



## 统计委员会

## 第四十七届会议

2016年3月8日至11日

临时议程\*\* 项目 3(c)

供讨论和决定的项目：官方统计

使用大数据问题

## 官方统计使用大数据问题全球工作组的报告

## 秘书长的说明

## 摘要

依照经济及社会理事会第 2015/216 号决定，秘书长谨转递官方统计使用大数据问题全球工作组的报告。全球工作组自 2014 年 10 月在北京举行第一次会议以来，交付了宣传材料、关于大数据的获取和伙伴关系问题的文件，以及关于大数据分类和方法及项目储存库的范围规划文件；进行了一次关于大数据的全球调查；开展了涉及卫星图像数据和移动电话数据的试点项目。短期的优先事项清单和工作方案中包括：关于大数据用于可持续发展目标各项指标的试点项目；利用“大数据沙箱”进行培训；启动大数据储存库；以及在获取私营部门所拥有数据的原则上取得进一步进展。报告还载列了 2015 年 10 月在阿布扎比召开的第二届官方统计使用大数据问题国际会议的成果。

邀请统计委员会就所取得的进展发表意见，并就有关数据获取和伙伴关系、能力建设和使用大数据制定可持续发展目标各项指标的计划发表评论。

供委员会讨论的要点载于本报告最后一节。

\* 由于技术原因于 2016 年 2 月 10 日重发。

\*\* E/CN.3/2016/1。



## 一. 引言

1. 2014 年,统计委员会第四十五届会议设立了官方统计使用大数据问题全球工作组。全球工作组按照其职权范围<sup>1</sup> 和第 46/101 号决定(第三节)<sup>2</sup> 提供有关官方统计(包括 2030 年可持续发展议程指标)使用大数据全球方案的战略愿景、指导和协调。工作组促进切实使用大数据来源,同时促进能力建设、培训和分享经验。最后,全球工作组在使用大数据方面促进政策应用的沟通和宣传,并就如何建立公众对使用私营部门大数据的信任提供咨询。

2. 全球工作组<sup>3</sup> 将其 2014 年和 2015 年工作方案分为 8 个流程,每个流程由一个专门的任务小组负责实施。全球工作组设立了分别负责移动电话数据、卫星图像和社交媒体数据的三个任务小组,通过试点项目草拟指导意见和制订做法。针对被视为重要问题的获取数据及与私营部门和其他社区建立伙伴关系事宜,还设立了另一个任务小组;该小组起草了与全球大数据提供者之间关于获取数据的临时总括协定。

3. 为了宣传大数据的好处和价值(其中包括使发展中国家能够积极参加试点项目的筹资战略),又成立了另外一个任务小组。按照《2030 年可持续发展议程》,全球工作组还商定,让一个任务小组专门负责跟踪可持续发展目标监测工作所需的指标和大数据应用之间的关系。最后,再设立了两个任务小组:一个针对需要不同于统计部门现有技能的技能而负责培训、技能和能力建设,一个负责处理方法、分类和质量框架等贯穿各领域的问题。上次报告(E/CN.3/2015/4)列举了全球工作组一些与 8 个任务小组工作有关的可交付成果。此外,工作组成员参加了一些关于官方统计使用大数据的试点项目,并继续致力于编制一个大数据库项目清单。

4. 本报告首先简要介绍了全球工作组的各任务小组所取得的进展(第二节),然后介绍了关于大数据战略和项目的全球调查的结果(第三节),随后介绍了第二届官方统计使用大数据问题国际会议的成果(第四节)。之后,本报告审查了全球工作组的优先事项和工作方案(第五节)。报告最后列出了一些讨论要点。

## 二. 全球工作组取得的进展

5. 如上文所述,全球工作组的目标是指导全球官方统计使用大数据方案,包括将其用于衡量《2030 年可持续发展议程》的各项指标、各种政策应用及促进有关

<sup>1</sup> 见 E/CN.3/2015/4。

<sup>2</sup> 见 E/CN.3/2015/34。

<sup>3</sup> 全球工作组最新成员清单见 <http://unstats.un.org/unsd/bigdata/>。

大数据问题的能力建设。该小组网站上的一份背景文件中全面概述了其 8 个任务小组所取得的进展。<sup>4</sup> 下文阐述了一些重点。

6. 全球工作组交付了关于宣传和沟通的产出，即一份战略文件、一个小册子、录像和一份通讯。为此编写了关于获取数据和伙伴关系的三份文件，<sup>5</sup> 包括关于获取大数据来源的原则草案，这些文件将与相关利益攸关方进行密切协商后进一步编写。两个任务组开展综合工作，编纂了一份大数据项目清单，其中还编制了可持续发展目标及其指标的项目。此外，就以下工作起草了范围规划文件：大数据来源的分类、建立大数据质量框架<sup>6</sup>（旨在汇集自下而上(范例)和自上而下(理论)的方法)及测试这一应根据具体的大数据来源和程序调整的框架。

7. 全球工作组还通过各任务组启动了两个利用移动电话数据的试点项目和其他一些使用社交媒体数据的试点项目。关于卫星图像、遥感技术和地理空间数据，工作组支持收集质量合格的典型培训数据的方法，用以制定和执行符合目的、包括用于预测模型的评估方法。作为理想目标的“统包”统计应用程序正在测试当中，用以预测作物的产量。

8. 全球工作组网站上的一份背景文件中载有关于 8 个任务组的完整的工作进度报告。

### 三. 全球大数据战略和项目调查结果

9. 统计委员会在其第 46/101 号决定第三节中批准了关于全球工作组就官方统计使用大数据问题进行全球调查的提议。调查的目的是评估统计机构迄今针对大数据采取的步骤的情况。它询问了各国统计局的战略愿景及其在大数据方面的实际经验。该调查问卷所载问题涉及大数据管理、宣传和沟通、将大数据与可持续发展目标挂钩、数据的获取、隐私和保密、技能和培训以及各统计局关于使用大数据的最紧迫需求。此外，该调查问卷载有关于大数据项目的详细问题，对象是那些参与了一个或多个项目的统计局。

10. 调查从 2015 年 6 月进行到 8 月。调查问卷发给每个国家统计局，请其与本国统计系统内的相关利益攸关方协商。共有 93 个国家填妥了问卷(32 个经济合作与发展组织(经合组织)成员国和 61 个非经合组织国家)。此外，还收到了欧盟统计局的答复。该问卷包括 18 个关于大数据管理的问题和 24 个关于每个已报告大数据项目的问题。大约一半的国家报告了至少一个大数据项目，少数几个国家报告了 4 个以上的项目。然而，应当指出，另一半国家没有报告任何项目。不过，

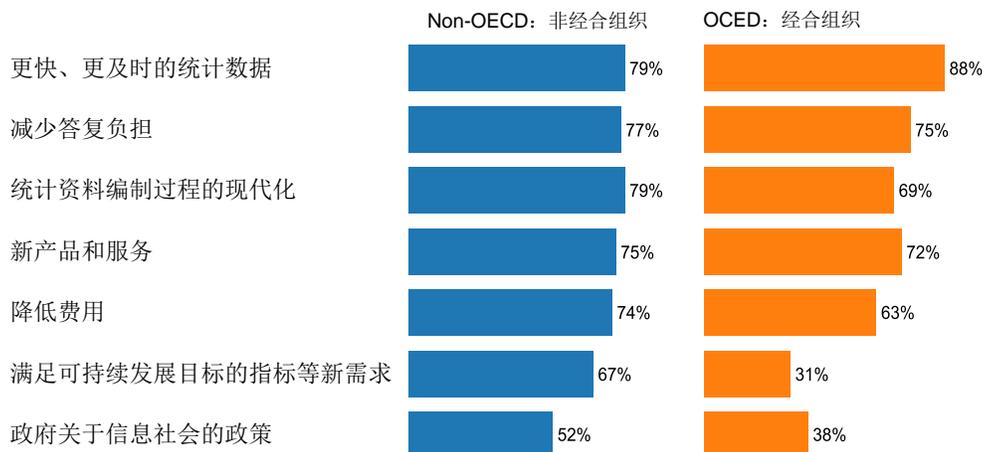
<sup>4</sup> <http://unstats.un.org/unsd/bigdata/>。

<sup>5</sup> 见 <http://unstats.un.org/unsd/trade/events/2015/abudhabi/gwg-agenda.asp>。

<sup>6</sup> 详情见 <http://www1.unece.org/stat/platform/display/bigdata/2014+Project>。

共报告了 115 个大数据项目：经合组织国家 89 个，非经合组织国家 22 个，欧盟统计局 4 个。结果如下图所示。

### 使用大数据的主要理由和商业利益



11. 如上图所示，各国统计局认为，“更快、更及时的统计数据”、“减少答复负担”和“统计资料编制过程的现代化”是使用大数据的主要好处，其次是“新产品和服务”及“降低费用”。有意思的是，三分之二的发展中国家认为满足可持续发展目标的指标等新的需求是一种好处，只有三分之一的经合组织国家赞同这一观点。

12. 许多统计局已经使用或考虑使用大数据来源，主要是扫描仪数据、网络采集数据、移动电话数据和卫星图像数据。这一结果大致符合实际报告的大数据项目；最常用的是移动电话数据(42 个项目)，其次是网络采集数据(31 个项目)和扫描仪数据(23 个项目)。然而，扫描仪数据和网络采集数据在经合组织国家的使用程度超过非经合组织国家。这些结果也反映在使用大数据最多的统计领域：价格统计(基于扫描仪数据)，其次是旅游业(移动电话数据)、人口(移动电话数据)以及运输和劳工统计数据(网络采集数据)。

13. 鉴于使用大数据一般涉及与其他统计局合作获得数据并掌握必要的技能，这份调查询问答复者建立了哪些挖掘数据项目伙伴关系。答卷者通常与政府机构结成伙伴，紧随其后的是研究和学术界。它们与数据提供者和信息技术公司建立的伙伴关系要少得多。

14. 此外，调查确定，所报告的大数据项目大都在探索和研究阶段。然而，如果某个项目打算投产，则是补充一个现有的数据来源。另外，调查结果显示了有趣的情况：发展中国家和发达国家都继续依赖传统的统计方法来处理和分析大数据来源。这一结果至少可以两种方式解释。首先，大数据项目显然可以使用传统方法处理和分析，这将大大降低实施大数据项目的门槛。其次，各国统

计局可能尚不十分熟悉先进的大数据处理和分析工作，这可能妨碍了大数据来源的充分利用。

15. 在回答“贵统计局亟需获得哪些技能以便更好地处理大数据？”的问题时，答复者表示，“关于大数据问题的方法论”、“数据科学”和“数学建模专业”是三个最需要的技能，而更以信息技术为本的“信息技术结构专业”、“数据可视化专业”和“网络安全专业”技能则得到较少的重视。虽然各国统计局清楚地看到有必要掌握这些新的、先进的大数据技能，但它们尚未开始聘用拥有这些技能的应聘者，也没有派其工作人员接受这种技能的培训。

16. 调查明确地询问了各国统计局可能迫切需要指导的情况。答复者表示，“大数据技能和培训”、“大数据质量框架”和“获取大数据”为三个最优先需要的指导，其次是“估算方法”、“使用网络采集数据”和“使用移动电话数据”。这一结果符合在大数据的方法论、估算方法和数据科学方面所指出的所需技能。

17. 从这些结果中可以得出以下结论：(a) 大数据专题的培训和能力建设(方法、估算和质量框架)是必要的；(b) 需要更多的试点项目，特别是发展中国家参与的项目。此外，更便捷地获取大数据将降低实施大数据项目的门槛。因此，培训、大数据方法和质量框架及数据使用已被选定为全球工作组工作方案的三个最高优先，并列入下届国际会议的主题。这次调查的完整报告已经作为背景文件提交给统计委员会。

#### 四. 第二届官方统计使用大数据问题国际会议的成果

18. 第二届官方统计使用大数据问题国际会议于2015年10月20日至22日在阿布扎比举办，主题是“从范例到指导”，出席会议的有来自世界各地的约250名统计员。这次会议由阿拉伯联合酋长国文化、青年和社区发展部长宣布开幕，他指出了大数据在《2030年可持续发展议程》中的意义，并提到了世界统计日的主题：“更好的数据，更好的生活”。会议展示了大数据项目利用移动电话数据、社交媒体数据和卫星数据而为各种统计应用程序取得的进展。在能力建设、数据获取和伙伴关系、质量和方法以及如何更加有效地宣传大数据的价值等领域也取得了进展。所有介绍均见会议网站。<sup>7</sup> 这届会议上的专题小组讨论中得出如下一些结论。

19. 大数据带来了巨大的机遇，但它不是“灵丹妙药”，且需要投入时间、金钱和能力。试点项目是确定可行方法的一个好办法。大数据的混合方法很可能是未来主要的和成功的模式，这意味着利用一系列来源的数据并作出符合目的的估算，其质量水平和不确定性得到确定和透明。归根结底，它仅涉及业绩：新的估算优于现有估算吗？

<sup>7</sup> <http://unstats.un.org/unsd/trade/events/2015/abudhabi/default.asp>。

20. 使用大数据符合统计编制现代化的议程，涉及到许多国家统计局的重大文化变革。在传播战略中，应当强调指出的是，各国统计局必须懂得，使用大数据是统计编制流程逐步变革进程的一部分。不久前，许多国家的统计局在企业普查的同时采用了企业调查，而企业调查则逐步成为收集商务统计数据的选用工具。此后，许多国家的统计局开始将行政(税务)数据与商业调查数据结合起来。虽然行政数据最初被认为不适合用于统计目的，但统计员今天几乎必须要用尽一切可能的现有行政数据，方可进行一项新的调查。随着时间的推移，在统计编制中使用某些大数据来源也会出现这种情况。

### 大数据方法和估算

21. 小组讨论得出的另一个结论是，大数据需要官方统计数据，正如官方统计数据需要大数据。这不仅是因为制作官方统计数据以国际商定的质量框架和方法以及专业独立性和信任原则为基础，而且是利用传统数据来源的官方统计数据使生成大数据来源统计数据的方法和技巧达到标准、“得到培训”并且最终得到证实。其他调查结果是，统计方法可通过取样等方式把大数据变为小数据，而数据转移并非总是必要的，因为这种方法或算法可在数据来源地点适用。

22. 此外，已经确定了卫星图像和地理空间数据在高时空分辨率环境监测方面对统计的潜力。在这方面，必须指出，在会议期间，全球政府间地理空间和地球观察机制——全球地理空间信息管理专家委员会和地球观测组织——为全球工作组探讨卫星图像和地理空间数据所带来的众多可能性的努力提供支助，包括如何利用现有方法来估算官方统计数据和可持续发展目标的各项指标。全球工作组也可在汇编关于联合国环境经济核算专家委员会主持下采取的生态系统程度、服务和条件措施的官方统计数据时，得益于在生态系统核算中进行的试验和测试。

### 培训和能力建设

23. 非常令人鼓舞的是，一些发展中国家正在探索大数据提供的机会(特别是卫星图像)，因为更为传统的统计领域中数据有限。都柏林的大数据“沙箱”是一次极好的低成本机会，使各国得以开始提高其对跨部门小组大数据的工具、方法和能力的熟悉程度。还应探索其他机会，包括利用私营部门和非营利组织的技能和能力，以利用如全球脉动等其人力资本和基础设施，并酌情建立长期伙伴关系。

### 数据获取和伙伴关系

24. 鉴于大数据项目从定义讲几乎涉及私营部门和研究机构的角色参与，故必须提出令人信服的企划案，说明为什么这种伙伴关系对所有参与方来说是双赢局面。长期伙伴关系，特别是与私营部门实体的这种关系，将需要对双方均有强大的刺激。同样，统计界在寻求伙伴关系，特别是与在全球开展业务的数据提供者的伙伴关系时，必须以统一的方式为自己定位。

## 五. 工作方案中的优先事项

25. 如上文第二节所述,全球工作组取得了一些进展。在以下方面已编写了文件:数据获取和伙伴关系的宣传、获取原则、良好做法,谅解备忘录的模板,大数据项目的分类、质量和方法及储存库。此外,项目的储存库是全球工作组的一项重要交付成果,应服务于统计界以及更广泛的受众。试点项目正在利用卫星图像、移动电话和社交媒体的数据而继续展开。然而,要从研究到生产及从范例到指导,还需要做更多的工作。

26. 全球调查问卷的答复者表示,迫切需要技能和培训、大数据的质量框架、获取大数据和估算方法方面的指导。会议得出的结论也反映了这种培训、方法指导和获取数据方面的需要。因此,全球工作组未来的工作方案将优先考虑这些专题。

27. 关于获取数据问题,全球工作组将继续着力于获取数据的范围和原则的制定,以及其形式和包装方式。原则的范围被认为是“获取专利数据”,包括私营部门实体、政府机构或其他组织持有的数据。预计这项工作将与各利益攸关方、尤其是统计界以外者密切协商而推进,其结果将在定于都柏林举行的下届大数据会议上介绍。

28. “边做边学”的方法是建设将大数据用于官方统计的能力的方法之一。全球工作组将继续积极参与若干试点项目,其中一些项目将用于衡量可持续发展目标的指标。发展中国家尤其应利用这一机会参与试点项目。它们可以通过都柏林“沙箱”提供的而积极参与。此外,还将进一步探讨使用雅加达和坎帕拉的“全球脉动”实验室的问题。

29. 鉴于表示需要建立大数据方法和质量保证框架,全球工作组同意更多地按照大数据质量框架的轮廓安排其工作,将其工作方案的内容列于“投入质量”(数据获取)、“通量质量”(大数据方法和估计方法、大数据项目的储存库和统计资料编制过程)和“产出质量”(大数据用于官方统计,包括可持续发展目标的指标)之下。各任务组这次工作重心调整和略微重新安排,还可促进分享试点项目的结果并将其纳入任务组关于数据获取、方法和培训的机会的工作。

30. 鉴于第三届官方统计使用大数据问题国际会议预计将于2016年6月28日至30日在都柏林举行,临时议程按大数据质量框架的方针编写,具体如下:专题1,“数据获取和伙伴关系:双赢的局面”,专题2:“统计数据编制:大数据的作用及所需技能和能力”;专题3:“官方统计使用大数据和可持续发展目标的各项指标”。

31. 为了进一步强调全球工作组对培训和能力建设的高度重视,已商定将在都柏林会议期间举办一个或多个讲习班。小组建议,根据全球调查得到的答复中表达的需求,举办一个关于大数据的方法的讲习班。还可以举办另一个关于“领导和管理大数据项目”的讲习班,讨论与解决问题、制定企划案、获取数据来源、项

目管理和统计程序相关的问题。最后，有人提出，也应举办一次题为“大数据的当代和新的信息技术工具和技术”的讲习班。

## 六. 讨论要点

32. 请委员会就以下问题发表看法：

- (a) 官方统计使用大数据问题全球工作组的工作进展；
- (b) 全球大数据战略和项目调查结果；
- (c) 第二届官方统计使用大数据问题国际会议的成果；
- (d) 全球工作组工作方案的优先事项：
  - (一) 官方统计使用大数据试点项目，包括衡量选定的可持续发展目标的指标的项目；
  - (二) 使用大数据的培训和技能涉及到更多的发展中国家，特别是因为为培训目的而使用都柏林的“沙箱”及全球脉动的实验室；
  - (三) 进一步拟订获取专利数据的指导原则，以增强数据的获取；
  - (四) 关于大数据方法和估算方法的指导；
  - (五) 官方统计使用大数据的质量框架。