

首先，要区分两个概念（公共部门履职要用到的数据和公共部门履职产生的数据）之间的差别。例如，煤气公司为了提供供气服务，需要用户的家庭地址、户主及其联系方式等，这些数据是履职所必需的；而用户缴费数据在完成缴费后，就不再是

履职所必需的了。

其次，对于公共部门履职必须要用的数据，如果其中一些数据是本部门产生的，相对比较简单，即自产自销；其他一些非本部门生产的数据，如果不作为公共数据来处理，那么这个公共部门就需要外购这些数据，会增加公共部门的成本。进一步地，如果这些数据是私人部门生产的，那么私人部门可以不提供这些数据，这将导致相关的公共职能不能履行。

第三，更多的情况是多个公共部门履职需要相同的数据，无论是从成本效率考虑，还是从公共需要考虑，将其纳入公共数据是比较合适的。例如，水电煤气等公共部门都需要用户的家庭地址、户主及其联系方式等数据。而政府数据共享（一网通办、最多跑一次）就是为了解决公共部门自身的数据供给和运行成本问题。

另外，第三类数据不是政府部门之间的数据共享就能够满足的，这类数据还会涉及私人部门。如果将这类数据纳入公共数据范畴，可以为公共部门履职能力升级（数字化转型）提供保障，当然也存在公共部门过度获取数据的可能性。这个问题有待进一步研究和实践。

综上，所谓公共数据就是指公共品数据，不是公共资源数据或者公共数据资源，可以从狭义和广义两个层次界定公共数据。

- 狭义公共数据：是社会公众无力独自提供（无论是否自愿），必须由公共部门提供的社会正常生存与发展所必需的（包括公众知情权所必需的，但不包括公共部门业务所必需的）、具有非竞争性和非排斥性的数据产品和服务。

- 广义公共数据：是社会公众无力独自提供（无论是否自愿），而必须由公共部门提供的社会正常存在与发展所必需的（包括公众知情权所必需的和公共部门业

务所必需的)、具有非竞争性和非排斥性的数据产品和服务。

现在再来看各地公共数据的界定:公共数据是国家机关、事业单位,经依法授权具有管理公共事务职能的组织和法人在履行公共管理和公共服务过程中收集和产生的数据。显然这个界定没有考虑到数据的用途,忽略了“公众必需的数据”这个公共数据的本质。这个提法可能是依据“公共部门资产是公共的”这样一个逻辑,确实公共部门形成的数据资产应该是公共的数据资产,公共数据资产经营形成的收益是公共的,应该上缴财政回馈公众。

3 公共数据的提供

满足公共数据定义的数据,只是学术上或理论上的公共数据。一类数据是不是公共数据,首先要看其是否符合公共数据的定义,即是否具备了公共数据的充分必要条件;然后,要由政府来决定是否将其列入公共数据,一旦政府将其列入公共数据,这类数据就需要由公共部门向社会提供。因此,政府将一类满足公共数据条件的数据列入公共数据的前提是,有相应的公共部门持续不断地向社会提供这类公共数据,这取决于政府当时的能力和财力是否具备。

公共数据的确认和提供是一个复杂的问题,需要探索和实践。再来看天气预报数据,每天的天气预报数据是公共品,但是历史天气预报数据则不是公共数据,因为大部分人不再使用十年、五年、一年前的天气预报数据,甚至前一天的天气预报数据都没有多少人使用。公共部门(气象局)没有义务向公众提供历史天气预报数据。这个例子说明,同样的数据,随着时间的变化,数据的公共性会发生变化,当然

其本质是不同的数据产品,即当天发布的天气预报数据和一年前的天气预报数据是两个完全不同的数据产品。

谁来提供公共数据呢?从现状来看,首先,一部分公共数据是现有公共部门服务社会的转型升级(例如天气预报数据的网络发布),这部分公共数据仍然可以由原公共部门提供。其次,随着公共数据需求的增大,需要有一个新的公共部门来提供公共数据,就像供水、供电、供气的公共部门一样,将来应该有一个供数的公共部门。目前各地的大数据中心可能可以承担这个角色。上海的“随申办”就是典型的公共数据提供部门。

4 问题解决

依照数据用途来界定公共数据能够有效地解决公共数据的开放开发利用和国家数据安全、个人信息保护形成冲撞的问题。下面针对3类公共数据进行讨论。

第一类,公众要知道的数据。这类数据就是政府数据开放目录中的数据。当前政府开放数据目录应该说已经杜绝了危及国家数据安全和个人信息保护的问题,不用再讨论。

第二类,公众要用的数据。这类数据是公众生产生活中必须要用到的数据,这类数据需要由现有的公共部门或者未来新的公共部门(供数部门)生产、管理,并向公众提供。在公众必需的约束下,其基本不会危及国家数据安全和个人信息保护。况且,针对具体的数据,政府是否将其列入公共数据、是否有能力提供、是否有财力提供,都还需要严格的行政程序管控。

第三类,公共部门要用的数据。这类公共数据的使用者是公共部门,并不向公众提供,加上公共部门原本就持有国家数

据安全和个人数据,因此这部公共数据并不会增加国家数据安全和个人信息保护的风险。

最后简单讨论一下公共部门持有的数据资产。前面讨论过,“国家机关、事业单位,经依法授权具有管理公共事务职能的组织 and 法人在履行公共管理和服务职责过程中收集和产生的数据”不是公共数据或者不完全是公共数据,而是公共数据资产,即数据资产是存在的,由公共部门持

有。公共数据资产需要保值、增值,需要运营变现、上缴财政、服务公共事业,即数据财政的内容之一。

综上,建议政府:

- 依照数据的用途来界定公共数据;
 - 逐步增加公共数据供给以便满足社会正常存在与发展所需,服务社会全面数字化转型后对公共数据的需求;
 - 运营公共数据资产,以增加财政收入。
-