



统计委员会

第四十八届会议

2017年3月7日至10日

临时议程* 项目 3(d)

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

依照经济及社会理事会第 2016/220 号决定和以往惯例，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。过去三年期间，全球工作组确定了公共和私营机构在国家和区域各级开展的许多大有前途的大数据举措，这些举措都旨在使数据、服务和应用程序更加便利，为推动研究和能力建设加快数据、服务和应用程序之间的协同增效。本报告介绍工作组取得的进展及其今后工作构想和建议。报告提议建立全球平台，汇总所提及的在全球大背景下对推动使用大数据和其他数据源有现实意义的所有举措。该平台可为数据、服务、应用程序和基础设施提供规模效应和可扩放性，促进使用大数据并使其与行政资源、地理空间信息以及传统调查和普查数据实现整合。该平台将在统计委员会主持下设立。邀请统计委员会对拟议全球平台以及全球工作组取得的进展和提出的建议发表意见。讨论要点载于本报告最后一节。

* E/CN.3/2017/1。



官方统计使用大数据问题全球工作组的报告

一. 引言

1. 2014 年举行的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。工作组按照其职权范围(见 E/CN.3/2015/4)和统计委员会第 46/101 号决定(见 E/2015/24-E/CN.3/2015/40)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。工作组还推动能力建设、培训和经验分享以及政策应用程序使用大数据的工作。

2. 工作组在统计委员会 2015 年和 2016 年举行的第四十六和四十七届会议上分别向统计委员会提出报告。工作组介绍了分别在北京和阿布扎比举行的第一届和第二届全球会议及其工作方案的重要内容,包括培训、技能发展和能力建设、将大数据与可持续发展目标挂钩、倡导和宣传、数据获取和伙伴关系以及分类和框架等共有问题,并探讨了官方统计使用的特定大数据源,包括移动电话及社交媒体数据和卫星成像。

3. 2016 年,统计委员会要求工作组完成一揽子举措,以改善大数据等专有数据的获取渠道;与区域培训机构合作,特别针对发展中国家开发大数据使用和应用培训课程,以及进一步构建大数据质量框架以确保官方统计汇编受到信赖。

4. 统计委员会还要求工作组在工作方案中增加以下内容:审查收集大数据所用网络采集办法;改进工作组网站和大数据储存库;在发展中国家宣传大数据;将工作组的工作与《2030 年可持续发展议程》统计伙伴关系、协调和能力建设高级别小组的工作挂钩。

5. 工作组应统计委员会的要求采取了行动,见第二节全球工作组 2016 年进展情况概览和第三节所述官方统计使用大数据问题第三次全球会议的成果。第四节特别着重介绍获取专有数据问题。第五节则提出了设立大数据、服务和应用程序全球平台的建议。报告在最后提出了一些讨论要点。

二. 全球工作组的进展

6. 2016 年,全球工作组在都柏林举行了第三次大数据问题全球会议。会议提出了设立全球数据、服务和应用程序协作平台的建议,并商定将获取专有数据问题纳入官方统计质量框架的全面审查工作。工作组还在其他问题上取得进展,包括协作开展欧洲经济委员会“沙箱”试点项目、编制若干卫星数据、移动电话数据和社交媒体数据使用手册以及更新工作组网站,包括在网站上提供大数据项目清单。本节详细介绍各任务小组在各自具体工作领域取得的进展。嗣后各节则摘要说明第三次全球会议的主要成果、设立全球平台提案以及获取专有数据的各项建议。

A. 卫星数据

7. 工作组的卫星成像、地理空间数据和遥感任务小组编制了一本手册，载列地球观测数据来源、使用卫星成像数据编制作物统计和其他统计的方法、任务小组试点项目概览以及各国统计局首次探索使用地球观测数据指南。试点项目包括澳大利亚统计局应用卫星成像数据编制农业统计、Terra Bella(谷歌子公司)使用专用卫星开展商品库存评估、墨西哥国家统计地理所对墨西哥各区域气候假设的环境分析以及哥伦比亚国家统计局利用卫星图像汇编地表覆盖和土地使用统计。该手册经与联合国粮食及农业组织和欧洲联盟委员会的作物生产和地表覆盖专家举行专家组会议审查后，预计于 2017 年定稿。

8. 任务小组还在开发实践课程，讲授如何利用地球观测数据编制农作物统计，并提供有关利用地球观测数据计量实现可持续发展目标进展情况的案例研究。课程教材提供了官方统计使用卫星成像数据的背景资料，并就数据源选择和生成统计产出方法提出建议。课程内容结合了情况介绍、讨论课程、如何使用“R”软件包的演示(即示范在谷歌地球引擎内处理免费卫星分析数据)以及协助实际操作利用卫星成像数据制作统计产出。这份材料经过进一步调整后可适用于农业地区的生态系统核算。2017 年 1 月举办的讲习班将对该教材进行审查，然后将在亚洲和拉丁美洲举行其他讲习班。

B. 移动电话数据

9. 移动电话数据任务小组借鉴欧统局的几个项目以及 Positium 公司参与的阿曼境内项目所取得的经验，着手编写一本手册，预计于 2017 年定稿。该小组还可借鉴欧洲统计系统英才中心和网络在比利时、法国、荷兰、斯洛文尼亚和西班牙开展的项目工作。这份手册将包括各国统计局与移动运营商之间的商业伙伴关系模式，以及(国际、国内和国家以下各级)旅游统计、日间人口统计(当日旅游统计)、事件统计、运输统计和人口密度统计等方面的应用程序。该手册将提供实际指导、国家实践资料和大量参考，对发展中国家确实特别有用。

C. 社交媒体数据

10. 同样，社交媒体数据任务小组也着手编制手册，预计于 2017 年完成。该手册将说明如何利用大数据技巧来使用社交媒体数据编制统计数据。该手册将包括以下方面的说明：各种社交媒体数据源(如推特和脸书)、包括自然语言处理在内的相应方法和技巧以及从社交媒体非信息部分和网页内容提取资料的方法。手册还将包括用于汇编消费者信心指数、报告公共舆论指标以及分析出行方式和情感的一些应用程序。除了这本手册之外，作为社交媒体任务小组成员的“联合国全球脉动”小组还与推特达成每月使用大量推文的协议，可以通过在“沙箱”中使用这些推文，与全球工作组成员共享。

D. 大数据项目清单

11. 除了全球工作组 2015 年大数据项目的调查外，大数据与可持续发展目标任务小组还就大数据使用情况，特别是大数据用于可持续发展目标的情况对各类机构进行了调查。此次调查收集的资料进一步扩大了大数据项目清单。该清单是一份目录，载列与编制官方统计数据 and 可持续发展目标指标以及用于管理和监测公共部门方案和其他类型统计相关大数据项目。清单维护和更新程序已实现标准化，按可持续发展目标及其具体目标安排大数据项目应该能进一步加强大数据在官方统计方面的作用。这份独特的大数据清单制定后，工作组将做出进一步努力，宣传这份官方统计界和更广泛的大数据业界使用大数据的实践目录。

E. 培训、技能和能力建设

12. 为了在多源统计编制流程中应用大数据，官方统计界在编制官方统计时使用大数据和新信息技术基础设施的情况越多，对新技能和新信息技术方法的需求就越大。大数据的来源不同，对如何处理和分析数据以及如何使用信息技术基础设施带来挑战。为了克服知识转让领域的这些挑战，培训、技能和能力建设任务小组着重就以下四个主要目标开展工作：

(a) 拟订方法和工具(包括提供协调和便利)，以便在国家统计系统开展大数据技能基线需求评估，并开展统计编制工作使用大数据机构就绪情况评估；

(b) 以技能和机构就绪状态基线需求评估成果为基础，指导编制模块式培训方案；

(c) 推动建立培训和研究机构全球网络，以便在使用大数据编制统计资料方面开展培训和能力建设；

(d) 将高级培训方案和工作组各任务小组开发的培训课程与官方统计使用大数据问题全球调查协调一致。

13. 关于 2017 年，该任务小组拟订了以下交付成果：

(a) 开展涵盖国家统计机构的新一轮大数据全球调查，评估基本技能需要以及各机构是否已准备就绪；

(b) 与各国统计局、学术机构和私营部门伙伴协作举办培训课程，以开发统计编制流程使用大数据所需初级基础技能和提高机构就绪水平；

(c) 建设由国家和区域培训和研究机构以及数据创新中心组成的全球能力建设网络。

三. 官方统计使用大数据问题第三次国际会议的成果

14. 第三次全球会议于 2016 年 8 月 30 和 31 日及 9 月 1 日在都柏林举行，世界各地的大约 250 名统计师出席了会议。爱尔兰负责金融服务、电子政务和公共采购的国务部长奥恩·莫菲先生宣布会议开幕，他指出大数据在《2030 年可持续发展议程》框架内具有重大意义。会议展示了各种统计应用程序使用移动电话数据、社交媒体数据和卫星数据的大数据项目取得的进展和成果。会议进一步展示了在能力建设、数据获取和伙伴关系、质量和方法以及如何更好地宣传大数据价值等领域取得的进展。会议网站载有所有介绍资料。¹

15. 会议涵盖三大主题，包括大数据获取和伙伴关系；官方统计转型和现代化能力建设战略；官方统计多源大数据统计编制环境，包括可持续发展目标监测。

A. 数据获取和伙伴关系

16. 数据获取和伙伴关系高级别小组强调，通过公私伙伴关系获取移动电话、社交媒体和其他大数据，将为私营伙伴和各国统计局的服务提供带来巨大好处。这一伙伴关系应建立在云计算技术基础设施、私营部门的数据和数据分析专门知识以及公众相信各国统计局能发挥独立专业作用提供高质量公平信息基础之上。该小组确定保持公众信任是官方统计使用大数据的一个基本要素。为此，收集、储存和使用移动电话或社交媒体数据必须保持充分的透明度，同时按照有关官方统计数字和数据保护的监管框架和法律框架开展行动。其好处包括减少合规费用和成本，提高统计资料的及时性和质量。但是，官方统计界应该认识到，必须探索创新性筹资机制，以满足伙伴关系中私营部门的营利目的。本报告第四节进一步阐述获取专有数据问题。

B. 官方统计创新能力建设

17. 鉴于有必要管理外部举措构成的下行风险、其他统计资料提供方的竞争和预算削减因素，同时通过快速的技术发展和数据革命收获不断变化的信息和通信技术和数据发展收益，目前无法再采用常规做法为国家和国际客户提供统计服务。

18. 能力建设小组指出，国家统计系统需要实现多层面转变，以应对新挑战和赢得新机会。应该扩大现有的能力建设方案，并尽可能关注技术架构转型、转变员工队伍、利用更多数据来源和调整产品和服务方向。小组指出，技术架构转型应推动从现场有形信息技术设备转向采用云计算环境，以及在数据收集、登记、元数据和数据管理、分析和传播方面采用共同服务和应用架构。

19. 在投资于技术架构转型的同时，应该开展能力建设方案，支持国家统计系统的工作人员，包括使用新的多源数据和现代技术的数据科学家和数据工程师、加

¹ 见 <http://unstats.un.org/unsd/bigdata/conferences/2016/default.asp>。

强法律环境的律师以及在不断提高质量标准基础上引领企业文化变革的管理人员，逐步掌握多种新的配套技能。有了这些新的能力就可以采取灵活且适应新需求的标准化企业业务架构。这些能力应该以流程而不是产品为基础，同时越来越多地使用行政来源和大数据源(而不再使用初始数据源)编制多种统计产出。此外，我们的传播和宣传战略也应升级并具有更大的适应能力，以便通过采用各种数据传播技术，包括移动设备应用程序和关键调查结果数据可视化，对不同用户部门具有针对性。

20. 区域统计培训方案拟逐渐拓宽自身的能力建设方案，增进国家统计局系统高级管理人员推动机构转型与现代化的能力，并提高这些系统的技术和信息技术工作人员从事数据工程、数据分析和数据科学工作的能力。为了通过这些新的培训方案开展能力建设，区域统计培训机构可采用具备多源大数据处理能力的区域和全球云计算平台。因此拟创建培训和 15 研究机构国际网路，通过区域和全球平台相互联系，从而为持续学习新的统计方法、新的管理技能和新的信息技术技能提供机构、技术和数据基础设施。联通全球平台的区域网络可以成为既有国家培训机构和新建国家数据和研究中心的补充。

21. 官方统计创新智慧投资小组表示，技术和数据格局瞬息万变，必须重新审视捐助界对国家、区域和国际统计系统的支持情况。在快速变化的环境下，首先要考虑把试行、检测、继而推广创新方法和技术的小型项目作为智慧投资的重点。此类新兴项目和方案要求在官方统计界、学术界、技术企业和专有数据所有者之间建立新的协作和战略公私伙伴关系。似乎还需要作出调整以处理持续获取私营大数据源和现代云端技术基础设施所需经常费用问题。这需要改变捐助支付方式，从对国家统计局结构性投资的前期供货改为对可扩放的可共同区域和全球平台的系列分期投资，通过战略伙伴关系获取数据、服务和应用程序。

C. 大数据和可持续发展目标的各项指标

22. 大数据和可持续发展目标专题组成员还强调，需要采取混合式多源办法满足可持续发展目标指标框架的要求。挑战在于这些指标不仅数目庞大，而且细致入微，要一一达标且“不让任何人掉队”，为了一一达标，需要从行政和大数据源获取资源。随着对获取技术、数据 and 应用程序配合实施主要统计项目的需求不断涌现，塞内加尔、佛得角和南非在建立国家、区域技术和数据中心方面出现了可喜动向。第三届全球会议的嘉宾普遍对打造全球平台的提议表示支持，认为这是在数据、服务和应用程序共享方面开展国际、区域合作的积极进展，有助于应对可持续发展目标各项指标监测工作所带来的挑战。

23. 有关第三届全球会议成果的更多信息载于本报告的背景文件。

四. 关于将获取大数据作为官方统计质量框架组成部分的建议

24. 官方统计基本原则的第五项原则涉及数据获取，该原则阐明，“可为统计目的从各种来源提取数据，不论是统计调查还是行政记录。统计机构在选择来源时应考虑到数据的质量、及时性、成本和给应答者造成的负担”。这项原则在解释中可被认为包括大数据的获取。

25. 工作组担心的是，在很多情况下，获取大数据指的是获取那些追求商业利益的私营专有数据所有人掌握的数据。这就决定了获取大数据不同于获取传统数据源，而且令人担心的是，统计机构在与此类数据所有人达成协议时，能否保持独立性，确保隐私和保密性以及编制优质统计数据。有待审议的事项包括：公平分配数据所有人的负担、数据提供所需成本与工作、数据对企业价值主张的作用、在公共和企业利益之间取得公允平衡以及统计机构和数据供应方之间的业务安排。为保证官方统计质量和公众信任，此类协议得规定：提供数据要透明，元数据的说明要清楚，使用数据要附带公开免责声明。

26. 2015年，工作组起草了一套关于官方统计使用大数据源的原则，²目的是帮助统计机构获得新数据源，同时在大数据所有者(通常在私营部门)的正当利益与大众需要获得更好的官方统计这一需求之间达成平衡。使用大数据源的原则与官方统计基本原则挂钩，吁请数据提供者抱有社会责任感。2015年，工作组在多个场合就大数据源使用原则的内容征求了利益攸关方的意见，收到的反馈基本上都是积极的。

27. 2016年8月在都柏林开会期间，工作组讨论了今后应采取哪些步骤让全球统计知晓和认可这些原则。首先，工作组同意“原则”一词可能过于强烈，不易为任何政府间机构所接受(通过这些基本原则花了20年)，最好改成“建议”。此外，还可通过定向调查更多地征求利益攸关方的意见。最重要的是，会上商定统计委员会应在官方统计基本原则、现行官方统计业务守则和国家质量保证框架的总体框架内审议大数据源使用建议。

28. 提交统计委员会第四十八届会议的关于官方统计基本原则的报告(E/CN.3/2017/9)要求统计委员会考虑成立“主席之友”小组，除其他外负责审查如何解释基本原则并将它们用于私营部门，特别是在使用专有数据方面。主席之友小组不妨考虑到工作组对获取专有数据的顾虑。有关获取专有数据的补充资料，见本报告背景文件。

² 见 <http://unstats.un.org/unsd/trade/events/2015/abudhabi/default.asp>。

五. 数据、服务和应用程序全球平台

29. 2014 年 11 月，秘书长转递了数据革命促进可持续发展问题独立专家咨询小组题为“世界需要数据：发动数据革命促进可持续发展”的报告。报告载有以下方面的建议：(a) 通过技术、创新和分析，搭建数据创新网络网，以利用和共享数据与数据研究；(b) 实施能力建设，开拓能力建设与技术转让资源，提高数据素养，并与私营部门结成伙伴关系，通过创新融资机制调动资源；(c) 加强治理，引领政府、私营部门、非政府组织、媒体与学术界之间形成伙伴关系并进行协调，以推广有关数据共享、开放数据和数据权利的良好做法和原则。

30. 过去三年中，工作组发现公共和私营机构在国家和区域层面实行了许多前景可观的大数据措施，力求使数据、服务和应用程序易于使用，并加快形成合力，促进研究与能力建设。工作组中负责卫星数据、移动电话数据和社交媒体数据等领域的各个任务小组，以手册为形式、用培训教材和培训课程作补充，力求使应用程序易于使用。在借助扫描数据和网络采集编制价格指数等有前景、能对使用官方统计数据起到示范作用的领域，目前已提出和成立新的任务小组。然而，旨在搭建与数据创新网络网类似平台的综合努力尚未落实到位。

31. 工作组继 2016 年 8 月 29 日都柏林会议作出决定之后，提出建立全球数据、服务和应用程序平台。全球平台可借鉴公共和私营部门大数据举措的最佳做法，在向官方统计界提供数据创新所需技术基础设施和网络的同时，满足下列对全球中枢的需求：

(a) 官方统计师、数据科学家以及公共和私营部门网域专家相互交流大数据处理、分析和可视化方面的构想和方法；

(b) 存储大数据相关处理、分析和可视化方法、服务和应用程序，供持续开发和循环利用；

(c) 通过使用官方统计进行试点研究和案例研究，展示大数据对改善决策的价值；

(d) 提供大数据促进能力建设方面的培训教材和讲习班资源。

32. 工作组还提议在统计委员会主持下设立这个平台。这意味着委员会将就平台的业务模式以及设立和落实该平台的路线图作出决定。在统计委员会的监督下，该平台将由主要为发达国家和发展中国家统计系统服务的官方统计界管理和享用。如上所述，建立全球平台的提案经过调整并得到 2016 年举行的第三届官方统计使用大数据问题国际会议和关于官方统计变革议程的系列区域会议各专家组的支持。该提案也与《2030 年可持续发展议程》统计伙伴关系、协调和能力建设高级别小组制定的可持续发展数据全球行动计划草案所载“国家统计系统创新

与现代化”战略领域的目标(见 E/CN.3/2017/3), 以及各次区域会议的成果文件(见 E/CN.3/2017/5)相一致。

全球平台暂行业务模式

33. 为帮助统计委员会思考全球平台的结构和内容, 以下段落简要说明平台的暂行业务模式, 其中提出了平台建设的办法。工作组成立了一个委员会, 负责探讨平台建设可能采取的其他办法。

34. 全球平台可在大数据的获取、大数据使用能力建设及其与行政资源、地理空间信息以及传统调查和普查数据相结合方面, 还可在相关服务、应用程序和基础设施的使用方面提供规模效应和可扩展性。在产品和服务方面, 全球平台能够:

- (a) 协助获取并使用信息技术基础设施和专有数据源;
- (b) 承接特定全球项目, 例如全球企业名册项目以及关于双边和多边数据不对称问题的项目;
- (c) 存储并提供新的开源应用程序, 以便利用多源数据集更及时、更灵活地生成官方统计数据;
- (d) 在使用新数据源码编制统计数据方面提供培训和能力建设服务。为了成为真正的官方统计全球平台, 设想数据获取也可包括获取可信赖的官方统计资料来源(微观)数据集, 以及获取公私伙伴组织提供的已经核实和校验的数据。全球平台的服务可以包括云存储和云计算服务、技术统计服务(包括独立服务或协助开发应用程序基础服务)和相关能力建设服务。

35. 下举一例说明可如何有效提供服务 and 应用程序。可以启动一个项目用于转让在发达国家或发展中国家计算农业统计数据时使用的方法, 并允许使用卫星图像数据和技术基础设施进行计算。该项目需要使用卫星数据进行培训和测试; 在安全和适当的信息技术环境中编制数据; 对方法进行测试; 编制农业统计数据以及培训工作人员执行例行相关行动。全球平台是一个包括统计机关、研究机构和私有企业在内的伙伴网络, 它将能够依靠自身的技术基础设施, 推出并提供这些服务, 包括与培训和能力建设有关的服务。

36. 业务制度环境也需要考虑。为敲定全球平台业务论证的细节和制定全球平台实施路线图, 工作组设立了一个由各国统计局包括近期成立国家数据中心的统计局代表组成的委员会。该委员会将运用成员积累的经验和最佳做法, 设计和落实全球平台。工作组还将与私有数据和技术企业以及国际和区域机构分别进行对话。在制定该平台实施工作的业务论证和路线图时, 委员会将审议技术基础设施的部署、从私营数据所有人获取专有数据、共享现有的和开发新的大数据应用程序、按照公共和私营部门的方针推出能力建设服务、治理和融资安排, 包括统计委员会和工作组的定位和作用)、与国家数据中心、欧洲经济委员会的“沙箱”项目、非洲区域中心和全球脉动实验室等国家、区域和全球举措对接。本报告的背景文件提出了一份初步路线图, 附带重要节点和时间表, 包括β版测试时间表。

37. 最后，工作组提出在统计委员会主持下建立一个数据、服务和应用程序全球平台。该平台可成为一个包括配备适当技术基础设施的统计单位、研究机构和私营企业在内的伙伴网络，为官方统计界提供和交付数据、应用程序和服务，包括培训和能力建设服务。该平台的建设工作将依照背景文件所述路线图逐步推进。

六. 讨论要点

38. 邀请统计委员会就下列问题发表意见：

- (a) 联合国官方统计使用大数据全球工作组各任务小组取得的进展；
 - (b) 第三届官方统计使用大数据问题国际会议的成果；
 - (c) 工作组关于在审议拟议官方统计基本原则主席之友小组时考虑到专有数据获取问题的建议；
 - (d) 工作组关于在统计委员会主持下逐步建立服务于官方统计界的数据、服务和应用程序全球平台的提案。
-