



统计委员会

第五十三届会议

2022年3月1日至4日

临时议程* 项目 3(u)

供讨论和决定的项目：大数据

官方统计使用大数据和数据科学专家委员会的报告

秘书长的说明

依照经济及社会理事会第 2021/224 号决定和以往惯例，秘书长谨转递官方统计使用大数据和数据科学专家委员会的报告。根据统计委员会第 52/112 号决定，专家委员会处理所提出的问题，并请统计委员会就专家委员会今后工作的拟议方向发表意见。专家委员会提议将大数据和数据科学的使用纳入各国统计局工作方案的主流，具体做法是促进专家委员会各工作组与各专题专家组加强合作，将大数据和数据科学培训纳入各国统计局的培训课程，并为专家委员会的专家以及整个官方统计界维持一个积极的交流平台。其他建议包括根据政策考虑确定新活动的优先次序，创建一个数据科学中心和大数据区域中心负责人论坛，制定更具战略性的与私营部门接触的办法，以及加强与地理空间信息界的合作。关于大数据和数据科学能力发展未来方向的建议主要包括增强区域中心的权能，将其作为每个区域各国统计局的培训和项目执行点。

* E/CN.3/2022/1。



官方统计使用大数据和数据科学专家委员会的报告

一. 引言

1. 统计委员会在第 52/112 号决定(见 E/2021/24-E/CN.3/2021/30, 第一章 B)中敦促官方统计使用大数据和数据科学专家委员会¹ 加强将大数据和数据科学用于官方统计的方法, 支持全球统计界所有成员利用联合国全球平台和各区域中心开展培训和项目活动, 并支持设立一个关于获取全球私营部门数据的工作组和一个财务委员会, 以便为全球平台开拓坚实的供资办法。

2. 专家委员会通过编写方法手册、增加全球平台上的项目数量以及进一步改进培训方案, 处理了统计委员会提出的问题。第二节介绍各工作组、全球平台和区域中心取得的最新进展。第三节说明专家委员会今后工作的方向, 并概述今后的步骤。第四节列出有待统计委员会采取的行动。

二. 官方统计使用大数据和数据科学专家委员会的进展情况

A. 工作组

3. 专家委员会通过九个工作组开展大部分工作, 这些工作组制定方法和准则, 积极参与全球平台的项目, 并与四个区域中心合作开展能力建设活动。2021 年, 各工作组专家为许多网络培训研讨会作出了贡献, 包括 2021 年统计新方法和新技术在线会议、² 国际统计学会第六十三届世界统计大会³ 和第三届联合国世界数据论坛⁴ 上的网络研讨会。专家委员会网站对各项活动作了更详细的概述。⁵ 本节介绍每个工作组的进展情况。

4. 统计委员会第五十二届会议设立了由荷兰牵头的全球便利获取私人持有数据工作组, 该工作组于 2021 年拟订了其职权范围。⁶ 它是一个联合工作组, 将向专家委员会和经济统计师网络报告工作。⁷ 联合工作组将推动数据获取用例的共同编制和共同投资, 并将与欧洲统计系统关于将私人持有的数据用于官方统计的小组以及其他相关国际团体和组织协调其工作并开展合作。

5. 联合工作组的目标是: (a) 在全球一级就获取私营数据所有者的数据源进行谈判, 这些数据源严格用于统计目的, 为国家、区域和全球各级的决策提供参考

¹ 以前称为官方统计使用大数据问题全球工作组, 见第 52/112 号决定。

² 见 <https://www.ntts2021.com/>。

³ 见 <https://www.isi2021.org/>。

⁴ 见 <https://unstats.un.org/unsd/undataforum/index.html>。

⁵ 可查阅 <https://unstats.un.org/bigdata>。

⁶ 见 <https://unstats.un.org/bigdata/task-teams/data-access/index.cshtml>。

⁷ 见 E/CN.3/2022/19。

信息，特别是推动落实《2030年可持续发展议程》；(b) 通过用例证明这些数据源与统计目的的相关性；(c) 在成功完成试验和测试后，就体制安排提供咨询意见。

6. 2022年初，联合工作组将组织一系列虚拟冲刺(即密集的工作会议)，以推进其工作。国家、区域和国际统计机构以及与学术界、研究和私营部门有关的机构都受邀参加这些冲刺。建议将获取与全球价值链和电子商务有关的数据源作为第一系列冲刺的重点。将在本报告的背景文件中介绍相关成果，该文件将与经济统计师网络联合编写。

7. 由加拿大牵头的地球观测数据工作组最近完成了一份经同行审查的报告，其中包括关于使用地球观测数据编制农业统计数据的项目集。该工作组还提供了地球观测数据产品及关于其潜在用途的指导意见，并通过联合国粮食及农业组织“手把手”地理空间平台提供了阿富汗和莱索托的现场数据。工作组即将提供一个提高认识的视频，概述地球观测数据及其在农业统计中的使用。工作组还在进一步开发一个可搜索的知名实体在线培训课程数据库和一个创建定制课程方案的界面，以满足个人在使用地球观测数据方面的培训需要。

8. 2021年，由国际电信联盟牵头的移动电话数据工作组编写了五本手册，分别涉及使用移动电话数据进行灾害期间流离失所情况统计、动态人口制图、信息社会统计、移民统计和旅游统计。这些手册已由领域专家进行同行审查，目前正在交叉检查，以确保一致性、连贯性和格式统一。工作组即将提供一个提高认识的视频，介绍使用移动电话数据进行各种官方统计的情况。

9. 工作组最近发布了一份文件，⁸介绍维持公众对将移动运营商数据用于政策目的的信任的指导原则。该文件提到2020年3月在爱沙尼亚、冈比亚和加纳分别开启的三个项目(每个项目都有一名工作组成员参与)，目的是在地方一级迅速、经常地监测政府干预措施的效果，以在冠状病毒病(COVID-19)大流行期间遏制病毒的传播。该文件提出了确保公众对在此类紧急情况下使用移动电话数据的信任的五项原则，即必要性和相称性原则、专业独立性原则、保护隐私原则、承诺质量原则和国际可比性原则。

10. 由大不列颠及北爱尔兰联合王国牵头的扫描仪数据工作组在起草关于价格(包括消费价格)的电子手册方面取得了重大进展，该手册涵盖了从获取数据到应用价格指数方法的端到端过程。还正在制定关于消费价格数据来源分类的指导意见，所选择的方法取决于具体的数据来源。此外，工作组正在创建一系列培训课程，约有10门，其中前两门将于2022年初发布。工作组与欧盟统计局于2021年10月举办了一次线上讲习班，⁹来自50多个国家的参与者交流了在消费价格统计中使用新数据源的实际经验。

11. 由秘书处经济和社会事务部统计司牵头的自动识别系统数据工作组在2021年升级了其自动识别系统平台，以便获得更好的用户体验和成本效益。在全球平

⁸ 见 <https://doi.org/10.1017/dap.2021.21>。

⁹ 见 https://ec.europa.eu/eurostat/cros/content/workshop-scanner-data-web-scraping_en。

台上使用此类数据的成本是前几年的主要问题,但 2021 年的重新设计非常成功,将成本降低了约 70%。这意味着平台现在接纳更多自动识别系统相关项目。工作组还发布了关于自动识别系统的基础电子学习课程,¹⁰ 并获得了载有船舶特征数据和业务信息的船舶登记册,该登记册根据船舶登记的修订情况或新的船舶登记情况定期更新。工作组还开发了教授与自动识别系统有关的一系列核心和具体技能的课程。全球平台上目前有几个项目,特别是跟踪太平洋岛屿国家贸易情况的应用程序、¹¹ 监测东非港口效率的应用程序以及生成关于巴拿马运河周围海洋排放情况的高频统计数据的应用程序。

12. 由丹麦牵头的可持续发展目标大数据工作组于 2021 年创建了一个可利用大数据计算的可持续发展目标指标库。这些指标是通过与广泛的利益攸关方、数据制作方和组织机构协商确定的。目前,指标库存有六个指标。已在专家委员会网站上提供了易于查阅的所确定指标的概述,并说明了方法和相关联系人。由世界银行牵头的衡量农村四季公路通行情况工作组正在与可持续发展目标大数据工作组合作,将在 2021 年底和 2022 年初举办两次活动,¹² 就指标 9.1.1 的计量提供实际指导。

13. 由荷兰牵头的隐私保护技术工作组于 2021 年编写了一本手册,这是其第二本手册。手册介绍了新的技术、标准和框架,重点是隐私保护技术对取自各国统计局和国际组织正在进行的的研究的具体统计用例的适用性。此外,工作组正在制定关于隐私保护技术的法律方面的准则。手册和准则将于 2022 年初发布。工作组还在制定培训方面取得了进展。工作组与 Openmined.org 社区合作,推出了关于“我们的隐私机会”和“隐私计算基础”的首批两门免费电子学习课程,¹³ 有 7 500 多名学习者注册。2022 年还会继续推出更多课程。

14. 在过去几个月里,工作组启动了联合国加强隐私技术实验室,以加快在官方统计界采用加强隐私技术。该实验室包括以下三个核心支柱:

(a) 试验:一系列正在开展的试点项目,重点是评价用于官方统计实际用例的增强隐私的技术;

(b) 外联和培训:通过培训、公开讲座和教育活动,促进整个统计界就使用增强隐私技术开展共同学习并传播相关见解;

(c) 支助服务:向希望利用增强隐私技术的统计界人士提供支持和咨询意见的机制。

¹⁰ 见 <https://learning.officialstatistics.org/course/index.php?categoryid=8>。

¹¹ 见 <https://www.imf.org/en/Publications/WP/Issues/2021/08/20/Tracking-Trade-from-Space-An-Application-to-Pacific-Island-Countries-464345>。

¹² 见 <https://unstats.un.org/bigdata/blog/2021/road-to-expo2020.cshtml>。

¹³ 见 <https://courses.openmined.org/>。

15. 此外，工作组与欧洲经济委员会官方统计现代化高级别小组下的一些项目小组¹⁴ 密切合作，特别是与从事输入内容隐私保护、合成数据和机器学习工作的小组密切合作。由联合王国牵头的机器学习小组已成为在将机器学习用于官方统计方面进行研究合作、知识交流和能力建设的重要平台，拥有来自 30 多个国家的约 250 名专家。重点工作包括发布道德准则、部署数据库以将机器学习模型投入生产，以及为估计美利坚合众国各州的消费性开支创建模型。2022 年 1 月，机器学习小组与专家委员会将在 2020 年世博会上共同组织一次活动。

16. 由波兰牵头的培训、技能和能力建设工作组提出了解决方案，以帮助正在或考虑在官方统计中使用大数据的机构建设能力。2021 年，该工作组越来越多地向专家委员会的其他工作组提供实际操作支持，以制定培训科目和设计培训课程。该工作组还制定了大数据总体课程草案，并将提高认识培训课程确定为优先事项。已开发大数据一般入门培训课程，向用户介绍主要概念和大数据在官方统计中的应用。预计该课程将成为不同工作组在各自领域开发的基础课程的先修课。

17. 已设立一个学习管理平台，其中包含专家委员会各工作组开发的所有培训课程。该平台包括已经开发或正在制作的非同步课程，以及未来的同步和混合课程。工作组根据专家委员会内外开发的现有资源，编制了一份培训课程和材料目录。此外，还开发了一个与能力框架各个方面和级别挂钩的网络应用系统，使用户能够发现相关课程，并创建关于大数据的个人学习路径。最后，工作组已开始与区域中心就培训活动的规划和协调进行对话，并使它们掌握在其区域内就建设大数据能力提供咨询和支持的技能。

B. 联合国全球平台

18. 除了上一份报告(E/CN.3/2021/14)中列出的项目外，2021 年在联合国全球平台上启动了以下项目：

(a) 经济合作与发展组织关于利用可持续性报告中的信息的项目。该项目的重点是利用机器学习技术系统分析(50 多个国家的)数千份可持续性报告所载的文本信息，以便制定各公司可持续性方向的新措施；

(b) 拉丁美洲海事技术合作中心与自动识别系统数据工作组共同开展的项目，以生成航运活动产生的排放情况的高频(每日和每周)统计数据，以及关于巴拿马领水的港口等待和服务时间及运河时间表等增效措施的数据；

(c) 加拿大统计局和巴西地理和统计研究所的项目，利用开放来源渠道(数据收集、标准化、计算)，估计加拿大温哥华的城市建成区中供所有人公共使用的开放空间的平均份额，按性别、年龄和残疾状况分列(可持续发展目标指标 11.7.1)。

19. 南非主持的全球平台财务委员会旨在为全球平台制定可持续的商业模式和营销计划，以便管控成本和筹集资金。2021 年，大幅降低了每月费用，特别是降低了自动识别系统相关服务的费用。这些费用的降低为 2021 年全年的供资提供

¹⁴ 见 <https://statswiki.unece.org/display/hlgbas/High-Level+Group+for+the+Modernisation+of+Official+Statistics> 和 <https://statswiki.unece.org/display/hlgbas/Timetable+and+Documents+HLG2021>。

了稳定基础，预计 2022 年供资充足。通过宣传该平台和吸引新的资金，将有可能扩大在全球平台上开展的活动，以提供更多更好的服务，并鼓励更多国家的统计局参与使用大数据和数据科学。

20. 财务委员会迄今已做了几个宣传视频、一次推销介绍和一本小册子，并正在与各工作组和区域中心密切合作，以制作更多的宣传视频，并确定哪些项目和培训活动需要资金。接下来，财务委员会将与发展合作界、慈善基金会和私营部门接触，讨论筹资和伙伴关系机会。

21. 为了推进环境经济核算和可持续性方面的工作，专家委员会支持为全球平台环境经济核算体系创建一个环境和可持续性人工智能部门中心。该中心是一个协作型和以行动为导向的知识、技术和创新中心，将把创新技术和使用人工智能和大数据的数据科学方法结合在一起，并提供一个急需的平台，以促进环境经济核算和可持续性领域的数据和模型的互操作性。环境经济核算体系的环境和可持续性人工智能部门中心是统计司、美国地质测量局、巴斯克气候变化中心和多诺斯蒂亚国际物理中心之间的合作成果，设在西班牙毕尔巴鄂的巴斯克气候变化中心。该部门中心向官方统计使用大数据和数据科学专家委员会以及环境经济核算专家委员会报告工作。

22. 部门中心将与这两个专家委员会合作，促进和支持以下活动：

(a) 制定数据科学的方法，以支持环境和可持续性领域的新统计方式和方法；

(b) 为研究人员、统计人员和数据科学家开展能力建设和培训；

(c) 进一步开发和维护环境经济核算体系的环境和可持续性人工智能应用系统以及将其作为全球平台一部分部署所需的环境和可持续性人工智能平台组件；

(d) 进一步发展环境经济核算体系的环境和可持续性人工智能应用系统，将其作为官方统计界及其利益攸关方的数据、方法和学习资源的权威来源；

(e) 实施互操作性战略，描述所有利益攸关方(数据提供方、研究和模型开发方以及平台托管方)的作用和责任，以支持在全球实施环境经济核算体系生态系统核算的标准。

C. 支持全球平台的大数据区域中心

23. 为了便利各国统计局参加培训讲习班和项目活动，2020 年和 2021 年在巴西、中国、卢旺达和阿拉伯联合酋长国创建了四个大数据区域中心。这些中心将在区域一级汇集官方统计师界，使语言、文化和情况相似的国家能够合作开展利用大数据和数据科学估计统计值和可持续发展目标指标的联合项目。

24. 2021 年 4 月，统计司与巴西地理和统计研究所签署了关于巴西区域中心的谅解备忘录。随后设立了一个咨询委员会和各技术工作组，并制定了 2021-2022 年度初步工作方案，其中包括评估拉丁美洲和加勒比地区正在进行的关于大数据和数据科学的工作，评估方式为网页抓取、就将大数据用于官方统计与各国统计

局进行协商以及组织网络研讨会和讲习班。2021年11月23日，拉丁美洲和加勒比经济委员会美洲统计会议宣布该中心开始运作。

25. 中国区域中心于2020年12月7日在杭州举行的仪式上正式启动。该中心由中国国家统计局管理，竭力利用大数据和数据科学开发新的统计方法、技术和管理方式，并为本国、本区域和全球统计人员和数据科学家提供能力建设、培训和分享最佳做法的平台。自中心成立以来，已举办两次大数据研讨会(2020年12月和2021年9月)；召开该区域中心国际咨询委员会第一次会议，14名国际和本国成员出席；开展了以扫描仪数据、遥感数据、移动通信数据、行政记录和电子商务平台交易数据为中心的创新数据解决方案研究，并在一些领域取得了进展和成果；确定了应用系统的开发框架，包括门户网站和数据融合计算平台；与数据科学企业和研究机构开展沟通，努力创建合作机制。此外，中国国家统计局表示希望，在摆脱疫情恢复正常后，统计界的许多专家能访问杭州的区域中心，开展联合项目和研究，并在这方面邀请各方参加将于2023年4月在杭州举行的联合国世界数据论坛。

26. 非洲区域中心是2020年3月通过经济和社会事务部、卢旺达政府和非洲经济委员会(非洲经委会)之间的谅解备忘录设立的，由卢旺达国家统计局与非洲经委会共同管理。其主要目的是推动非洲大陆官方统计和可持续发展目标指标统计的创新。2021年，卢旺达国家统计局和非洲经委会进一步确立了中心的治理结构，并制定了数据科学和大数据路线图，其中包括研究项目、能力建设活动和指导方案。例如，为非洲国家统计系统的高级管理人员举办了数据科学领导课程，还组织了Python编程讲习班，有11个国家统计局参加。计划开展更多的能力建设活动，并在全球平台上开展合作。

27. 阿拉伯联合酋长国的区域中心也成立于2020年3月，是通过经济和社会事务部与阿拉伯联合酋长国政府之间的谅解备忘录设立的，由阿拉伯联合酋长国联邦竞争力和统计中心管理。该区域中心的主要目标是助力使用大数据和数据科学的项目，分享关于新开发的方法、算法和工具的知识，并为中东和北非区域的官方统计师界提供培训。

28. 官方统计使用大数据和数据科学专家委员会和联邦竞争力和统计中心将于2022年1月在2020年世博会上举办为期三天的“动员大数据和数据科学为可持续发展目标服务”的活动。¹⁵ 活动包括在阿拉伯联合酋长国启动中东和北非区域中心的仪式，以及全体会议、对话和炉边谈话，与专家委员会各工作组和各区域中心举行实践会议，并在“青年圈”举行主题为“COVID-19对人们生活的影响”的特别会议。

¹⁵ 见 <https://www.expo2020dubai.com/>。

三. 今后工作的方向和下一步行动

29. 将大数据和数据科学纳入各国统计局日常工作的主流被视为专家委员会今后工作方向的主要驱动因素。这些努力包括各工作组与现有各专题专家组之间加强合作，包括共同举行会议或就联合产出进行合作。已经有一些例子，例如扫描仪数据工作组与欧盟委员会价格统计工作组之间的合作，以及地球观测数据工作组与联合国粮食安全、农业和农村统计专家委员会之间的合作。此外，隐私保护技术工作组正在与官方统计现代化高级别小组就输入内容隐私保护项目密切合作，记录实际用例的应用情况。主流化工作还包括将大数据和数据科学课程纳入国家统计局的常规培训课程，并创建一个专家委员会成员与其他群体之间积极沟通交流的平台。

30. 作为下一步工作，专家委员会打算根据与主要政策议程或新出现的问题有关的战略考虑，更好地确定其各工作组、全球平台和各区域中心的案例研究、项目、活动、方法和应用的优先次序。例如，可以与 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(见 E/CN.3/2022/4)和可持续发展目标各项指标机构间专家组(见 E/CN.3/2022/2)协商确定优先事项。此外，专家委员会希望创建一个由各国统计局和各区域中心数据科学领域负责人组成的网络，以制定就共同目标、社会效益和开放源码工具使用问题与私营部门接触的战略办法，加强与地理空间信息界的合作，特别是通过在各工作组的工作中应用全球统计地理空间框架¹⁶ 加强合作，并组织更大规模的关于大数据和数据科学的活动，以促进与其他界别的合作。

31. 专家委员会力求继续开展其正在进行的能力发展工作，其中包括制定和推广成熟度矩阵、能力框架、培训课程和目录以及个性化学习系统。作为下一步工作，建议在统计界内创建一个全球数据科学培训师网络，并增强区域中心向各国统计局提供培训、指导、辅导、建立伙伴关系和支持执行联合项目的的能力。最后，最近设立的国际数据科学指导方案被认为非常成功，建议扩大该方案，吸收所有工作组专家。

四. 有待统计委员会采取的行动

32. 请统计委员会：

- (a) 确认并表示支持各工作组、联合国全球平台和各区域中心所做的工作；
- (b) 支持官方统计使用大数据和数据科学专家委员会今后工作的拟议方向，包括通过以下方式提供支持：
 - (一) 将大数据和数据科学纳入各国统计局日常工作的主流，并将相关课程纳入各国统计局的定期培训课程；

¹⁶ 见 https://unstats.un.org/unsd/statcom/51st-session/documents/The_GSGF-E.pdf。

- (二) 根据与主要政策议程或新出现的问题有关的战略考虑，确定案例研究、项目、活动、方法和应用的优先次序；
 - (三) 创建一个由各国统计局数据科学领域负责人和各区域中心负责人组成的网络；
 - (四) 制定就共同目标、社会公益效益和开放源码工具使用问题与私营部门接触的战略办法；
 - (五) 加强与地理空间信息界的协作；
- (c) 支持大数据和数据科学能力发展方案的方向，包括通过以下方式提供支持：
- (一) 在统计界内创建一个全球数据科学培训员网络；
 - (二) 扩大国际数据科学指导方案，吸收所有工作组专家；
 - (三) 增强区域中心向各国统计局提供培训、指导、辅导、建立伙伴关系和支持执行联合项目的能力；
- (d) 支持为支持全球平台的环境经济核算体系创建一个环境和可持续性人工智能部门中心，以此促进环境经济核算和可持续性领域数据和模型的互操作性，并在官方统计使用大数据和数据科学专家委员会和环境经济核算专家委员会的指导下，为研究人员、统计人员和数据科学家开展这方面的能力建设和培训；
- (e) 支持创建联合国加强隐私技术实验室，以展示将保护隐私技术用于官方统计的价值，了解实际挑战，并传播隐私技术知识。
-