联合国 A/72/540

大会

Distr.: General
19 October 2017
Chinese
Original: English

第七十二届会议
议程项目 72(b)

促进和保护人权：人权问题，包括增进人权和基本自由切实享受的各种途径

隐私权*

秘书长的说明

秘书长谨向大会递移由人权理事会隐私权问题特别报告员约瑟夫·卡纳塔西依照人权理事会第 28/16 号决议提交的报告。

* 在最后限期之后才提交本报告的原因是为了反映最新事态发展。

17-18460 X (C) 311017 021117

人权理事会隐私权问题特别报告员的报告

摘要

本报告分为两部分：第一部分是有关 2016 年和 2017 年所开展活动的执行摘要；第二部分是有关人权理事会隐私权问题特别报告员设立的大数据和开放数据专题行动流工作队所开展工作的中期报告。

目录

页次

一. 隐私权问题特别报告员 2016—2017 年开展的活动概述 4

A. 关于监控和隐私的国际法律文书草案 4

B. 指控函 5

C. 其他函件：公共领域；日本 5

D. 其他正在进行的有关监控的举措 5

E. 对隐私的更好的理解 5

F. 健康数据工作队 5

G. 公司对个人数据的使用 5

H. 正式国家访问 6

I. 提供资源 6

二. 大数据和开放数据工作队 6

A. 界定问题 6

B. 数据 7

C. 大数据 9

D. 先进的分析 10

E. 算法 11

F. 开放数据 15

G. 开放政府 16

H. 大数据的复杂性 17

I. 思考现实：商业大数据和隐私 19

J. 未来的原则：控制数据公开 21

三. 支助文件 23

四. 结论 23

五. 建议 24

一. 隐私权问题特别报告员 2016-2017 年开展的活动概述

1. 2016-2017 年对特别报告员的任务来说是特别繁忙的一年，通过在四个大陆上 15 个国家举行的 26 项活动，与民间社会、政府、执法部门、情报部门、数据保护当局、情报监督当局、学术界、公司和其他利益攸关方进行了接触。为了与之接触，特别报告员造访了 30 多个城市，有些位于亚洲、北非和中美洲，25%的接触发生在美国，在欧洲的超过 50%。

A. 关于监控和隐私的国际法律文书草案

2. 安全和监控是重要的问题，这一问题导致在 2015 年创建了联合国人权理事会隐私权问题特别报告员的任务。

3. 人权理事会第 28/16 号决议所规定的隐私权问题特别报告员的任务明确指出其义务：“查明在增进和保护隐私权方面可能存在的障碍，查明、交流和推动国家、区域和国际层面的原则和最佳做法，为此向人权理事会提交提案和建议，包括针对数字时代的特有挑战提交提案和建议”。[1]

4. 关于网络空间的监控和隐私的国际法存在空白是斯诺登披露事件的实质性核心问题，该问题已被特别报告员确定为隐私方面的一个严重障碍，也是特别报告员目前的主要关切问题。特别报告员认为，促进和保护隐私面临的障碍不仅包括缺乏实质性规则，还包括缺乏适当的机制。[2]

5. 特别报告员根据其任务，将强烈建议人权理事会支持在联合国范围内讨论和通过一项法律文书，以实现两个主要目的：

(a) 向成员国提供一套可被纳入其国家立法的原则和示范条款，体现和执行人权法的最高原则，特别是在涉及监控时的隐私权；

(b) 向成员国提供一系列备选方案，以帮助它们填补国际法方面的差距和空白，尤其是有关网络空间的隐私和监控的国际法。

6. 虽然对法律文书的需求是明确的，但确切范围和形式尚不清楚。尽管通过正在进行的研究和利益攸关方的协商，其实质内容正在明确地形成，但实现这些目的的最佳工具还有待确定。

7. 人们早已认识到，几个无法绝对保障隐私权的领域之一就是犯罪的侦查、预防、调查和起诉，同样，国家安全领域也是如此。然而，维护民主需要制衡办法，以确保采取的任何监控都是为了保护一个自由的社会。如果一个民主社会想要维护定义其性质的各项自由，那么监控的事先授权和监控活动的事后监督就是其所需规则、保障和补救办法的关键部分。

---

[1] A/70/53, 第三节，A 部分，第 28/16 号决议，第 4 (c)段。

[2] 隐私权问题特别报告员提交人权理事会的报告，2017 年 3 月（网上有未经编辑的预发本，见 A/HRC/34/60）。

8. 2017 年 3 月特别报告员提交人权理事会的报告包含对一项法律文书的中期结论，该文书对网络空间的监控作出规定，补充了现有的网络法，例如 2001 年欧洲委员会部长理事会通过的《网络犯罪公约》（《布达佩斯公约》）。一项原有举措，即欧洲联盟支持下的隐私、财产和互联网治理替代管理方案项目，正在探讨规范网络空间监控的法律文书的备选方案。民间社会和国际各公司正在对一份案文草案进行讨论，该草案将在 2018 年春季之前公布。

9. 支助文件五[3]对该程序进行了更加详细的描述。

B. 指控函

10. 特别报告员向政府发送的一些与监控有关的指控函将由人权事务高级专员办事处(人权高专办)根据特别程序任务负责人来文报告予以发布。

C. 其他函件：公共领域；日本

11. 2017 年 5 月 18 日，特别报告员向日本国政府发表了一封函件 (见支助文件三)[4]。在这封函件中，特别报告员对拟议立法的缺陷表示关切，该立法允许在没有必要保障的情况下实施监控，表面上是为了允许日本国批准 2000 年《联合国打击跨国有组织犯罪公约》。特别报告员将继续尝试参与这一事项，并且将在 2018 年 3 月特别报告员提交人权理事会的报告中提到这一点。

D. 其他正在进行的有关监控的举措

12. 还有其他一些举措，其任务是对监控、安全和隐私进行探讨。可酌情在稍后阶段公布细节。

E. 对隐私的更好的理解

13. 特别报告员正在对隐私进行分析，在分析过程中，除其它外，将隐私视为一项必不可少的权利，以此实现人格自由且不受阻碍的发展的首要基本权利。隐私和人格工作组主席 Elizabeth Coombs 是澳大利亚新南威尔士州的前隐私事务专员，她以欣然接受从事这一工作，并尤为关注性别和隐私问题。

14. 关于工作队开展的活动的更多信息，可查阅支助文件四。[5]

F. 健康数据工作队

15. 特别报告员的健康数据工作队在来自美国的 Steve Steffensen 医生的领导下已开始其工作。预计在 2018 年春季和夏季举行协商。

G. 公司对个人数据的使用

16. 特别报告员继续在商业模式和公司使用个人信息过程中的隐私问题(包括独立使用和在隐私、财产和互联网治理替代管理方案项目内使用)方面开展工作，

[3] 见 www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx。

[4] See www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx。

[5] 见 www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx。

以此为特别报告员有关这一议题的工作队的启动进行准备，其时限公布于特别报告员网站(http://www.ohchr.org/EN/Issues/Privacy/SR/Pages/ThematicReports.aspx)。

H. 正式国家访问

17. 已经进行或计划进行了下列国家访问：美国(2017 年 6 月 19 日至 28 日)、[6]法国(日期已确认，2017 年 11 月 13 日至 17 日)；大不列颠及北爱尔兰联合王国(日期已确认，2017 年 12 月 11 日至 17 日)；德国(日期已确认，2018 年 1 月 29 日至 2 月 2 日)；大韩民国(日期已确认，2018 年 7 月 3 日至 15 日)。

I. 提供资源

18. 只有对美国的正式国家访问，以及特别报告员和其他发言者前往中国香港特别行政区参加数据保护和隐私专员国际会议及有关亚洲的人格与信息流动讨论的行程才由特别报告员的任务预算提供资金，该预算由人权高专办管理。其他访问接受外部供资，资金大多来自有关活动的东道国。

二. 大数据和开放数据工作队

19. 特别报告员设立的大数据和开放数据工作队由 David Watts 领导。[7] 本报告的主要撰稿人是 David Watts 和 Vanessa Teague。[8]工作队的许多成员也为本报告做出了贡献，包括 Christian d'Cunha(欧洲联盟的欧洲数据保护监督员)、Alex Hubbard(联合王国信息专员办公室)、Wolfgang Nejdl 教授(德国汉诺威大学)、Marty Abrams(美国信息问责基金会)和 Marie Georges(法国)。Sean McLaughlan、Elizabeth Coombs 和 Joe Cannataci 也为本报告做出了贡献。

20. 关于大数据和开放数据报告起草程序的更多资料可查阅支助文件七[9]。

A. 界定问题

21. 二十一世纪信息社会面临的最严重挑战之一是协调新信息和通信技术所带来的社会福利与保护隐私权等基本权利这一任务。这些新技术有可能协助各国确保尊重、保护和履行它们的人权义务，但是也存在破坏某些人权的风险，特别是隐私权。

22. 收集和分析数据的新方法——大数据现象——以及全世界各国政府为了带来经济增长和促进科学研究越来越愿意公开发布它们所掌握的个人信息，虽然是

---

[6] 对美利坚合众国正式国家访问的最后报告预计将于 2018 年春季发布。任务结束说明可查阅：http://www.ohchr.org/Documents/Issues/Privacy/SR_Privacy/VisitUSA_EndStatementJune2017.docx。

[7] David Watts 是拉筹伯大学和迪肯大学法律系的副教授。2017 年 8 月 31 日之前，他是澳大利亚维多利亚州隐私和数据保护专员。

[8] Vanessa Teague 博士是澳大利亚墨尔本大学计算机与信息系统系的高级讲师。

[9] 见 www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx。

以去识别化的形式——开放数据现象——给构成我们理念基础的许多假设带来了挑战，包括隐私是什么、它必然涉及什么以及如何最好地保护它。

23. 随着人权理事会认识到隐私作为一项对于实现享有尊严的权利以及个人人格自由和不受阻碍地发展来说至关重要的权利（见人权理事会 2017 年 3 月 23 日第 34/7 号决议），大数据和开放数据构成的挑战范围已经扩大。

24. 关于大数据和开放数据的一些主张被贴上了“乌托邦”的标签。[10]这些主张认为，大数据提供了为气候变化、恐怖主义威胁和公共卫生等棘手的公共政策问题提出新见解的途径。观点的另一端是那些持反乌托邦意见的主张者，他们对国家和非国家行为者越来越多的监视、不正当的侵入私人领域以及隐私保护的破坏感到困扰。

25. 在编写本报告时遇到的一项主要挑战是查明并评估这些和其他参与围绕大数据和开放数据的复杂辩论的利益攸关方提出的主张。尽管这两个问题都引起了大量的评论和学术研究，我们对技术及其对未来影响的认识存在差距：自相矛盾的是，缺乏数据会阻碍我们对大数据和开放数据潜在益处和危害的认识。

B. 数据

26. 每天我们的数字活动生成大约 250 万兆字节的数据。[11]这是 2.5 后面跟着 18 个零[12]字节的数据。换个角度看，一部平均 300 页的小说包含大约 3 后面跟着 5 个零字节的数据。全世界所有数据中有 90%是在过去两年里生成的，[13]而且数据生成的速度在持续增长。

27. 在当今相互联接的世界里，数据是普遍和无处不在的。无论何时，只要我们使用计算机、智能手机或者甚至是包含能够记录信息的传感器的日常设备，数据就作为一个副产品生成。它以字符或符号的形式出现，最终被计算设备简化为二进制代码，然后作为电子信号被处理、存储和传输。

28. 大数据使用的数据来源因使用互联网进行的活动不同而各异：“数据来自许多不同的来源，包括科学仪器、医疗设备、望远镜、显微镜、卫星；数字媒体，包括文本、视频、音频、电子邮件、博客、推特订阅、图像采集、点击流和金融交易；动态传感器、社交网络和其他类型的网络；科学模拟、模型和调查；或观测数据的计算分析。数据可以是时间的、空间的或动态的；结构化或非结构化的；

[10] Danah Boyd and Kate Crawford, “Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon”, Information, Communication and Society, vol 15, No. 5。

[11] 见 https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html。

[12] 这是美利坚合众国使用的计算。在大不列颠及北爱尔兰联合王国，此单位是 1 后面跟着 30 个零。

[13] 见 www-01.ibm.com/software/data/bigdata/what-is-big-data.html。

从数据中获得的信息和知识在表示、复杂性、粒度、上下文、来源、可靠性、可信赖性和范围方面存在差异。数据也可能在生成和访问的速率上有所不同。”[14]

29. 一些生成的数据与个人无关。它是从分析气候模式、太空探索、材料或设计的科学测试或金融市场证券交易相关风险等活动中产生的数据。但是，很大一部分是我们自己生成的数据或者是生成的关于我们的数据。本报告的侧重点是这一类数据——个人信息——不论是通过提供、观察、推导或推断产生的。[15]

30. 个人信息捕捉我们作为人类的个性。正是这一确认每一个人的能力使得个人信息如此宝贵。

31. 我们自己生成的数据涉及到我们自己的代理。它包括我们的电子邮件和文本信息，以及我们生成和分享的图像和视频。其他关于我们的数据是由第三方生成的，但是我们至少在一定程度上参与了它的生成，例如，电子健康记录或电子商务交易。

32. 但是其他关于我们的数据则是通过不明显的方式产生的，因为它发生在幕后，在不透明和对我们来说在很大程度上未知——和不可知的情况下。它包括“电子琐碎信息”、[16]电子产品和作为我们线上和线下活动产物的其他电子轨迹。这一数据可包含我们的移动设备连接到移动电话信号塔或全球定位系统（GPS）卫星的时间和地点、我们访问网站的记录，或由数字闭路电视系统采集的图像。这些我们留下来并且可能在计算机服务器上永久保留的“电子琐碎信息是关于我们是谁、我们做什么以及我们想要什么的线索。这使个人数据——关于个人的数据——对于公共利益和私营公司都非常宝贵。”[17]

33. 一个被数据、计算机处理和即时数字通信吞没的世界提出了隐私权如何能够与新技术共存的问题，后者使个人信息能够以一种在起草 1948 年《世界人权宣言》和 1966 年《公民权利和政治权利国际公约》时无法想象的方式被收集、处理和分析。

34. 由于无处不在的计算机中介，几乎世界的每一个方面都以新的符号维度呈现，如事件、物体、过程，人们以一种新的方式变得可见、可知和可共享。由于数据和电子文本在规模和范围上是普遍的，世界获得了重生。[18]

---

[14] United States, National Science Foundation, “Critical techniques and technologies for advancing big data science and engineering (BIGDATA)”, Program Solicitation NSF 14-543, 可查阅：www.nsf.gov/pubs/2014/nsf14543/nsf14543.pdf at p3。

[15] Martin Abrams, “The origins of personal data and its implications for governance”, 可查阅：http://informationaccountability.org/wp-content/uploads/Data-Origins-Abrams.pdf。.

[16] Evan Schwartz, “Finding our way with digital bread crumbs”, *MIT Technology Review*, 18 August 2010。可查阅：www.technologyreview.com/s/420277/finding-our-way-with-digital-bread-crumbs/。

[17] Julie Lane and others, eds., *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (New York, Cambridge University Press, 2014)。

[18] Shoshana Zuboff, “Big other: surveillance capitalism and the prospects of an information civilization”, *Journal of Information Technology*, vol. 30, No. 1 (March 2015)。

35. 信息和通信技术使个人通过分析他们的数据变得可知，其方式涉及到“将一个人的本质视作是由此人的信息所构成的”。[19]使这一点成为可能的现象被广泛称为大数据。

C. 大数据

36. “大数据”一词通常用于描述数量巨大且在不断增加的数据以及用来搜索、关联、分析并从中得出结论的高级分析技术。

37. 对于“大数据”没有一致的定义。美国国家标准和技术研究所将其描述为传统的数据架构无法高效处理新的数据集。大数据推动新架构的特征有：

(a) 数量(即数据集的大小)；

(b) 种类(即来自多个储存库、域或者类的数据)；

(c) 速率(即流动的速度)；

(d) 变率(即其他特征发生的变化)。

38. 这些特征——数量、种类、速率和变率——也被称为大数据的四个“V”属性。[20]

39. 美国国家标准和技术研究所的描述以及很多其他试图明确大数据现象的说法(如欧洲联盟称“‘大数据’是指由大量不同来源快速产生的大量数据”)[21]把注意力引向了共同使收集、处理和分析海量数据成为普遍现实的各种技术。然而，这些描述的高度概括及其对于技术的重点关注并不足以说明大数据现象。

40. 一些专家对大数据进行了比“V”属性更为广泛的详尽描述。有一种较为详细的实用说法是大数据：

(a) 数量大，含万亿字节或千万亿字节数据；

(b) 速率高，实时或接近实时创建；

(c) 种类多，分结构化和非结构化两种性质；

(d) 范围穷尽，力求捕捉整个群体或系统；

(e) 分辨率细粒度高，唯一索引识别；

(f) 本质上具有关联性，具有有利于不同数据集结合的公用区；

[19] Luciano Floridi, “Four challenges for a theory of informational privacy”, *Ethics and Information Technology*, vol. 8, No. 3 (July 2006)。

[20] 还有其他的“V”属性，这四个是关键的驱动因素。见 http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf。

[21] 见 https://ec.europa.eu/digital-single-market/en/policies/big-data。

(g) 灵活，增加新域方便，可快速扩大规模。[22]

41. 特定大数据并不一定包含上述所有特征。

42. 其他方法则认为大数据不仅仅是一种技术现象：“我们把大数据定义为一种取决于以下几个方面相互作用的文化、技术和学术现象：

“(a) 技术——使采集、分析、关联和比较大型数据集的计算能力和算法精度最大化；

“(b) 分析——利用大型数据集识别模式，以提出经济、社会、技术和法律主张；

“(c) 神话学的——普遍认为大型数据集产生更高的智慧和知识形式，可以形成前所未有的见解，带有真实、客观和准确的光环。”[23]

43. 大数据的支持者提出的一大主张是：它可以解决由于缺乏经验证据(即缺乏数据)而使研究受到的限制问题，为我们提供关于各种情况或现象的客观真实。这些认识论主张往往把大数据提升为一种新的科学方法，是许多人对大数据的局限性和带来的风险表示不安的核心所在。

44. 人们普遍认为，大数据能够产生包括个性化服务、增加获取服务的机会、提高健康成果、技术进步和无障碍改善等在内的社会效益。[24]欧洲联盟委员会指出，“理解‘大数据’的必要使技术创新，并使新工具和新技能得到开发。”[25]

45. 欧洲联盟委员会把信息当作一种经济资产，同劳动力和资本一样对社会具有重要的意义。[26]值得注意的是，这个市场由少数大技术公司主导，其市场份额依赖于数据的使用情况。

D. 先进的分析

46. 关键的变化是大量使用数据以通知算法，而算法的后续行为取决于它所访问的数据本身。

“机器学习这个术语指的是自动检测数据中有意义的模式。几十年来，它几乎成了任何需要从大型数据集中提取信息的任务中的常用工具……

[22] Rob Kitchin, “Big data, new epistemologies and paradigm shifts”, *Big Data and Society*, vol. 1, No. 1 (April-June 2014)。

[23] Boyd and Crawford, “Critical questions for big data”。

[24] 也有截然不同的观点。例如，见 2014 年 9 月 16 日关于开发有关个人保护的大数据对欧洲联盟处理其个人资料的影响的欧洲联盟 29 条数据保护工作组声明：“虽然大数据的真正价值尚有待证实，但人们期望通过开发大数据获得众多个人和集体利益。工作组自然会支持欧洲联盟或国家一级为欧洲联盟的个人实现这些利益所作出的真正努力，无论是个人还是集体”。见 http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf。

[25] 见 https://ec.europa.eu/digital-single-market/en/making-big-data-work-europe。

[26] 同上。

所有这些应用程序有一个共同的特点，即相对于传统的计算机用途，在这些情况下由于需要检测的模式的复杂性，人工程序员无法清楚、详细地说明应该如何执行这样的任务……

机器学习工具关系的是为程序赋予学习和适应的能力。”[27]

47. “现在”和“当时”之间的主要区别是新技术的自主和半自主性。

48. 最常用的分析技术之一就是所谓的“数据挖掘”。这是一个从大型数据集中提取数据并进行分析以确定模式或相关性是否存在的过程。数据挖掘有助于大量原始数据[28]的简化和总结并从显示的模式中推断知识。

49. 驱动这些技术和工具的引擎是算法。

E. 算法

50. 算法并不是新出现的。它们“从一开始就一直存在，并且在创造出特定词汇来描述它们之前就已存在。”[29]

51. 算法不仅限于数学。巴比伦人利用算法来决定法律条文，拉丁语教师用算法来确保语法正确，而且算法在所有文化中都被用于预测未来，用来决定医疗措施或准备食物。现今，人们在按照菜谱烹饪、使用某种针织花纹或是操作家庭用具时，通常都在无意识地使用这样或那样的算法。[30]

52. 与大数据的其他要素一样，“精确地表述出算法的特性异常困难。”[31] 就本报告来说，有用的工作性定义是：

“一套执行某种程序或解决某个问题的具体指令集，通常要求程序在某一时刻终止。具体的算法有时也被称为方法、程序或技术……将某种算法应用于输入以获得输出的过程被称为计算。”[32]

53. 将用于烘培蛋糕的算法与评估个人信贷价值的算法区分开来的是所处理数据涉及的自动化程度、其自主性、非线性、性质和数据量。

54. 透过算法，我们越来越多地了解我们自身以及我们与世界的关系。算法现在是信息社会的重要组成部分，越来越多地支配着“之前应由人类进行的操作、决

[27] Shai Shalev-Shwartz and Shai Ben-David, *Understanding Machine Learning* (New York, Cambridge University Press, 2014)。

[28] 仅与一人有关的数据。

[29] Jean-Luc Chabert, ed., *A History of Algorithms: From the Pebble to the Microchip* (Berlin, Springer-Verlag, Berlin, Heidelberg, 1999)。

[30] 同上。

[31] Felicitas Kraemer, Kees van Overveld and Martin Peterson, “Is there an ethics of algorithms?”, *Ethics and Information Technology*, vol. 13, No. 3 (September 2011)。

[32] 见 http://mathworld.wolfram.com/Algorithm.html。

定和选择”。[33]它们在交友网站上推荐匹配的人，[34]决定最佳出行路线[35]以及评估人们是否为低风险信用对象[36]。它们被用于分析-识别个人特征和行为模式以进行个性化预测，如我们可能倾向于购买的商品或服务。它们确定应如何解读数据，以及应因此采取何种行动。它们“调解社会进程、商业交易、政府决策以及我们如何看待和理解我们自身与环境及之间的互动。”[37]

55. 从个人角度来看，算法处理产生的建议和决定似乎来自于一个神秘、未知的黑盒子，这是一种二十一世纪的德尔斐神谕，看起来似乎让无可争辩和权威的声明脱离于人的能动性。解开算法处理的机制并进而评估它们构成的风险，这是非常复杂的，需要考虑大量的问题。这种复杂性阻碍了我们了解算法如何运作以及它们如何影响我们生活的能力。

56. 越来越多的文献强调算法可能带来的问题，敦促我们在还没有思考我们为管理风险所需要的保障措施的情况下就一头扎进算法的未来之前保持谨慎。

1. 算法负载有价值观

57. 与让它们呈现出客观性的算术结构相反，算法“不可避免地负载了价值观”。[38] 它们所包含的价值观通常反映了设计它们的软件工程师的文化或其他设想，工程师将这种设想作为未明确说明的观点嵌入了算法的逻辑结构中。

58. 例如，信用评分算法可能被设计为去询问某人的出生地点、学校、居住地点以及就业状况。这些测算代替物的选择涉及到价值判断，即这些问题的答案与评估是否应提供贷款有关，如果可以，那么以何种条件提供贷款。不管怎样，贷款申请人往往都无法了解某个特定贷款决定的理由，也无法确定所采用的价值判断。

59. 尽管这些数据代替物在某些社会中可能与信用评估有关，但在其他社会中，它们最多也就是毫无益处的干扰，或甚至会带来损害。例如，在一些可能大部分人口都没有固定地址、几乎未受过正规教育以及可能自营职业的发展中国家，采用这种算法可能会让人们永远无法获得贷款。

---

[33] Brent Mittelstadt and others, “The ethics of algorithms: mapping the debate”, *Big Data and Society*, vol. 3, No. 2 (July-December 2016)。

[34] 例如，见 Rebecca Harrington, “Dating services tinker with the algorithms of love”, *Scientific American*, 13 February 2015。可查阅：www.scientificamerican.com/article/dating-services-tinker-with-the-algorithms-of-love/。

[35] 见 https://motherboard.vice.com/en_us/article/4x3pp9/the-simple-elegant-algorithm-that-makes-google-maps-possible。

[36] 见 Michael Byrne, “The simple, elegant algorithm that makes Google Maps possible”, 22 March 2015。可查阅：http://mitsloan.mit.edu/media/Lo_ConsumerCreditRiskModels.pdf。

[37] Mittelstadt and others, “The ethics of algorithms”。

[38] 同上。

60. 另一方面，分析非传统形式数据的算法可能表明某个没有常规信用记录的人反而具有良好的信用风险，从而促进人类发展。[39]

2. 不完善的数据带来的问题

61. 支撑算法的原材料是数据，但并非所有数据都是精确、足够全面、最新或可靠的。[40] 一些数据的来源，例如税收记录，通常都可以便利地建立起来，但其准确性在某个州内和各州之间的税务机构都各不相同。其他数据源可能是从过时的从未真正整理过的数据库中抽取的，或是来自于不安全的来源，或是出现了不恰当数据输入和记录-保存标准的情况。

62. 算法的作用是处理数据，因此，它们“面临所有数据处理类型都面临的局限性，即输出永远不会超过输入。”[41] “无用输入，无用输出”原则在这里也适用。

3. 数据的选择

63. 选择数据方面的风险与上文第 62 段所述风险类似。正如低质量数据产生低质量结果一样，选择不恰当或不相关的数据也会产生不可靠和误导性结果。

64. 大量的算法处理涉及归纳推理和识别明显不相干数据片段之间的相关性。如果使用了错误的数据，任何建议或决定都将存在缺陷。

4. 偏见、歧视和固化弱势状态

65. 虽然一些专家在偏见和歧视之间划分了界限，[42]但在大数据背景下它们带来的风险非常类似，有足够理由将它们放在一起来讨论。

66. 算法可用于罪犯的特征分析，也就是利用机器学习能力，“在群体层面确认相关性并对行为做出预测，虽然群体(或特征)在不断变化，算法也在重新定义”：

> “无论是动态的还是静态的，个体都因与算法确定的其他人之间的联系而不是实际行为而被捕。 根据群体的信息固化了个体的选择。罪犯特征分析无意间带来了造成歧视的证据依据。”[43]

67. 一些评论家认为，高级分析技术，如罪犯特征分析，强化了人们的弱势状态。例如预测警务，它利用犯罪统计和基于算法的分析来预测犯罪热点地区并将其作

---

[39] United States, Federal Trade Commission, “Big data: a tool for inclusion or exclusion—understanding the issues” (2016)。可查阅：www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf。

[40] 例如，少数群体的利益在特定数据集中没有得到很好的代表，可能受到随后基于此类信息做出的决定和预测的影响。

[41] Mittelstadt and others, “The ethics of algorithms”。

[42] 偏见被认为是持续或反复表达某种决策倾向、价值观或信仰。歧视是算法决策可造成的负面、不相称影响。

[43] Mittelstadt and others, “The ethics of algorithms”。

为执法机构的优先重点。[44] 由于在热点地区部署了大量警力维持治安，并且热点地区通常位于社会弱势地区，而不是发生白领犯罪的地区，因此更多的警务往往会将逮捕和定罪局限在某一个地方，而这又导致一种恶性循环，同一热点地区一再得到确定并得到强化，使那些弱势人群面临被依照刑法逮捕和受到惩罚的更高风险。

68. 政府可能利用这类工具控制、针对或以其他方式伤害某些社区，这也引起了人们的关切。[45]

5. 责任和问责

69. 算法处理带来的危害可大体归因于处理大量不相干数据集所面临的困难以及用于处理数据的算法的设计和实施。由于涉及如此多的变量，很难查明谁应对造成的任何伤害负责。通常，大数据分析基于发现和探索，而不是测试特定的假设，因此在一开始很难预测(以及对个人来说，阐明)数据使用的最终目的。

70. 算法不透明不一定是“既定事实”；在技术上是可以在算法处理的每个阶段保留使用的数据和应用后的结果的。

6. 对隐私的挑战

71. 经济合作与发展组织(经合组织)在 1980 年出版了《保护隐私和个人数据跨界流动指导方针》[46]。经合组织指导方针中的八项原则以及 1981 年欧洲委员会《关于在自动处理个人数据方面保护个人的公约》和大会在其 1990 年 12 月 14 日第 45/95 号决议中通过的《电脑个人数据档案的管理准则》中的类似原则，影响了全世界的信息隐私法律。

72. 经合组织指导方针和欧洲委员会的《数据保护公约》中的基本原则，即收集限制原则，指的是个人信息只能合法、公平地收集，并酌情征得相关个人的知情和同意。[47] “具体说明目的原则”要求在收集时应说明收集个人信息的目的，随后的信息使用应仅限于收集的目的或与之相容的目的，而且在目的发生变化的任何时候，都应当予以说明。[48] “使用限制原则”限制出于不相容的目的披露个人信息，除非个人同意或合法当局授权。[49] “数据质量原则”受到了数据大量收集以及仅处理适当、相关及不过度个人信息要求的挑战。1990 年联合国

---

[44] 例如，见 www.predpol.com/how-predictive-policing-works/。

[45] Lee Rainie and Janna Anderson, “Code-dependent: pros and cons of the algorithm age”, Pew Research Center, 8 February 2017。可查阅：www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/。

[46] 经济合作与发展组织(经合组织), “OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data”。可查阅：www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm。

[47] 见经合组织隐私原则。可查阅：http://oecdprivacy.org/。

[48] 同上。

[49] 同上。

《电脑个人数据档案的管理准则》提出了在出于数据处理目的而保留数据中的相称原则。

73. 大数据对这些原则提出了挑战，同时带来了因算法使用考虑不周造成的道德问题和社会困境。这非但没有解决公共政策问题，反而无意中带来了削弱人权的后果，例如免受一切形式歧视的自由，这包括针对妇女、残疾人和其他人的歧视。

74. 同时，有迹象表明，在算法设计上出现了思维模式的变化，为大数据算法带来了更好的算法解决方案，如电气和电子工程师学会标准协会关于道德化设计的举措。[50]

75. 在隐私方面，相关国际文书将隐私权的含义扩展到信息隐私权之外，而信息隐私权是经合组织指导方针中的原则和欧洲委员会《数据保护公约》的重点。鉴于将隐私认定为对享受其他人权至关重要的一项授权权利，以及一项与人类尊严及自由和不受阻碍地发展人格密切相关的权利（见人权理事会第 34/7 号决议），大数据对隐私构成的挑战扩大到各类人权的纳入。大数据通过将个人信息细节透露给那些收集和分析个人数据踪迹的人从而侵入人们生活的倾向，完全违背了隐私权和保护隐私权所认可的原则。

76. 监管的影响与新出现的行业和政府做法中显而易见的变化一样重大。

F. 开放数据

77. 开放数据是与先进的分析同步流行起来的一个概念。它旨在鼓励私营和公共部门向公共领域发布数据，以鼓励透明和公开，特别是在政府当中。

78. 开放数据的定义是：

> “……一类可以被任何人免费使用、再利用、再分发的数据——在其限制上，顶多是要求署名和使用类似的协议再分发。”[51]

79. 开放数据事实上可包含任何类别的数据。开放知识基金会将其总结为：

(a) 文化：关于文化作品和艺术品的数据——例如标题和作者——通常由美术馆、图书馆、档案馆和博物馆收藏及持有；

(b) 科学：作为科学研究一部分产生的数据，范围包括从天文学到动物学等不一而足；

(c) 金融：如政府账户(支出和收入)等数据以及金融市场信息(股票、股份、债券等)；

(d) 统计：统计局产生的数据，如人口普查和主要社会经济指标；

---

[50] 电气和电子工程师学会(IEEE), IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, *Ethically Aligned Design: A Vision for Prioritizing Wellbeing with Artificial Intelligence and Autonomous Systems*, ver. 1 (IEEE Press, 2016)。可查阅：http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf。

[51] 见 http://opendatahandbook.org/guide/en/what-is-open-data/。

(e) 气象：用于了解和预测天气及气候的诸多类型信息；

(f) 环境：与自然环境有关的信息，如污染物的存在和水平，河流和海洋的水质。[52]

80. 为满足定义的要求，开放数据通常根据知识共享许可进行发布。只要满足归属要求，知识共享许可 CC BY 4.0 允许对经许可的材料进行无限制复制、再分发和改编(包括用于商业目的)。[53]

81. 政府持有的有关其公民的数据不属于任何这些类别。开放数据和开放政府旨在提供获得有关政府本身和我们所处世界的数据的途径。它并没有打算包含政府收集的公民数据。因认识到这一点，某些司法系统将“个人”和其他如商业或内阁机密信息等明确排除在开放数据之外。[54] 在面对诸如“分享”和“联系”这样的术语时，我们不应忽视情况已经发生逆转，即政府正在发布有关其公民的数据，而不是发布有关政府如何运作以及公众可用于对政府问责的数据。

G. 开放政府

82. 奥巴马政府采取的首批行动之一就是颁布一项行政命令，鼓励发布政府拥有的信息，以获得公众信任并促进透明、参与和协作。[55]

83. 在此之后，开放政府伙伴关系得以成立。开放政府伙伴关系于 2011 年 9 月发布了《开放政府宣言》。[56]《宣言》侧重于向个人提供更多有关政府活动的信息，并且强调需要加强公民参与和提升政府透明度、打击腐败、增强公民权能以及利用“新技术力量使政府更为高效和负责”。

84. 《开放政府宣言》让其成员在没有约束力、自愿的基础上承诺：

(a) 增加有关政府活动信息的可获得性；

(b) 支持公民参与；

(c) 在整个行政部门实施最高标准的职业操守；

(d) 增加获取开放和问责的新技术。[57]

85. 奥巴马政府发布第一道行政命令之后，又于 2013 年 5 月 9 日发布了另一道行政命令，争取让所有美国政府信息公开并默认机器可读。[58] 其重点与之前 2009

[52] 见 https://okfn.org/opendata/。

[53] 见 https://creativecommons.org/licenses/by/4.0/。

[54] 澳大利亚，新南威尔士州政府，“Open data policy”, Department of Finance and Services, 2013。

[55] 奥巴马总统，“透明与开放政府”，2009 年 1 月 21 日，给行政部门和机构首长的备忘录。可查阅：https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government。

[56] 见 https://www.opengovpartnership.org/open-government-declaration。

[57] https://www.opengovpartnership.org/open-government-declaration。

[58] 奥巴马总统，2013 年 5 月 9 日行政命令，“Making open and machine readable the new default for Government information”。可查阅：https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-。

年的命令相比已经发生变化。关于开放政府数据，这项命令称：“促进向公众提供高效和有效的服务，为经济增长做出贡献。作为开放政府的一项重大好处，让信息资源易于搜寻、获取和利用可以激励企业家精神、创新和科学发现，从而改善美国人的生活，大力促进创造就业机会。”[59]

86. 在接下来的几年，开放数据已经发展到这样一个阶段，即在 2017 年，开放数据的目标已经不仅仅是要在公共领域发布从不属于个人信息的数据或不是从个人信息衍生出的数据，而且还寻求发布去识别化后的个人信息。这种做法的支持者声称，政府数据库或其他信息储存库掩藏了许多“有价值”的信息，让这类信息公诸于众将有助于刺激信息经济的增长。

87. 因此，源自个人信息的开放数据完全依赖于“去识别化”过程的效果，以防止重新识别并将信息与其起源的个人再次联系起来。关于去识别化是否提供了保护隐私和“有益于研究”的数据的辩论，已被证明存在巨大争议。

H. 大数据的复杂性

88. 2015 年，澳大利亚记者威尔·奥肯登(Will Ockenden)在网上公布了他的电信元数据，并请人们告诉他从中可对他的生活做出怎样的推断。元数据包含所有电话呼叫和短信的确切时间，以及距离最近的信号塔。虽然他用假名代替了电话号码，但是仅根据通信和位置模式就可轻松准确地回答出像“我的母亲住在哪里”这样的问题。这并不复杂——观众们只是(正确地)猜测出他的母亲居住在他圣诞节曾拜访过的地方。

89. 这是隐私研究的一个重要主题：数据中的模式，即使没有姓名、电话号码或其他明显的标识，也可用于识别一个人，从而从数据中提取更多有关他们的信息。当这些模式被用于将许多不同的数据集联系起来以逐步形成某人的复杂形象时，这就尤为强大。

90. 一些数据必然会暴露出来。电话公司知道每个客户正在拨打什么号码，医生知道他们的病人的检验结果。因此，向其他人(如公司或研究者等)公开这些数据，以及政府可利用信息进而影响其公民行使人权的方式，都会引起争议。

91. 其他数据是故意获取的，通常个人并不知情或做出同意。电子前沿基金会的研究人员发布了“panopticlick”实验的结果，这个实验表明，可以根据简单的特征(如插件和字体等)鉴别个人的网络浏览器。[60] 他们警告称，网络浏览隐私是存在风险的，除非对这些特征的存储及其与浏览历史的联系设置限制。没有重大的政策变化。在 2017 年，网络浏览隐私已不复存在。许多公司通常出于商业原因例行地对人们进行蓄意跟踪。网络跟踪现在几乎无处不在，只有付出巨大的努力才能规避。

[59] 同上。

[60] Peter Eckersley, “How unique is your web browser?” in *Privacy Enhancing Technologies*, Mikhail Atallah and Nicholas Hopper, eds. (Berlin, Springer-Verlag, 2010)。

92. 现代互联网经济大部分依赖于收集有关潜在客户的复杂数据，以向他们出售商品，这种做法被称之为“监控资本主义”。[61] 然而，与童工相对工业经济相比，监控相对于数据驱动效率看起来并没有更为合理。这只是利用信息的最便捷途径。它绝少像隐私权那样被视为一项基本权利。事实上，如果最低标准和经过改进的技术迫使公司和政府迈入一个普通人对其自身数据有更大掌控能力的世界，数据驱动的经济是可以存活和繁荣起来的。[62]

93. 各国政府利用更为合法的许可，也能够进行革新。社区对政府的信任程度强烈地影响着他们如何看待开放数据和开放政府举措的可能影响。信任政府的人更有可能认为开放数据会带来好处。[63] 研究表明，人们在很大程度上可以接受政府提供关于其社区的在线数据，尽管当数据触及切身利益时他们会发出警告。公民的接受程度不同，视正在讨论哪个数据收集领域而定。[64]

94. 大多数信息隐私法对收集和处理个人信息做出了规定：如信息不是“个人信息”，则不受信息隐私法管理。许多这样的法律承认，个人信息可去识别化，这样就可出于公共利益研究的目的，以不干涉个人信息隐私权的方式，对其加以利用或处理。各国政府和其他人通过保证去识别化，力求维持他们所收集数据的人的信任。

95. 这促成了重要的思考“去识别化过程提供的数据不会干涉个人信息隐私权吗”？

96. 如汇总统计等这类简单数据类型是可以真正进行隐私保护处理的，如差分隐私。差分隐私算法进行规模化处理效果最好，正被纳入商业数据分析。实现差分隐私的随机算法是隐私工具库中的一项宝贵工具，但它并没有为单元记录[65]级别个体数据高度复杂的数据集提供一揽子去识别化方式。苹果公司 2016 年使用这些技术就是如何大范围利用差分隐私的一个实例。[66]

---

[61] Shoshana Zuboff, “Big other: surveillance capitalism and the prospects of an information civilization”, *Journal of Information Technology*, vol. 30, No. 1 (March 2015)。

[62] 公司和政府并不一定需要被迫提供隐私保护。例如各公司采取的道德做法，见信息专员办公室，“Big data, artificial intelligence, machine learning and data protection”, ver. 2.2 (2017)。可查阅：https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf。

[63] John Horrigan and Lee Rainie, “Americans’ views on open Government data”, Pew Research Center, 21 April 2015。

[64] 同上。

[65] 仅与一人有关。

[66] Andy Greenberg, “Apple’s ‘differential privacy’ is about collecting your data — but not your data”, 13 June 2016。可查阅：https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/https://techcrunch.com/2016/06/14/differential-privacy/ https://arxiv.org/abs/1709.02753。。

97. 如果不充分降低其实用性，高维单元记录级别数据则不能安全地去识别化。这是纵向跟踪个人健康、出行、网络搜索等数据形成的数据类型。支助文件一[67]提供了去识别化工具和争议的概述。

开放政府数据

98. 存在许多成功在政府公布的数据中再次识别个人的例子。[68] 这种“公开再次识别”的公开有两个层面的意思：结果是公开的，以及再次识别仅是利用公开的辅助信息。

99. 可用辅助信息越多，越容易再次识别大量的个人。由于将更多的数据集联系起来，再次识别所必需的辅助信息就会减少。数据集公开和相互建立联系将同一地点关于个人的大量辅助信息聚集起来，使得更为容易再次识别任何与他们有关的数据。

100. 开放数据的再次识别仅是一个更大问题的缩影，即例行出售、分享和交易的“去识别化”商业数据集的可再次识别性。

101. 在大数据和开放数据时代，反对隐私权是强大的力量。允许最小可能的去识别化可能是所有出于商业或其他目的涉足数据的人在经济上最倾向于选择的，而且各国政府不仅在开放获取个人数据方面面临压力，而且在监管这类获取行为方面也面临压力。

102. 非政府组织对在未适当考虑个人参与的情况下大数据的增长以及因个人私人信息管理或监管不当带来的道德和法律问题表达了关切。[69] 这些组织将继续倡导进行充分的保护和采取适当的行动。

I. 思考现实：商业大数据和隐私

103. 数据收集的指数级增长和急于在表面上将每个对象都联网而不充分考虑数据安全，为个人和团体带来了风险。在努力向消费者和个人保证识别信息的安全的过程中，已经在公共域传播了许多观念。例如，从用户对匿名的错误感受中受益的行业培养了对高度复杂的“匿名”数据的观念。[70]

---

[67] 见 www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx。

[68] 在 2005 年 6 月 15 日向国土安全部隐私与廉政委员会提供证词时，斯威尼称，那是在 1997 年，她“能够展示仅根据生日、性别和邮政编码，时任马塞诸塞州州长威廉•维尔德的医疗记录是如何被再次识别的。事实上，美国人口的 87%通过出生日期（如年、月、日）、性别及其 5 位数的邮政编码是可以被唯一识别的。重点是看起来匿名的数据并不一定是匿名的”。见 www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_06-2005_testimony_sweeney.pdf; 另见 Latanya Sweeney, “Matching known patients to health records in Washington state data”, Harvard University, 2012。可查阅：http://dataprivacylab.org/projects/wa/1089-1.pdf 和 http://dataprivacylab.org/index.html; Latanya Sweeney, “Achieving k-anonymity privacy protection using generalization and suppression”, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, No. 5 (2002)。

[69] 见 www.privacyinternational.org/node/8。

[70] 即使匿名，这也不能消除与隐私原则和如“同意”这样的考虑的关联。

104. 大量数据是在普通用户并不了解或不知情的情况下进行采集的。这些数据可以出售和与其他来源数据建立联系，从而生成有关个人生活方方面面的复杂记录。这种信息有许多用途，包括政治控制，如美国一个政治组织无意间暴露的数据集所展现出的那样。[71] 数据集包含近 2 亿美国选民的个人详细资料，以及采集(或猜测)的有关他们政治信仰的惊人细节。在中国，有一个社会信用项目，它不仅要评价公民的金融信誉，而且还旨在跟踪他们的社会行为和可能的政治行为。这个项目依赖各种来源的数据，主要是一段时间的在线来源。[72]

105. 数据代理(收集消费者个人信息并向他人转售或分享这类信息的公司)是大数据经济的重要参与者。在开发他们的产品时，数据代理获得来自各种来源的大量各类有关消费者的详细具体信息；[73]分析信息并对消费者做出判断，其中一些可能是敏感信息；并与各行业客户分享信息。所有这些活动都是在消费者不知情的情况下进行的。[74]

106. 虽然数据代理的产品有助于防止欺诈、增进产品供应和提供个性化服务，但数据代理收集和使用数据的许多用途给消费者带来风险。存在的关切是缺乏透明度、收集关于年轻人的数据、无限期保留数据，以及出于确定资格或非法歧视性目的而使用这类数据。[75]

107. 欧洲议会最近关于欧洲隐私监管的报告草案建议：“应向最终用户提供一系列隐私设置选项，范围从较高(例如，‘绝不接受任何“饼干”’)到较低(例如，‘始终接受“饼干”’)以及中等……。”[76]

108. 增加个人对因特网隐私的控制的需求正得到广泛讨论。个人利用他们自己的装置和他们的数据，来获得他们需要的信息，例如地图和方向，并查看他们感兴

---

[71] Sam Biddle, “Republican data-mining firm exposed personal information for virtually every American voter”, *The Intercept*, 19 June 2017。可查阅：https://theintercept.com/2017/06/19/republican-data-mining-firm-exposed-personal-information-for-virtually-every-american-voter/。

[72] “China invents the digital totalitarian state”, *The Economist*, 17 December 2016。可查阅：https://www.economist.com/news/briefing/21711902-worrying-implications-its-social-credit-project-china-invents-digital-totalitarian; Lucy Hornby, “China changes tack on ‘social credit’ scheme plan”, *Financial Times*, 4 July 2017。可查阅：www.ft.com/content/f772a9ce-60c4-11e7-91a7-502f7ee26895。

[73] 报告中存在许多从智能设备如电视、“私人电器”、儿童玩具和“联网汽车”拼车应用程序等获得大型商业数据的例证。

[74] 美国参议院商业、科学和运输委员会，“A review of the data broker industry: collection, use, and sale of consumer data for marketing purposes”, staff report, 18 December 2013。可查阅：http://educationnewyork.com/files/rockefeller_databroker.pdf。

[75] 美国联邦贸易委员会，“Data brokers: a call for transparency and accountability”, May 2014。可查阅：www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf。

[76] Marju Lauristin, “Draft report on the proposal for a regulation of the European Parliament and of the European Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC”, European Parliament, Committee on Civil Liberties, Justice and Home Affairs, 2017。

趣的广告。在这方面，必须要问的是，虽然促进最终用户控制的技术很重要，但个人能在多大程度上运用充分的保护控制？这些工具的采用与当前主导互联网的经济力量存在冲突。[77] 政府在开发和采用这些工具方面是否发挥作用？

控制数据收集的技术

109. 控制(包括终止)数据收集与个人不希望分享的数据有关。在“传统”技术下，这不是一个考虑因素，用户必然处于控制地位，因为技术能做的就只有用户确认：在摄像头上有物理覆盖物的设备或是可手动拔下的仅通过以太网连接的因特网连接。现在有内部无线网络和无盖摄像头。电视机有不能关闭的麦克风。手动禁用功能已经消失，但是也有阻碍数据收集的技术。[78] 非常成功的“无处不在的传输层安全”运动指的是大多数因特网流量现在都已加密，不大可能在传输中被用户不知道的实体收集信息。需要进一步探讨和支持这类技术带来的好处。

110. 模糊身份和职业的想法也并不新鲜——如一些社交网络的“实名”政策与有些人为捍卫其假名注册权利所作努力之间的较量。要模糊化则需要允许用户提交“保留”简介并使之与他们选择呈现的其他简介区别开来的工具。

111. 研究始终表明，如果个人担忧与他们打交道的组织对个人信息采取的做法，那么他们更有可能提供不准确或不完整的信息。[79] 因为隐私和数据保护产生信任，它们有利于数据质量和数据分析。用户对隐私的信任对于机器学习算法的稳定性和准确性也很重要。普通的机器学习很容易受到故意的人为混淆输入的影响。[80] 如果大量的人因担忧隐私问题而故意采用模糊自身的工具，会发生什么呢？

112. 无视察觉到的隐私管理商业做法、隐私方面的信任以及个人行为方式之间的复杂相互影响，对大数据-开放数据采取简单化办法，不会促进“大数据”，只会造成潜在的不精确和低质量的决策。

J. 未来的原则：控制数据公开

113. 隐私法倾向于以能够让足够的灵活性在出现隐私风险时能加以应对的原则为基础。考虑是否需要额外原则来补充现有隐私原则，以保护个人数据免受技术上的隐私侵犯，是有其价值所在的。

---

[77] 例如，通过自动点击向某人呈现的所有广告以混淆用户真正阅读过的广告的方式，AdNauseum 打败了跟踪系统。这已被谷歌 Chrome 浏览器阻止。其他网站检测和阻止安装有广告拦截器的浏览器的个人访问其网址。见 Daniel Howe and Helen Nissenbaum, “Engineering privacy and protest: a case study of AdNauseam”。可查阅：https://adnauseam.io/。

[78] Tor（匿名网络）路由器隐藏谁与谁进行联系(例如电信元数据)，但并未得到广泛使用。一些浏览器(如 Firefox 和 Brave)包含“无痕浏览”模式，从而阻止数据收集。电子前沿基金会的 Privacy Badger 和纽约大学的 TrackMeNot 都非常有效，但并未得到广泛采用。

[79] Office of the Australian Information Commissioner, Australian community attitudes to privacy survey, 2017 and 2013; Deloitte, “Trust starts from within: Deloitte Australian privacy index 2017”, 2017。

[80] Ian Goodfellow, Jonathon Shlens and Christian Szegedy, “Explaining and harnessing adversarial examples”, ArXiv preprint, 2014。

114. 一种表述提出了数据共享的七个原则：[81]

1. 将算法移至数据：分享结果，而不是直接共享数据。

2. 开放算法：开放检查和公开审查数据共享和隐私保护的所有算法，以便确认和纠正错误或弱点。

3. 允许使用：尊重(明确或暗示)允许使用数据或“上下文完整性”。[82] 在医疗背景下，明确赋予撤消同意的权力已在动态同意界面中付诸实践。[83]

4. 始终发回“安全答复”：实践中的差分隐私。

5. 数据始终处于加密状态：仅有那些知道解密密钥的人可读取加密数据。[84]

6. 网络化协作环境以及审计和问责的区块链。

7. 社会和经济激励。

115. 这些原则本身并不一定是完整的解决方案，因为它们反过来又导致了更多的问题。例如，当用于保护隐私的技术非常复杂，仅有少数人有能力理解时，透明度又特别具有挑战性。“开放算法”原则是至关重要的第一步，但正在使用的确切算法及其影响在实践中仍具有挑战性。

116. 已经提出了其他“原则”办法，如“代理”和“透明度”，“代理”包含修改、模糊以及尝试提炼数据等权利。[85] 潜在的动力是赋予个人权力，以及在数据公司/持有者及用户之间引入权力平衡。其他人提出了模糊、防止或决定退出数据收集机会的原则。

117. 总的来说，透明度和用户控制的原则非常重要，这样用户才能选择他们公布何种数据才不会对设施或服务造成不合理的损失。

---

[81] Alex Pentland and others, “Towards an Internet of trusted data: a new framework for identity and data sharing”, 2016。

[82] 隐私的定义是“关于人的信息(‘个人信息’)适当流动的要求，这里适当指的是根据信息规范……社会环境构成了这种隐私做法的背景……”。见 Solon Barocas and Helen Nissenbaum, “Big data's end run around anonymity and consent”, in *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Julia Lane and others, eds. (Cambridge University Press, 2014)。

[83] Jane Kaye and others, “Dynamic consent: a patient interface for twenty-first century research networks”, *European Journal of Human Genetics*, vol. 23, No. 2 (2014)。

[84] 近期在密码学上取得的进展允许多方一起计算他们私下输入的一个功能，然后仅展示明确的成果。有非常普遍的工具，基于多方计算(如，见 Ivan Damgård and others, “Multiparty computation from somewhat homomorphic encryption”, *Advances in Cryptology — CRYPTO*, vol. 7417 (2012); and homomorphic encryption，可查阅：www.microsoft.com/en-us/research/project/homomorphic-encryption/#) 。对于大型数据集，大多数运算并不够快，但未来可能会出现更为简单的改进型。有许多专门的协定解决有关大型数据集的问题。有关加密数据计算的一般概念对于简单的一个数据集的计算效果很好，但对于复杂计算或分散在若干地方的数据集则行不通。

[85] Andreas Weigend, Data for the People: How to Make our Post-Privacy Economy Work for You (New York, Basic Books, 2017)。

118. 最重要的是，尝试制定尊重隐私的大数据-开放数据原则，为讨论提供了有益的起点。无论采用什么原则，包括民间社会组织在内的所有利益攸关方应进行充分磋商，从而确保任何此类原则的适当性。

119. 因实施这些原则而导致的问题包括政府的作用以及激励和监管的类型，这将促进保护隐私和其他人权以及评估“它们对道德和政治价值观的相对影响，如公正、正义、自由、自治、福利和讨论背景下其他更为具体的情况”。[86]

120. 如果能看到各国政府和公司遵守获取、分享和控制人们数据的强制性规定，那么创新型信息经济可能会获得更大的社会支持。

三. 支助文件

121. 支持本报告的以下文件可在特别报告员网站上查阅[87]：

(一) 了解历史：去识别化工具和争议；

(二) 特别报告员在非洲、美洲、亚洲和欧洲的接触；

(三) 致日本国政府公开信的背景；

(四) 隐私与人格工作队的活动；

(五) 关于监控的法律文书草案进程的说明；

(六) 承认协助；

(七) 关于大数据和开放数据专题报告的程序说明。

四. 结论

122. 本报告中确定的问题并不仅限于少数几个国家。大量的新数据收集的可获得性使得全球的个人、公司和国家能够更多和更好地做出合理的决策，但隐私管理不善使它们的潜在价值面临风险。

123. 需要审慎理解和成功减缓隐私、其他相关人权以及自治和公平的道德政治价值观所面临的风险。

124. 数据是并将仍然是一项重要的经济资产，与资本和劳动力一样。隐私和创新可以做到相互协调。了解如何有效利用大数据以及如何在不削弱人权保护的情况下分享其好处会很困难，但最终是值得的。

---

[86] Solon Barocas and Helen Nissenbaum, “Big data's end run around anonymity and consent”, in *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Julia Lane and others, eds. (Cambridge University Press, 2014)。

[87] 见 http://www.ohchr.org/EN/Issues/Privacy/SR/Pages/SRPrivacyIndex.aspx; see also www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx。

五. 建议

125. 在等待直到 2018 年 3 月磋商期间的反馈和当前调查及对政府指控函的结果之际，特别报告员正在考虑将以下建议纳入将在 2018 年或之后发布的本报告更新版本：

126. 开放数据政策需要明确阐述基于国际标准和原则对利用个人信息的限制，包括具有约束性要求的个人信息豁免类别，以确保去识别化过程的可靠性，从而使这类信息适宜作为开放数据予以发布，而且还需要强有力的执行机制。

127. 涉及个人信息的任何开放政府举措，不管是否去识别化，都需要对隐私保护进行严格、公开和科学的分析，这包括隐私影响评估。

128. 除非有合理证据确认已进行去识别化并在未来完全不会面临再次识别，否则有关个人的敏感高维单元记录级别数据不应在线发布或进行交换。

129. 应建立框架，管理向研究者提供的敏感数据面临的风险。

130. 政府和企业应积极支持创造和使用隐私增强技术。

131. 处理大数据时要考虑以下选项：

治理

(a) 责任：确认应负责任、决策过程并酌情确认决策者；

(b) 透明度：个人数据在公开发布前发生了什么、何时发生以及如何发生，以及个人数据的使用，包括“开放算法”；

(c) 质量：数据和处理质量的最低保证；

(d) 可预测性：当涉及机器学习时，结果应是可预测的；

(e) 安全：采取适当措施防止数据输入和算法受到未经授权的干扰；

(f) 开发新工具来识别风险并具体说明风险缓减；

(g) 支持：向雇员提供与个人信息有关的法律、政策和行政要求培训(以及适当的认证)；

监管环境

(h). 确保为负责保护公民数据的监管者确立明确的重点、责任和权力；

(i) 监管权力应与大数据带来的新挑战相称，例如监管者能够仔细检查分析过程及其成果的能力；

(j) 审查隐私法，确保这些法律在应对技术进步(如机器生成个人信息)和数据分析(如去识别化)带来的挑战方面“适用”；

纳入反馈机制

(k) 正式成立协商机制，包括有专业人员、社区和其他组织及公民参与的道德委员会，以保护权利免遭削弱和确认合理的做法；

(l) 就本报告提出的建议和问题进行基础广泛的磋商，例如对禁止提供政府数据集的兴趣；

研究

(m) 技术：调查相对较新的技术，如差分隐私和同态加密，以评估其是否提供充分的隐私处理和输出；

(n) 调查公民对政府和企业数据活动、个人信息使用(包括用于研究和技术机制)的认识，以加强个人控制其数据和增强他们利用这些数据以满足其所需的能力。