



## 统计委员会

### 第四十五届会议

2014年3月4至7日

临时议程\* 项目3(j)

供讨论和作出决定的项目：海量数据和统计系统现代化

## 海量数据和统计系统现代化

### 秘书长的报告

#### 摘要

本报告是根据经济及社会理事会第 2013/235 号决定编写的，其中提供有关目前官方统计使用海量数据的评估。报告将概述官方统计界最近的活动，并介绍统计司 2013 年开展的官方统计使用海量数据的全球评估结果。

报告将通过以下办法制订下一步的行动：基于共享方法进展、发展战略问题方面的最佳做法和培训机会的现有区域举措，在全球层面上建立一个工作组。该工作组还将促进建立国际伙伴关系，以推动在向发展中国家转让技术方面使用海量数据和支持 2015 年后发展议程。请统计委员会就报告最后一节中的提议提出意见。

\* E/CN.3/2014/1.



## 一. 引言

1. 关于 2013 年新出现问题的星期五研讨会的主题是“政策、发展和官方统计方面的海量数据”。<sup>1</sup> 私营部门的发言者与首席统计师一道就决策海量数据的相关性以及国家统计系统在利用新信息来源方面应确保发挥的作用发表意见。下午会议由澳大利亚统计师主持。这次活动得出的结论是，海量数据是官方统计人员不容忽视的信息来源，官方统计人员必须组织起来，紧急采取行动，以有效方式利用各种可能性和驾驭各种挑战。

2. 作为这次会外活动的立即后续行动，澳大利亚代表在统计委员会第四十四届会议上要求对官方统计使用海量数据的情况进行评估，并提交给委员会第四十五届会议。本报告是对这一要求作出的回应。

3. 因此，电子装置使用普遍，数字信息全面生成和提供，这从根本上改变了现在不断大量生成、我们称之为海量数据的数据性质，其具有的特性使其有别于传统来源数据。这些创新来源提供的数据发布范围广，结构不严密，数量巨大，且往往是实时提供。海量数据的数据来源可表述如下：“数据的数量之大，速度之快，种类之多，要求以符合成本效益的创新形式进行处理，以提高见解和加强决策”。

4. 在各国的家庭和企业调查答复率不断下降的时代，海量数据可以在定价、就业、经济产出和发展以及人口方面为决策者提供实时证据。<sup>2</sup> 海量数据可以提供的统计数据要比传统的官方统计数据来源如调查和行政数据来源更相关和及时。海量数据的多数来源来自私营部门，多数国家尚未颁布旨在允许海量数据用于官方统计的立法。因此，需要进一步开展有广泛基础的对话，不仅要考虑到法律、保密和观念的问题，而且还要考虑到有关及时、可靠和可信地利用海量数据的问题。通过将海量数据来源纳入编制官方统计数据，各国家、区域和国际统计组织能够在及时性、成本效益和减轻资源负担方面得到改善的情况下更好地获得有关经济、社会 and 环境的官方统计数据。

5. 鉴于技术创新以及实时信息的数量和多样性以爆炸式出现，联合国秘书长 2009 年启动了全球脉动倡议，利用数字数据、数据快速收集和分析方面的创新，帮助决策者能够实时了解危机对弱势群体的影响。全球脉动编写的白皮书虽然强调海量数据并非灵丹妙药，但其中指出，“海量数据是提高我们共同能力的历史性机会，以便通过了解其以电子形式提供越来越多的信息，支持和保护人类社会”。<sup>3</sup>

<sup>1</sup> 研讨会是 2013 年 2 月 22 日星期五举行的统计委员会第四十四届会议的一次会外活动，详见以下网站：[http://unstats.un.org/unsd/statcom/statcom\\_2013/seminars/Big\\_Data/default.html](http://unstats.un.org/unsd/statcom/statcom_2013/seminars/Big_Data/default.html)。

<sup>2</sup> 经济合作与发展组织(经合组织)，“探讨以数据驱动的创新作为新的增长来源：界定‘海量数据’提出的政策问题”，(DSTI/ICCP(2012)9/FINAL)，2013 年。

<sup>3</sup> 全球脉动，“海量数据促进发展：挑战与机遇”，2012 年，5 月。

6. 2015年后发展议程知名人士高级别小组在题为“新型全球伙伴关系：通过可持续发展消除贫困和促进经济转型”的报告中向秘书长提出建议，呼吁进行数据革命。海量数据被认为是数据革命的重要部分，可能有助于提高统计数据质量的某些方面，如及时性和全面性，同时不影响统计数据的相关性、公正性和方法合理性。高级别小组的报告指出：

更好的数据和统计将有助于各国政府跟踪进展情况，并确保进行循证决策；还可加强问责制。这不仅事关政府。国际机构、民间社会组织和私营部门也应参与其中。一次真正的数据革命将利用现有和新的数据来源，以便在决策中充分纳入统计数据，促进统计数据开放，并确保给予统计系统更多的支持（第4章，第24页）。

7. 更广泛地说，海量数据已表明有可能改变社会的许多方面。TechAmerica基金会的海量数据委员会在最近一份报告中指出，“海量数据有可能改造政府和社会本身。<sup>4</sup> 当今编制的海量数据数量巨大、种类繁多，速度极快，其中隐藏着新的信息、事实、关系、指标和指示，这一切过去几乎无法发现，或根本不曾存在。<sup>5</sup> 这一新信息经过有效地收集、管理和分析，完全有能力大大提高政府的效率”。<sup>6</sup>

8. 本报告第二节将概述有关海量数据和统计系统现代化的活动。第三节将简要介绍海量数据的来源和带来的挑战。第四节将提供统计司2013年7月至10月进行的有关官方统计使用海量数据的全球评估结果。第五节将提供一些有关海量数据使用的个案研究。报告最后提出关于今后行动方向的建议，这符合伙伴组织的倡议，并强调有关发展中国家官方统计使用海量数据以及2015年后发展议程的问题。

## 二. 最近海量数据相关活动概览

9. 越来越多国家的政府承认海量数据的重要性，并建立了同业交流圈和工作组，以研究其使用和可能产生的影响。统计界正逐步认识到有可能出现根本性的转变。意大利国家统计研究所确认，“为利用海量数据山透出的(可能巨大的)财富，国家统计研究将必须攀登这座山”。<sup>7</sup> 荷兰统计局承认，谈到探索海量数据提供的机会，官方统计界只是抓到表皮。<sup>8</sup>

<sup>4</sup> TechAmerica 基金会的海量数据委员会题为“揭秘海量数据：转变政府业务的实用指南”的报告执行摘要的开启句(华盛顿特区，2012年10月)。

<sup>5</sup> 同上，第二句。

<sup>6</sup> 同上，第三句。

<sup>7</sup> Monica Scannapieco、Antonino Virgillito 和 Diego Zardetto，“将海量数据置于官方统计数字中：一项巨大挑战？”，2012年，12月21日。

<sup>8</sup> Piet Daas 和 Mark van der Loo，“海量数据(和官方统计)”，为统计信息系统管理会议(MSIS 2013)编写的工作文件，巴黎和曼谷，2013年4月23-25日。

10. 欧洲经济委员会(欧洲经委会)、欧洲委员会统计局(欧统局)、经济合作与发展组织(经合发组织)以及亚洲及太平洋经济社会委员会(亚太经社会)举办了一次关于统计信息系统管理的会议,分别于2013年4月23至25日在巴黎和曼谷同时举行。<sup>9</sup> 这次会议的重要结论是,统计组织必须共同努力面对有关使用海量数据的共同问题,并努力寻找共同解决办法,如编制时效性短的统计数据、对海量数据采取多学科办法以及就不同类型的海量数据商定共同分类。

11. 2013年8月25至30日在中国香港行政特区举行的国际统计学会第五十九届世界统计大会非常关注海量数据问题,期间举行一次主题为海量数据的特约文件会议、一次关于因特网的潜力、海量数据及官方统计方面的有机数据的特别专题会议以及一场有关海量数据和统计计算的配套活动。在上述特别专题会议上,爱尔兰中央统计局的John Dunne介绍一份关于“海量数据很快进入……你附近的国家统计局”,<sup>10</sup> 其中介绍了一些海量数据来源,如移动电话漫游、电力消费或电子付款,并介绍可通过第三方提供商的云计算处理这些大量及时数据的方式。海量数据还将成为即将于2015年7月26至31日在巴西里约热内卢举行的第六十届世界统计大会科学方案的主要议题。<sup>11</sup> 国际统计学会大会为官方、学术和商业统计人员进行合作和交流知识提供机会。

12. 欧统局正在调查官方统计在价格统计(采用因特网价格数据)以及信息和通信技术(信通技术)使用统计等领域可能使用海量数据的情况。此外,欧洲国家统计机构总干事2013年9月在海牙举行年度会议,期间专门举行一次会议讨论海量数据的主题。<sup>12</sup> 这促使签署《关于海量数据和官方统计的斯海弗宁恩备忘录》(欧洲统计系统委员会于2013年9月27日通过),其中鼓励欧洲统计系统的成员制订一项海量数据战略,分享经验,并在欧洲统计系统内外举行协作。应在2014年年中之前通过一项行动计划和路线图,并将其纳入欧统局的工作方案中。

13. 海量数据也是欧洲经委会2013年9月25至27日在日内瓦举行的关于统计数据收集问题研讨会讨论内容的一部分。<sup>13</sup> 在研讨会上,荷兰统计局的Peter Struijs和Piet Daas介绍了一份题为“海量数据,重大影响?”的工作文件(第31号),其中指出,很明显,海量数据将对统计界和官方统计产生重大影响。这一影响的具体情况只会逐渐变得清楚,但其中一些特点已显现或可预见。国家统计系统将面对外界行为体更多的竞争。他们将需要调整开展统计工作的方式,并利用海量数据提供的新可能性找到新的平衡。这可能需要进行根本性的转变,即从

<sup>9</sup> 见: <http://www1.unece.org/stat/platform/display/msis/MSIS+2013>。

<sup>10</sup> 可从以下网站查阅: <http://www.statistics.gov.hk/wsc/STS018-P3-S.pdf>。

<sup>11</sup> 见: <http://www.isi2015.ibge.gov.br>。

<sup>12</sup> 见: <http://www.cbs-events.nl/dgins2013/>。

<sup>13</sup> 见: <http://www.unece.org/stats/documents/2013.09.coll.html>。

注重调查转向更注重二级数据，其中基于模型的方法将成为常态。随着进入海量数据时代，统计机构必须有数据科学家加入其队伍，使其成为一种常态而非例外，并相应改变组织文化。由于一早就认识到海量数据带来的机会和挑战，官方统计界将能够大大加强其在提供信息方面的作用，超越这些新来源。

14. 在2013年10月21和22日举行的第二次会议上，欧洲统计员会议的2013/2014年度主席团、欧洲经委会的统计理事机构深入审查了海量数据的主题。<sup>14</sup> 审查提出的主要建议是，国际统计界应具体列明海量数据的主要优先领域，并开展协作活动加以应对，还应建立一个机制，分享有关使用海量数据的知识和经验方面的信息。主席团还批准了一个有关海量数据作用的项目提案，<sup>15</sup> 其目标如下：

(a) 确定海量数据提供的各种主要可能性和为统计组织提供指导，并对海量数据给官方统计界带来的主要战略和方法问题制订一项协调一致的对策；

(b) 展示有效编制新产品和利用海量数据来源编制的‘主流’官方统计的可行性以及在各国不同背景下推广这些做法的可能性；

(c) 促进各组织分享有关利用海量数据来源编制统计数据的知识、专门知识、工具和方法。

15. 海量数据和统计系统现代化也列入亚洲最近一些活动的议程。统计司和中国国家统计局在亚太经社会的支持下于2013年10月24至26日在中国天津举办了一次有关官方统计现代化的国际研讨会。这次研讨会着重讨论统计组织现代化问题，并将一些有关海量数据的介绍列入其中。亚太经社会的亚洲及太平洋统计研究所(亚太统计所)于2013年11月21日和22日在日本千叶举行亚洲及太平洋的各国家统计局负责人第11次管理研讨会，<sup>16</sup> 主要讨论数据革命倡议以及海量数据在2015年后发展议程的潜在作用。

### 三. 海量数据：来源与挑战

16. 统计编制和服务现代化问题高级别小组在欧洲经委会的协调下设立了一个非正式工作组。小组在一份题为“海量数据对于官方统计意味着什么”的报告中总结了有关海量数据使用的海量数据来源和挑战(2013年3月10日)。该报告已提交给2013年6月10日至12日举行的欧洲统计员会议第61届全体会议。有关该报告的定论将在下文进行讨论。统计司在2013年第三季度进行的全球调查问卷中采用了该报告介绍的欧洲经委会工作队提出的细目(见第四节)。

<sup>14</sup> 见 ECE/CES/BUR/2013/OCT/2。

<sup>15</sup> 这一项目是欧洲经委会协调的统计编制和服务现代化问题高级别小组下统计编制现代化倡议的一部分。

<sup>16</sup> 见：[http://www.unsiap.or.jp/training/6\\_Leading\\_MQS/6\\_ms11.php](http://www.unsiap.or.jp/training/6_Leading_MQS/6_ms11.php)。

17. 现在不断用多种不同方式以数字形式自动生成数据。官方统计可能有意采用这些不同数据来源，以便更准确和及时地衡量某些社会、环境、金融或经济趋势。总体而言，海量数据来源可分类如下：

- 无论政府与否的方案管理提供的来源，如电子病历、医院就诊、保险记录、银行记录和粮食银行
- 两个实体之间的交易提供的商业或交易来源，如信用卡交易以及联机交易(包括移动装置交易)
- 传感器网络来源，如卫星图象、道路传感器和气候传感器
- 追踪装置来源，如从移动电话和全球定位系统追踪数据
- 行为数据来源，如网上搜索(产品、服务或任何其他类型的信息)和网页浏览
- 意见数据来源，如社交媒体评论

18. 行政数据是国家统计系统编制官方统计的其中一个主要数据来源。传统上，这些数据是公共行政当局以一种非常有序的方式提供的，然后再由统计机构处理、储存、管理和使用。行政数据目前不属海量数据来源，但如果速度和数量增加，如统计研究所通过实时、每天或每周进行收集，而不是按通常的每年或每月进行一次收集，开始更广泛地使用行政数据，这些数据就可能成为“海量”。

19. 官方统计使用海量数据带来许多挑战，主要分为以下类别：

- 立法方面，即有关获得和使用数据
- 隐私，即管理公众对数据再用的信任和接受及其与其他来源的连接
- 金融方面，即采购数据的潜在成本与效益
- 管理方面，如有关数据管理和保护的政策和指示
- 方法论方面，即数据质量和统计方法的适当性
- 技术方面，即与信息技术有关的问题

#### 四. 海量数据使用情况全球调查结果

20. 根据欧洲经委会任务小组报告提供的说明，统计司拟定了一份关于官方统计使用海量数据情况的调查问卷。问卷有三个主要部分：来源、挑战和使用领域。有关来源与挑战的问题包含上述各点，其中对管理方面作了一些详细说明，具体措辞是，“管理问题，例如遵守新政策和条例，以及开发具备必要技能和专门知识的人力资源”。开展这一全球评估旨在了解官方统计在目前使用或计划使用海量数据方面的国家优先事项、发展状况及经验。

21. 2013年7月至9月，向200多个经济体的统计部门发出了调查问卷：7月份发出英文问卷，8月份是西班牙文，9月份是法文。可在线或以可移植文件格式回答问卷。截至2013年11月2日，共收到107份答复。作为背景文件，向统计委员会提供了全面调查结果。下文简要介绍调查结果。

#### A. 海量数据来源

22. 关于海量数据来源这一问题的确切措辞是：“请说明在今后12个月里，贵机构或贵国国家统计系统中的其他机构有可能使用下列哪种海量数据来源。”如果回答“是”，则需具体说明数据来源。50%以上的经济体对使用行政来源的海量数据给出了肯定答复，而对其他五种数据来源的肯定答复率各约为25%。一些国家就数量庞大的行政记录是否为海量数据提出了疑问。更广泛地说，行政数据是制作许多统计产品的基石，但是否应将其放在海量数据这一范畴内加以考虑，仍是一个有待讨论的问题。

23. 总体而言，许多国家对使用一种或多种海量数据来源给出了肯定答复。但进一步分析表明，其中一些肯定答复不符合欧洲经委会任务小组参考报告中使用的海量数据的定义。为了使此次关于海量数据使用情况的调查得出的结论更切合实际，对收到的答复进行了仔细审查，并将一些肯定答复缩减归为一个组，作为“相关做法”。未达标的肯定答复都是未作任何解释、解释非常有限或所做解释没有相关性。做了这种调整后得出的结果不再是最初的肯定答复率为25-50%，而是行政数据来源几乎为20%，交易来源、传感器网络和跟踪装置大约10%，行为和意见来源不到5%。将在统计司网站上提供详细结果，对给出肯定答复的所有国家和被视为“相关做法”的国家将分列显示。

24. 提及最多的行政数据来源是来自税务和增值税办公室的数据。其他来源包括个人登记和商业登记册。提到的交易来源是超市扫描数据、电话通信、信用卡交易及收费数据。关于传感器或卫星来源，一些国家指出，他们使用卫星图像进行调查设计。还提到以卫星图像为来源进行作物生产统计研究。道路传感器用于进行交通统计（荷兰用于测算交通指数）和环境统计（大不列颠及北爱尔兰联合王国用来衡量空气质量）。

25. 若干国家使用追踪来源或全球定位系统来源。如爱沙尼亚通过跟踪移动电话来改进旅游统计；下一节将以此为案例研究加以介绍。加拿大利用有关卡车、铁路、船只和空中运输的全球定位系统数据来补充交通统计。尽管许多国家并没有使用行为数据，但一些国家计划利用这一数据来源。例如，意大利可能使用因特网上提问(a)作为辅助信息来源，以减少与调查相关的费用；(b)进行预测；(c)增加估计数的地理细节（特别是劳工市场指标）。

26. 同样，很少有国家利用意见或社交媒体作为海量数据来源。荷兰指出，其根据公开获得的社会媒体信息开展了一些研究工作（见下一节的案例研究）。新加坡目前正在研究如何利用这一海量数据来源尽早发现食品安全问题。

## B. 海量数据使用方面存在的挑战

27. 这部分的确切问题是，“[问题]对贵国国家统计局而言是否是一项重大挑战？”可能的答案选项为：否(不是挑战)、不确定(没有讨论过)、是(是一项挑战)。对所有六项挑战，大多数国家选择“是”(是一项挑战)，其次是“不确定”(意指尚未充分讨论过这些问题)，只有少数国家回答“否”(不是一项挑战)。总体而言，提及最多的是方法、信息技术和管理方面的挑战，其次是立法和隐私方面的挑战。

28. 答卷人都认为，主要挑战之一是解决方法问题，认为在这一领域需要开展更多研究。数据的质量、代表性、波动性和多面性是利用海量数据时遇到的方法难题。许多国家指出，信息技术问题也是一项重大挑战，因为现有数据处理和分析系统的规模不足以应对海量数据。海量数据的性质(特别是数量之大及速度之快、种类之多)给开发必要的信息技术工具和信息技术架构带来了重大挑战，对发展中国家尤其如此。在大多数国家，遵守新政策和条例及发展具备必要技能和专门知识的人力资源等管理问题也是一项重大挑战。

29. 对政府而言，有系统地使用非政府来源的数据是利用海量数据方面存在的一个立法挑战。一些国家已经立法，确保获得政府和非政府来源的数据；其他国家则指出，没有法律保障获得非政府来源的数据是一项挑战。隐私问题也是许多国家在使用海量数据方面的一个主要关切问题。保护个人资料在这方面非常重要。因此，有必要在保护个人资料和统计使用之间找到最佳平衡。

30. 一些国家认为，财政问题也很棘手，特别是采购数据所涉及的费用被认为是使用海量数据所面临的一项重大挑战。可能没有足够多的资金来购置处理海量数据所必备的最新硬件和软件。此外，私营公司似乎日益认识到它们所掌握的数据的价值，因此要价越来越高。

31. 还提到了国家统计局认为与海量数据使用有关的其他主要挑战。特别是，许多答卷人提到如何将海量数据纳入统计系统的问题。如何使这一新的数据来源适应现有的统计方案？大多数答卷人认为，海量数据不会取代官方统计，而是作为一种补充，例如迅速提供统计数据 and 关键统计的主要指标。由于海量数据是官方统计领域中一个较新的概念，有必要培养官方统计人员对海量数据的认同和接受。

## C. 现有使用领域及潜在使用领域

32. 全球调查第三和最后一部分的问题涉及未来12个月海量数据的使用(或研究使用)领域。所提领域包括“人口和社会统计”、“生命统计和民事登记统计”、“经济和金融统计”、“价格统计”、“运输统计”、“环境统计”和“其他官方统计领域”。答案选项为“是”或“否”。如选“是”，则需加以解释。肯定答复最多的使用领域是“人口和社会统计”及“经济和金融统计”。不过，如同对是否使用海量数

据这一问题的回答一样，需要对调查问卷这一部分的肯定答复进行重新评价，确定“相关做法”。做了重新评价后发现，使用海量数据的相关做法是在价格统计及经济和金融统计领域。特别是，一些国家提到使用扫描数据和(或)网络浏览技术创建时常更新的价格指数，以补充标准消费价格指数。总体而言，使用领域的相关做法所占比例对价格统计、经济和金融统计及人口和社会统计均为略高于10%，对其他使用领域均为大约5%。

33. 在人口和社会统计领域中，有几个答卷人提到移动电话数据可用于分析旅行和旅游行为，社交媒体情绪用于分析消费者信心，在线职位空缺数据用于就业分析，并提到谷歌的潜在使用趋势。其他例子包括与移动电话运营商合作，计算有地理参照的通勤统计数据。在经济和金融统计领域，提到研究使用卫星图像和地面遥感数据来估计作物产量，并使用移动装置位置跟踪数据来计算服务人口流动统计。若干国家解释了它们在运输统计领域如何使用海量数据。也使用来自公路收费站和道路传感器的数据编制交通统计。

## 五. 海量数据与统计系统现代化

34. 为应对全球调查答复中提到的各种挑战，统计系统需要实行现代化。有人建议应开展更多研究，以克服阻碍使用海量数据来源的各种方法难题。考虑到海量数据的性质，数据代表性和人口覆盖范围可能是棘手问题。此外，从官方统计的角度看，数据种类繁多及数据的潜在时间性(比如，人们使用“脸书”的时间会持续多久?)会带来问题，因为数据可比性、连续性和一致性是数据质量的主要方面。方法的改变可以包括更经常利用建模，可能需要与学术界开展更密切的合作。需要开展更多的研究和实验研究，探讨海量数据在官方统计中的潜在用途，而且应将研究作为统计制作过程的一部分。

35. 必须加强现有的信息技术基础设施和能力，以支持处理来自各种来源的大量和高速数据集。此外，应考虑对国家统计局使用的方法和工具进行标准化。必须获得最新技术(软件、硬件和人力)，以满足收集、处理和使用海量数据的技术需求。国家统计局可能需要考虑云计算，而不是试图把大量数据迁移到自己的服务器上。另外，可要求海量数据来源的监护人进行初始计算，向统计部门提供更集中、组织结构更好的数据。需要有足够的资源来建立必要的信息技术基础设施，以便更频繁地收集相关数据，链接和分析数据，并以适当格式向用户提交数据。

36. 发展和保留掌握先进的数据分析和计算信息学方面必要技能的工作人员将是一个重大挑战。为利用新的海量数据来源，需要改变数据收集和处理的组织工作。大多数国家统计局没有具备适当技能的工作人员，因此，必须在人力资源方面开展能力建设和密集培训，以便能够利用这些数据来源。以传统的统计知识为基础的专门知识不足以确保充分处理这种新类型的数据。因此，有必要提供资源，以购置和培训适当的新型专业人员(如数据科学家)。

37. 需要进一步采取法律步骤，以便使用海量数据编制官方统计。许多国家目前的统计立法框架并不涉及在政府内部和从私营部门获取和使用海量数据。因此，很难获得由其他各方收集和保存的海量数据。此外，需要建立隐私框架，确定有关集合、保护、共享、披露、分析和保存海量数据集的基本规则。这将解决公众对政府适当使用个人资料的信任问题。保持公众的信任至关重要：必须让人们放心，有关他们的个人资料将得到良好保护，不会被披露或滥用。例如，在移动电话位置数据领域即是如此，这些数据是海量数据的一个潜在来源。但即使隐去了个人识别信息，人们仍非常担心移动通信运营商会将此类信息转移给其他方用于各种目的，也担心这些信息会被多方共享。同样，移动设备提供者需要保证它们向政府提供数据时，不会侵犯隐私权。

38. 公众对官方统计使用海量数据作何反应是海量数据未来使用情况的一个重要决定因素。政府、更重要的是私营部门必须赢得(而不是破坏)公众对个人数据使用的信任和信心。政府部门需要做出极大努力，建立一个授权框架，其中规定有关获取和使用自动生成的包含个人内容的数字资料的规则，明确规定数据监护人及相关政府机构的作用。只有在这种环境下，才能确保个人信息的隐私和保密，才能建立公众对官方统计使用海量数据的信任。

## 六. 案例研究

### A. 澳大利亚：卫星图像用于农业统计

39. 卫星图像正在应用于土地覆盖领域和一项利用农业土地使用特征估计作物特定类型的研究方案中。将利用空间生成的特征输入建立的统计模型中，以便在农业统计汇编中进行计算。这些特征也可能应用于渔业、林业以及环境统计的汇编工作。原始数据由统计组织的计算环境以外的供应商系统云服务器查阅和处理。提取空间土地使用特征的图像分析算法是与大学合作开展的联合研究项目开发的。这涉及算法、空间和时间序列模型、地理编码系统和统计估算程序的开发。

### B. 巴西：高频监测数据用于水资源账户

40. 根据国家地理和统计研究所(统计机构)、国家供水署(水资源联邦机构)以及环境部水资源和城市环境秘书处(负责水资源政策协调)2012年签署的一项协定，设立了一个委员会，其任务是制订巴西的水资源账户。国家供水署通过由近22 000个监测站组成的国家水文气象学网络监测水资源的数量和质量，这些监测站收集有关水流、降雨、水质、蒸发和沉积物的高频数据。换言之，国家供水署每日收集有关水的大量数据，处理这些数据，发布在该署的机构网站<sup>17</sup>或国家水资源信息系统(信息系统)<sup>18</sup>上供免费查阅。信息系统是一个收集、处理、储存和

<sup>17</sup> <http://www2.ana.gov.br>。

<sup>18</sup> 见 <http://www.snirh.gov.br/telemetria/consultardadoshidrovisualizacaogeo/inicial.action>。

提供有关水资源管理的数据和信息的全面系统。国家地理和统计研究所将信息系统提供的数据与该所定期收集的经济和环境统计数据合并，汇编巴西水资源账户。

### C. 不丹：移动通信技术用于消费价格指数

41. 在不丹，消费价格数据的收集工作得到现代通信技术的支持和协助。消费价格指数每月根据当地统计人员使用安卓手机应用程序从各地区收集的数据进行编制。统计人员从各所在地区的相关商店收集消费价格指数篮子所有物品的数据，并自动向廷布国家统计局总部的服务器提交数据。总部合并各地区的数据，编制月份全国消费物价指数和通货膨胀数字。

42. 尽管不丹目前的应用并不涉及巨量数据，移动设备和无线传输的利用可得到进一步发展，并应用于需要更大量数据的价格收集工作。不丹采用的从使用小规模电子数据收集起步的方法，也可被视为其他发展中国家一种可借鉴的途径。数字数据输入自动化和向中央服务器无线传输数据是挖掘海量数据源潜力方面迈出的重要第一步。

### D. 哥伦比亚：卫星图像用于农业和运输统计

43. 哥伦比亚一直在若干领域中利用卫星图像作为海量数据源。利用卫星图像为全国农业普查的调查工作人员提供补充的试验是这方面一个实例。该试验的诊断阶段旨在确定通过卫星成像可能捕获并能支持普查行动所收集的信息的变量类型。另一个例子是，联合国毒品和犯罪问题办公室以及内政和司法部通过非法作物综合监测系统，使用卫星图像衡量和监测哥伦比亚的古柯作物。

44. 哥伦比亚国家公路局使用全球定位系统数据改善交通流动，并将其作为运输统计的输入。利用这一方法，汽车不必在收费站内停车，在进入收费站时车辆上安装电子追踪装置已被读取。这种动态的收费确保高水平的舒适、安全和服务。追踪装置还记录有关车辆的所有资料，可对国家单一过境登记的资料提供补充。到目前为止，这一新方法已在哥伦比亚 10 个收费站进行了测试，加强了对交通流量的控制，并促使运输统计得到加强。

### E. 爱沙尼亚：移动定位数据用于国际旅行统计

45. 使用被动移动定位汇编国际旅行统计数据的方法是由爱沙尼亚中央银行、塔尔图大学和 Positium LBS 共同开发的。Positium 是一家公司，使用统计模型收集和处理移动运营商提供的无记名数据，并将这些数据扩大为一般数据集。这一方法对经过爱沙尼亚边界前往国外（出境旅行）或进入爱沙尼亚（入境旅行）的自然人提供了一个可靠的概况。移动定位数据是无记名的。数据处理、分析和列报所采用的方法不允许以任何形式确定电话号码或电话所有人。

46. 中央银行利用国际旅行统计数据计算国际收支经常帐户旅行服务的进出口。根据国际收支编制方法，移动定位根据长期居住的标准确定乘客的居住地，而无

论该居民的公民身份。推定的电话所有人居住地通过电话漫游和当地移动网络的使用情况确定。爱沙尼亚居民和非居民访问的次数、时间和性质根据基于地点的手机使用方式确定，因使用的国家和地区而不同。

47. 出境旅行统计数据反映居民到外国当日访问和过夜访问。爱沙尼亚居民出境旅行可能包括访问若干国家。目的地国指距爱沙尼亚最远的国家或至少连续逗留两天的国家(若干国家)。到访但未过夜或不符合目的地国标准的其他国家列为过境国。居民和非居民访问的次数和时间通过统计模型进行计算,并将其扩大为一般数据集,同时也考虑到其他官方数据来源(塔林港和塔林机场、爱沙尼亚东部边界各过境点和正式住宿统计的乘客人数)。

48. 无记名移动定位数据的使用遵循爱沙尼亚和欧洲联盟委员会关于数据保护和个人数据保护的立法(爱沙尼亚个人数据保护法案;欧洲议会和欧洲理事会1995年10月24日关于在处理个人数据方面保护个人及个人数据自由流通的第95/46/EC号指令;欧洲议会和欧洲理事会2002年7月12日关于电子通讯领域处理个人数据和保护隐私的第2002/58/EC号指令)。自2006年以来,爱沙尼亚与数据保护监察局在个人数据保护问题方面开展合作,已在各个领域开发和使用移动定位方法。

#### F. 荷兰: 社交媒体作为潜在的官方统计数据来源

49. 越来越多的人活跃在社交媒体上。在这里人们自愿交流信息,讨论关心的话题,联系家人和朋友。荷兰大量人口使用社交媒体;约占人口70%的人报告说,他们活跃在一个或多个社交媒体平台上。由于对荷兰统计局调查问卷的答复继续减少,将社交媒体上自愿生成和交流的信息用作官方统计数据来源的可能性正在得到探讨。为此,对推特和脸书等各种社交媒体平台上生成的对外公开的社交媒体信息以及新闻网站、网上论坛和博客上张贴的公开信息进行了研究。这些信息由一家商业公司提供,该公司经常大量收集荷兰语网站上对外公开的荷兰语信息。

50. 对信息的内容和看法进行了研究。对生成最大量公开对外的荷兰语信息的社交媒体平台荷兰语推特信息的内容的研究表明,这些信息有接近50%由“无意义的唠叨”所组成。其余信息主要讨论业余活动(10%)、工作(7%)、媒体(电视和广播)(5%)和政治(3%)。比较不严肃的“唠叨”信息阻碍了这些比较严肃的信息的使用。确定所有可使用平台上生成的所有信息中的看法表明,这些数据用于统计的可能性相当令人关注。已发现这些信息中的看法与荷兰消费者信心高度相关,尤其是与有关经济形势的看法相关。后一种联系按月和按星期为单位衡量表现稳定。但按日衡量的数字表现出高度的波动性,表明有可能按月和按星期编制与消费者信心可比拟的看法指标。消费者信心指标可在进行研究这一周后的第一个工作日编制,表现出迅速交付成果的能力。

## 七. 结论和下一步行动

51. 统计界已经认识到海量数据用于官方统计的潜力。海量数据有助于更好地完成决策、研究和公共辩论提供及时和协调一致的经济、社会和环境统计数据任务。此外，随着经济、社会和环境单位地理编码技术的进步和提供的最低水平地理分辨率信息日益增多，预计海量数据的使用将在发达国家和发展中国家早日取得进展。因此，如果要发挥海量数据的优势，方法的发展、战略问题上的最佳做法和培训机会都必须跨越各种区域倡议在全球统计界之内共享，包括处理有关立法、隐私、资金、管理、方法和技术的问题。在解决海量数据问题的同时，各国政府和企业部门将需要创造一种环境，保证用于官方统计目的的个人信息的隐私权和保密性，建立公众对官方统计使用海量数据的信任。

52. 官方统计使用海量数据的全球评估产生了一些很好的例子，尤其是在交易、跟踪、遥感和卫星数据来源方面。使用扫描仪信息汇编价格统计数据，用电子卡交易信息汇编消费支出，用全球定位系统追踪装置信息汇编旅行和运输统计数据，用社交媒体数据汇编福利统计数据，用卫星图像数据汇编农业统计数据。全球调查还揭示了许多明显的挑战，特别是海量数据汇编方法上的困难、信息技术要求和隐私关切方面。

53. 显然，面对海量数据的挑战意味着统计系统的现代化。例如，统计系统的研究机构将需要加强，信息技术能力将需要提高，人力资源结构将需要调整以增加数据科学家，需要与私营部门建立伙伴关系，在一个满足隐私关切并遵守保密规定的环境下分享自动生成的数字信息。

54. 海量数据和统计系统现代化带来的挑战和机遇对大多数国家来说非常相似。因此，国家统计机构可以平等地分享经验、做法和解决方案。这方面的一些倡议已在区域一级启动。在欧洲统计员会议的框架内，欧洲经委会一个发展海量数据在统计编制现代化中的作用的的项目提案得到批准。亚洲及太平洋区域正通过亚太经社会提出类似的倡议。此外，其他区域也在考虑启动反映了类似期望的项目。

55. 设想建立一个全球工作组，在全球工作方案内补充区域期望和倡议。这一方案将使各区域委员会之间可以分享做法。各区域委员会可在区域一级与数量较少的国家就具体项目开展合作，而全球工作组将交流和管理各种问题，以便就方法事项、战略问题最佳做法、培训和共同平台的使用达成协议。作为其工作方案的一部分，全球工作组应促使国家统计系统在验证用于政策用途的海量数据或将海量数据来源与现有的官方统计数据相结合方面比其他各方具有相对优势。

56. 全球工作组可以进一步解决有关海量数据促进 2015 年后发展议程的作用的具体问题，并协调与其他联合国组织和国际组织、民间社会组织和新兴数据革命倡议全球伙伴关系其他成员的伙伴关系。对千年发展目标指标的一项投诉是指标的及时性差。海量数据往往实时生成，可能提供非常及时的指标。

57. 为启动全球工作组，2014 年第二或第三季度将举办一次国际会议，其目的是确定一项关于海量数据用于官方统计的共同方法、培训、平台和其他相关问题的工作方案。

## 八. 讨论要点

58. 邀请统计委员会对以下问题表示意见：

(a) 需要进一步调查在全球一级将海量数据用于官方统计的来源、挑战和领域，特别是发展中国家的情况以及与 2015 年后发展议程和数据革命倡议的联系；

(b) 建立一个海量数据用于官方统计问题全球工作组，其活动将补充各区域委员会所开展的工作，并管理全球相关问题。