

# DỮ LIỆU LỚN VÀ HIỆN ĐẠI HÓA HỆ THỐNG THỐNG KÊ

*Báo cáo của Tổng thư ký tại Phiên họp thứ 45 của Ủy ban Thống kê Liên hợp quốc*

## I. Giới thiệu

1. Chủ đề phiên họp vào Thứ 6 là các vấn đề đang nổi trong năm 2013 “Dữ liệu lớn cho chính sách, phát triển và thống kê nhà nước”<sup>21</sup>. Các diễn giả đến từ khu vực tư nhân cùng với các nhà thống kê trưởng trình bày quan điểm của họ về sự liên quan của dữ liệu lớn (Big data) đến hoạch định chính sách và vai trò của hệ thống thống kê quốc gia giả định liên quan với việc khai thác các nguồn thông tin mới. Các nhà thống kê trưởng của Australia chủ trì các phiên họp buổi chiều. Các kết luận rút ra ở sự kiện này là những dữ liệu lớn tạo thành một nguồn thông tin không thể bỏ qua của các nhà thống kê nhà nước và các nhà thống kê nhà nước phải tổ chức và hành động nhanh chóng để khai thác các khả năng và khai thác có hiệu quả.

2. Tại phiên họp 44 của Ủy ban Thống kê Liên hợp quốc (UNSD), đại diện của Úc được yêu cầu đánh giá việc sử dụng Big data cho thống kê nhà nước, và sẽ được trình lên tại phiên họp thứ 45. Báo cáo hiện tại đã đáp ứng yêu cầu đó.

3. Như là kết quả sự lan tỏa của việc sử dụng các thiết bị điện tử và thông tin kỹ thuật số xung quanh chúng ta, đã dẫn đến một sự thay đổi cơ bản trong bản chất của dữ liệu, mà bây giờ được tạo ra liên tục và với số lượng rất lớn, và cái này chính là dữ liệu lớn. Chúng có những đặc điểm rất riêng để phân biệt từ nguồn dữ liệu thông thường. Các dữ liệu từ các nguồn mới phân phối ở mức độ cao và có cấu trúc lỏng lẻo, khối lượng lớn và thường có sẵn trong thời gian thực. Big data là những nguồn dữ liệu có thể được mô tả như: “khối lượng lớn, tốc độ cao, loại dữ liệu đòi hỏi chi phí-hiệu quả, hình thức sáng tạo để nâng cao sự hiểu biết và quyết định”.

4. Trong thời đại suy giảm sự hưởng ứng đối với các cuộc điều tra quốc gia hộ gia đình và doanh nghiệp, Big data có thể cung cấp bằng chứng cho việc hoạch định chính sách ở thời gian thực trong các lĩnh vực như giá cả, việc làm, doanh thu kinh doanh, và nhân khẩu học<sup>22</sup>. Big data có tiềm năng để sản xuất số liệu thống kê liên quan và kịp thời hơn so với các nguồn dữ liệu truyền thống của số liệu thống kê nhà nước, chẳng hạn như khảo sát và các nguồn dữ liệu hành chính. Hầu hết các nguồn

<sup>21</sup> Hội thảo, sự kiện bên lề tại kỳ họp 44 của Ủy ban Thống kê Liên hợp quốc, thứ 6, ngày 22/2/2013, xem tại [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>22</sup> Tổ chức Hợp tác Kinh tế và Phát triển (OECD), “Tìm hiểu các hướng đổi mới dữ liệu như là một nguồn tăng trưởng mới: lập bản đồ các vấn đề chính sách đưa ra bởi dữ liệu lớn” (DSTI / ICCP (2012) 9 / FINAL), năm 2013.

Big data có trong khu vực tư nhân, và các đạo luật quy định cho phép sử dụng Big data cho mục đích thống kê nhà nước vẫn chưa được ban hành ở hầu hết các quốc gia. Vì vậy, mở rộng đối thoại hơn nữa với các bên liên quan là rất cần thiết nhằm xem xét không chỉ về pháp lý, bảo mật và các vấn đề nhận thức, mà còn là đối tượng của việc sử dụng kịp thời, đáng tin cậy và đáng tin cậy của Big data. Bằng cách kết hợp nguồn Big data vào sản xuất số liệu thống kê nhà nước, các Cơ quan thống kê quốc gia, khu vực và quốc tế để có vị thế cao hơn để có được số liệu thống kê nhà nước về kinh tế, xã hội và môi trường trong việc cải thiện tính kịp thời và chi phí-hiệu quả, và giảm gánh nặng tài nguyên.

5. Với những cải tiến mới trong công nghệ và sự bùng nổ về số lượng và sự đa dạng của thông tin thời gian thực, Tổng thư ký Liên hợp quốc đưa ra, trong năm 2009, các sáng kiến Trung tâm động đất toàn cầu (Global Pulse) để thúc đẩy sự đổi mới trong dữ liệu kỹ thuật số, thu thập và phân tích dữ liệu nhanh chóng để giúp người ra quyết định đạt được một sự hiểu biết thời gian thực của các cuộc khủng hoảng nào ảnh hưởng như thế đến quần thể dễ bị tổn thương. Trong khi nhấn mạnh rằng Big data không phải là một loại thuốc có thể chữa bách bệnh, nhưng Global Pulse đưa ra là “dữ liệu lớn thiết lập một cơ hội lịch sử để thúc đẩy khả năng chung của chúng ta để hỗ trợ và bảo vệ cộng đồng con người bằng sự hiểu biết các thông tin được sản xuất ở dạng kỹ thuật số”<sup>23</sup>.

6. Trong các khuyến nghị của Ban cấp cao về nhân vật nổi tiếng ở Chương trình nghị sự phát triển năm 2015 trong báo cáo “Một quan hệ đối tác toàn cầu mới: xóa đói giảm nghèo và chuyển đổi nền kinh tế thông qua phát triển bền vững”, có một cuộc cách mạng dữ liệu. Big data được ghi nhận như là một bộ phận quan trọng của cuộc cách mạng dữ liệu, trong đó có thể có tiềm năng đóng góp vào việc cải thiện một số khía cạnh về chất lượng của số liệu thống kê, chẳng hạn như tính kịp thời và đầy đủ, mà không ảnh hưởng đến sự thích hợp, công bằng và lành mạnh về phương pháp thống kê. Các báo cáo ban cao cấp nói rằng:

Dữ liệu tốt hơn và các số liệu thống kê sẽ giúp Chính phủ theo dõi tiến độ và đảm bảo rằng các quyết định của họ là dựa trên bằng chứng; làm tăng trách nhiệm giải trình. Đây không chỉ là vấn đề của các Chính phủ. Tổ chức quốc tế, tổ chức xã hội dân sự và khu vực tư nhân cũng cần được tham gia. Một cuộc cách mạng dữ liệu thực sự sẽ diễn ra trên nền tảng hiện có và các nguồn dữ liệu mới để tích hợp đầy đủ các số liệu thống kê vào việc ra quyết định, thúc đẩy tiếp cận mở, sử dụng dữ liệu và đảm bảo tăng cường hỗ trợ cho các hệ thống thống kê (Phần 4, đoạn 24).

<sup>23</sup> Global Pulse, “Dữ liệu lớn cho phát triển: những thách thức và cơ hội”, tháng 5/2012.

7. Trong bối cảnh rộng lớn hơn, các trường hợp đã được thực hiện mà dữ liệu lớn có khả năng thay đổi nhiều khía cạnh của xã hội. Theo Hiệp hội Công nghệ Mỹ, ghi trong một báo cáo gần đây rằng “dữ liệu lớn còn có tiềm năng làm thay đổi Chính phủ và xã hội”<sup>24</sup>. Ấn trong các khối lượng lớn, đa dạng và tốc độ của dữ liệu được sản xuất ngày nay là những thông tin, sự kiện, các mối quan hệ, các chỉ số và lời gợi ý mới, mà một trong số đó có thể không được phát hiện trong quá khứ, hoặc chỉ đơn giản là trước đây không tồn tại<sup>25</sup>. Các thông tin mới này, giành được hiệu quả, quản lý, và phân tích, có khả năng đề cao một cách sâu sắc tính hiệu quả của Chính phủ”<sup>26</sup>.

8. Phần II của báo cáo này đưa ra cái nhìn tổng quan về các hoạt động liên quan đến Big data và hiện đại hóa hệ thống thống kê. Phần III cung cấp tóm tắt các nguồn và thách thức của Big data. Mục IV trình bày kết quả việc đánh giá toàn cầu về việc sử dụng Big data cho thống kê nhà nước, được tiến hành bởi các Phòng Thống kê từ tháng 7-10/2013. Phần V cung cấp một số nghiên cứu trường hợp việc sử dụng Big data. Kết luận báo cáo này đưa ra đề nghị trong tương lai, đó là phù hợp với các sáng kiến của các tổ chức đối tác và trong đó nhấn mạnh các vấn đề liên quan đến việc sử dụng Big data cho thống kê nhà nước ở các nước đang phát triển và các Chương trình nghị sự phát triển năm 2015.

## **II. Tổng quan về các hoạt động liên quan đến Big data trong thời gian gần đây**

9. Ngày càng có nhiều Chính phủ của các quốc gia đã thừa nhận tầm quan trọng của Big data và đã thành lập các cộng đồng, nhóm nghiên cứu việc sử dụng và những tác động của Big data. Cộng đồng thống kê đang dần nhận ra một sự thay đổi lớn trong mô hình quá trình sản xuất thông tin hiện nay. Viện Thống kê quốc gia Ý thừa nhận rằng “để khai thác thông tin (có tiềm năng rất lớn) từ kho báu Big data, thì Viện thống kê quốc gia sẽ phải có những thay đổi tích cực để khai thác được kho báu đó”<sup>27</sup>. Thống kê Hà Lan thừa nhận rằng thống kê nhà nước rất khó khăn khi tiếp cận, khám phá những cơ hội từ Big data<sup>28</sup>.

<sup>24</sup> Mở đầu phần tóm tắt của Hiệp hội Công nghệ Mỹ trong Báo cáo mang tên “Làm rõ dữ liệu lớn: một hướng dẫn thực tế để chuyển đổi kinh doanh của Chính phủ” (Washington, tháng 10/2012).

<sup>25</sup> Như trên

<sup>26</sup> Như trên

<sup>27</sup> Monica Scannapieco, Antonino Virgillito và Diego Zardetto, "Đặt dữ liệu lớn trong số liệu thống kê chính thức: có phải là một thách thức lớn không?", ngày 21/12/2012.

<sup>28</sup> Piet DAAS và Mark van der Loo, "dữ liệu lớn (và số liệu thống kê chính thức)", bài trình bày tại Hội nghị về quản lý hệ thống thông tin thống kê (MSIs 2013), tại Paris và Bangkok, 23-ngày 25/4/2013.

10. Ủy ban Kinh tế Châu Âu (ECE), Ủy ban thống kê châu Âu (Eurostat), Tổ chức Hợp tác Kinh tế và Phát triển (OECD) và Ủy ban Kinh tế và Xã hội châu Á và Thái Bình Dương (ESCAP) đã tổ chức một cuộc họp về Quản lý hệ thống thông tin thống kê được tổ chức đồng thời tại Paris và Bangkok từ ngày 23-25/4/2013<sup>29</sup>. Kết luận chính tại cuộc họp là các tổ chức thống kê phải biết kết hợp các nguồn lực để đối phó với các vấn đề chung liên quan đến việc sử dụng Big data và hướng tới các giải pháp phổ biến, như là sản xuất số liệu thống kê trong thời gian ngắn, cách tiếp cận đa ngành với Big data, và thống nhất chung về phân loại khác nhau của Big data.

11. Đại hội Thống kê Thế giới lần thứ 59 do Viện Thống kê Quốc tế (ISI) tổ chức tại Hồng Kông, Trung Quốc từ ngày 25-30/8/2013, đã dành nhiều sự quan tâm đến chủ đề Big data, với một phiên hội nghị toàn thể về chủ đề Big data; một phiên họp đặc biệt với chủ đề về tiềm năng của mạng, Big data và dữ liệu thứ cấp cho thống kê nhà nước; và một sự kiện được truyền hình vệ tinh về Big data và Toán thống kê. Tại phiên họp đặc biệt, John Dunne làm việc tại Văn phòng Thống kê quốc gia Ireland, đã trình bày một bài báo cáo có tiêu đề "Big data đến sớm... để Viện thống kê quốc gia gần bạn"<sup>30</sup>, trong đó mô tả một số nguồn Big data, như dữ liệu về chuyển vùng điện thoại di động, tiêu thụ điện hoặc giao dịch điện tử, và trình bày cách thức xử lý khối lượng lớn dữ liệu kịp thời thông qua công nghệ điện toán đám mây bởi nhà cung cấp bên thứ ba. Big data cũng là một chủ đề chính trong chương trình hội nghị tại Đại hội Thống kê Thế giới lần thứ 60 được tổ chức từ ngày 26-31/7/2015 tại Rio de Janeiro, Brazil<sup>31</sup>. Đại hội do ISI tổ chức là nơi tạo ra những cơ hội cho sự hợp tác và trao đổi kiến thức giữa các nhà thống kê chính thức, các học giả và nhà kinh doanh.

12. Eurostat đang điều tra tiềm năng sử dụng Big data cho thống kê nhà nước trong các lĩnh vực như thống kê giá (sử dụng dữ liệu mạng Internet thu thập thông tin về giá) và số liệu thống kê công nghệ thông tin và truyền thông (ICT). Hơn nữa, cuộc họp thường niên của chủ tịch Viện Thống kê quốc gia châu Âu vào tháng 09/2013 tại Scheveningen, Hà Lan đã dành một phiên họp về chủ đề Big data<sup>32</sup>. Điều này được ghi lại trong Bản kết luận cuộc họp về Big data và thống kê nhà nước (thông qua bởi Eurostat vào ngày 27/9/2013), trong đó khuyến khích các thành viên trong hệ thống thống kê châu Âu phát triển một chiến lược Big data, chia sẻ kinh nghiệm và hợp tác trong hệ thống thống kê châu Âu và xa hơn nữa là một kế hoạch hành động và lộ

<sup>29</sup> Xem tại <http://www1.unece.org/stat/platform/display/msis/MSIS+2013>.

<sup>30</sup> Có sẵn tại <http://www.statistics.gov.hk/wsc/STS018-P3-S.pdf>.

<sup>31</sup> Xem tại <http://www.isi2015.ibge.gov.br>.

<sup>32</sup> Xem tại <http://www.cbs-events.nl/dgins2013/>.

trình cần được thông qua vào giữa năm 2014 được tích hợp vào chương trình làm việc của Eurostat.

13. Big data là một phần trong các cuộc thảo luận tại Hội nghị của ECE về thu thập dữ liệu thống kê được tổ chức tại Geneva từ ngày 25-27/09/2013<sup>33</sup>. Tại Hội thảo, Peter Struijs và Piet DAAS, Cơ quan thống kê Hà Lan đã trình bày báo cáo (Số 31) mang tên "Big data, tác động lớn?". Trong đó đề cập cụ thể những tác động của Big data đến cộng đồng thống kê và thống kê nhà nước. Các tác động cụ thể này sẽ dần trở nên rõ ràng, nhưng một số thì đã được nhìn thấy hoặc có thể dự đoán. Hệ thống thống kê nhà nước sẽ chịu sự cạnh tranh nhiều hơn từ các hệ thống ngoài thống kê nhà nước. Hệ thống thống kê nhà nước sẽ cần phải điều chỉnh lại phương pháp thực hiện và tìm một hướng đi mới, đó là sử dụng các khả năng mới được cung cấp bởi Big data. Điều này có thể đòi hỏi thay đổi định hướng mô hình từ một cuộc điều tra thống kê hiện nay cho đến dữ liệu thứ cấp nhiều hơn là tập trung vào việc định hướng cách tiếp cận dựa trên mô hình điều tra chuẩn mực. Trong kỷ nguyên Big data, Cơ quan thống kê phải kết hợp với các nhà khoa học dữ liệu trong đội ngũ nguồn nhân lực của mình như là một điều cần thiết chứ không phải là một ngoại lệ, và đồng thời công nhận những cơ hội và thách thức do Big data mang lại, cộng đồng thống kê nhà nước sẽ đi tắt đón đầu với vai trò của mình trong việc cung cấp thông tin từ các nguồn dữ liệu mới.

14. Tại cuộc họp lần thứ hai được tổ chức vào ngày 21-22/10/2013 trong Hội nghị các nhà thống kê châu Âu, cơ quan chủ quản ECE rà soát kỹ chủ đề về Big data<sup>34</sup>. Các khuyến nghị chính dựa trên báo cáo tổng quan về các lĩnh vực ưu tiên quan trọng của Big data cần được xác định và giải quyết như một hoạt động hợp tác của cộng đồng thống kê quốc tế, và cần thiết lập một cơ chế chia sẻ thông tin về kiến thức và kinh nghiệm của việc sử dụng này. Tổ chức cũng đã chấp thuận đề xuất dự án<sup>35</sup> vai trò của Big data, với mục tiêu:

(a) Xác định các khả năng chính mà Big data có thể làm được, và thực hiện hướng dẫn cho các tổ chức thống kê nhà nước, phát triển hợp tác đối phó với các vấn đề chiến lược và phương pháp mà Big data đặt ra cho thống kê nhà nước;

(b) Chứng minh tính khả thi và hiệu quả của sản xuất số liệu thống kê trong hệ thống thống kê quốc gia thông qua số liệu thống kê "chính thống" và số liệu thống kê sử dụng nguồn dữ liệu lớn, và khả năng tiếp cận Big data dựa trên bối cảnh của mỗi quốc gia;

<sup>33</sup> Xem tại <http://www.unecce.org/stats/documents/2013.09.coll.html>.

<sup>34</sup> Xem tại See ECE/CES/BUR/2013/OCT/2.

<sup>35</sup> Dự án này là một phần của sáng kiến về hiện đại hóa sản xuất thống kê thuộc Nhóm cấp cao về Hiện đại hoá sản xuất và dịch vụ thống kê phối hợp bởi ECE.

(c) Tạo thuận lợi cho việc chia sẻ kiến thức, chuyên môn, công cụ và phương pháp giữa các tổ chức để sản xuất số liệu thống kê từ các nguồn Big data.

15. Big data và hiện đại hóa hệ thống thống kê cũng có mặt trong Chương trình nghị sự của một số sự kiện gần đây ở châu Á. Phòng Thống kê và Cục Thống kê Quốc gia Trung Quốc với sự hỗ trợ của ESCAP đã tổ chức Hội thảo quốc tế về Hiện đại hóa số liệu thống kê nhà nước từ ngày 24-26/10/2013 tại Thiên Tân, Trung Quốc. Hội thảo này tập trung vào việc hiện đại hóa các tổ chức thống kê và bao gồm một số bài trình bày về Big data. Hội thảo lần thứ 11 về đào tạo quản lý dành cho những người đứng đầu các Cơ quan thống kê quốc gia ở khu vực châu Á - Thái Bình Dương<sup>36</sup> do Viện Thống kê khu vực châu Á - Thái Bình Dương (SIAP) tổ chức từ ngày 21-22/11/2013 tại Chiba, Nhật Bản, tập trung vào các sáng kiến cách mạng dữ liệu và vai trò tiềm năng của Big data trong Chương trình nghị sự phát triển 2015.

### III. Big data: Nguồn dữ liệu và thách thức

16. Một nhóm nghiên cứu không chính thức được thành lập bởi Nhóm cấp cao về Hiện đại hoá sản xuất và dịch vụ thống kê được điều phối bởi ECE, nhóm nghiên cứu đã tóm tắt các nguồn Big data và những thách thức liên quan đến việc sử dụng các Big data trong một báo cáo có tên "Big data có ý nghĩa như thế nào đối với thống kê nhà nước"(10/03/2013). Báo cáo được trình bày tại phiên họp toàn thể lần thứ 61 trong Hội nghị các nhà thống kê châu Âu tổ chức từ ngày 10-12/06/2013. Các điểm mới của báo cáo này sẽ được thảo luận ở phần sau đây. Báo cáo cũng đề cập đến những vấn đề khó khăn và đưa ra những đề xuất, đã được sử dụng đưa vào các câu hỏi trong các điều tra toàn cầu của Phòng Thống kê Liên hợp quốc tiến hành trong Quý3/2013 (xem phần IV).

17. Ngày nay, dữ liệu được tạo ra tự động và liên tục ở dạng kỹ thuật số với nhiều cách khác nhau. Các nguồn dữ liệu khác nhau có tiềm năng lớn đáng quan tâm để sử dụng trong thống kê nhà nước với mục đích đo lường xu hướng xã hội, môi trường, tài chính hoặc kinh tế với độ chính xác cao hơn và kịp thời hơn. Nói chung, các nguồn Big data có thể được phân loại như sau:

- Nguồn hồ sơ hành chính, như hồ sơ y tế điện tử, hồ sơ khám bệnh, hồ sơ bảo hiểm, hồ sơ ngân hàng, v.v...

- Các nguồn thương mại hoặc các giao dịch phát sinh từ các giao dịch giữa hai thực thể, ví dụ, các giao dịch thẻ tín dụng và các giao dịch trực tuyến (bao gồm từ các thiết bị di động)

<sup>36</sup> Xem tại [http://www.unsiap.or.jp/training/6\\_Leading\\_MQS/6\\_ms11.php](http://www.unsiap.or.jp/training/6_Leading_MQS/6_ms11.php).

- Nguồn dữ liệu từ các thiết bị cảm biến, ví dụ, ảnh vệ tinh, cảm biến đường và các cảm biến khí hậu
- Nguồn thiết bị theo dõi, ví dụ, theo dõi dữ liệu từ điện thoại di động và các Hệ thống định vị toàn cầu (GPS)
- Các nguồn dữ liệu hành vi, ví dụ, tìm kiếm trực tuyến (về một sản phẩm, một dịch vụ hoặc bất kỳ loại thông tin khác), v.v...
- Các nguồn dữ liệu chung, ví dụ, các ý kiến nhận xét trên phương tiện truyền thông xã hội.

18. Dữ liệu hành chính là một trong những nguồn dữ liệu chính để sản xuất số liệu thống kê nhà nước của hệ thống thống kê quốc gia. Theo truyền thống, những dữ liệu nhận được từ các cơ quan hành chính là những dữ liệu có cấu trúc, sau đó được xử lý, lưu trữ, quản lý và sử dụng bởi các Viện nghiên cứu thống kê. Dữ liệu hành chính hiện nay không phải là một nguồn Big data, nhưng nó có thể trở thành "lớn" khi vận tốc và khối lượng tăng - ví dụ, khi các Viện thống kê bắt đầu sử dụng dữ liệu hành chính nhiều hơn bằng cách thu thập chúng trong thời gian thực, hoặc theo ngày hay theo tuần, thay vì bình thường hiện nay hay làm đó là thu thập một năm một lần hoặc mỗi tháng một lần.

19. Việc sử dụng Big data trong thống kê nhà nước gặp nhiều thách thức, chủ yếu gồm các loại sau:

- Luật pháp, liên quan đến việc tiếp cận và sử dụng các dữ liệu
- Bảo mật, liên quan đến việc quản lý dữ liệu công khai tin cậy và chấp nhận sử dụng các dữ liệu và kết nối với các nguồn dữ liệu khác
- Tài chính, liên quan đến chi phí của nguồn dữ liệu so với lợi ích tiềm năng mà chúng mang lại
- Quản lý, như là, chính sách, các văn bản về quản lý và bảo vệ dữ liệu
- Phương pháp, đề cập về chất lượng dữ liệu và tính phù hợp của các phương pháp thống kê
- Công nghệ, liên quan đến các vấn đề về công nghệ thông tin.

#### **IV. Kết quả từ cuộc điều tra toàn cầu về sử dụng Big data**

20. Sử dụng các thông tin mô tả trong báo cáo của nhóm công tác ECE, Phòng Thống kê Liên hợp quốc xây dựng bảng câu hỏi về việc sử dụng các Big data cho thống kê nhà nước. Bảng hỏi chia thành ba phần chính: nguồn dữ liệu, thách thức và phạm vi sử dụng. Các câu hỏi về nguồn dữ liệu và thách thức bao gồm tất cả những điểm nêu trên, cùng với một số câu hỏi được xây dựng trên khía cạnh quản lý, ví dụ

“vấn đề quản lý, chẳng hạn như tôn trọng các chính sách và các quy định mới, phát triển nguồn nhân lực với các điều kiện cần thiết về kỹ năng và chuyên môn”. Đánh giá toàn cầu này được thực hiện với mục tiêu cung cấp thông tin về các vấn đề ưu tiên phát triển quốc gia và kinh nghiệm liên quan đến việc sử dụng Big data cho thống kê nhà nước trong hiện tại hoặc tương lai.

21. Các câu hỏi đã được gửi đến các Cơ quan thống kê quốc gia của hơn 200 nền kinh tế trong giai đoạn từ tháng 7-9/2013, các câu hỏi đã được gửi đi bằng tiếng Anh vào tháng 7/2013, tiếng Tây Ban Nha vào tháng 8/2013 và tiếng Pháp vào tháng 9/2013. Hoặc có thể trả lời trực tuyến ở định dạng PDF. Hạn cuối là ngày 02/11/2013, cuộc khảo sát này nhận được 107 bảng kết quả trả lời. Các kết quả này được ECE dùng như một tài liệu báo cáo và bản tóm tắt các kết quả này được đưa ra dưới đây.

### **A. Nguồn Big data**

22. Các phân đoạn chính xác của các câu hỏi liên quan đến các nguồn Big data là: “Hãy chỉ ra những nguồn Big data sau đây có thể sẽ được sử dụng trong 12 tháng của cơ quan bạn hoặc các cơ quan khác là một phần của hệ thống thống kê quốc gia”. Nếu câu trả lời là có, thì cần có thêm một vài giải thích cụ thể. Trong khi hơn 50% câu trả lời là có khi sử dụng dữ liệu hành chính như là nguồn Big data, thì khoảng 25% cho rằng từ các nguồn dữ liệu khác. Một số quốc gia đã nêu vấn đề cho dù hồ sơ hành chính có là nguồn Big data hay không. Cụ thể hơn, dữ liệu hành chính là nền tảng của việc sản xuất nhiều sản phẩm thống kê, nhưng liệu có nên xem xét dữ liệu hành chính trong bối cảnh như là Big data thì vẫn là một điểm còn phải thảo luận.

23. Nhìn chung, nhiều nước trả lời có đối với việc sử dụng một hoặc nhiều các nguồn Big data. Tuy nhiên, kết quả phân tích cho thấy một số trong những câu trả lời đó không phù hợp với các định nghĩa của Big data, như trong báo cáo tài liệu tham khảo của nhóm công tác ECE. Để củng cố giá trị kết quả của cuộc điều tra này dựa trên Big data, thì các câu trả lời được xem xét kỹ lưỡng và thu hẹp vào một nhóm các lựa chọn gồm “hoạt động có liên quan”. Những câu trả lời tích cực có phải là tất cả các câu không có giải thích đi kèm, hoặc giải thích rất hạn chế hay giải thích không phù hợp. Kết quả là, thay vì con số ban đầu từ 25-50% cho những câu trả lời có, thì các số liệu đã thay đổi là gần 20% đối với các nguồn dữ liệu hành chính, khoảng 10% cho các nguồn giao dịch, mạng cảm biến và các thiết bị theo dõi, và gần 5% cho các nguồn hành vi và ý kiến. Kết quả chi tiết có tại website Phòng Thống kê Liên hợp quốc, với từng cột cho tất cả các nước có câu trả lời tích cực và các nước có “hoạt động có liên quan”.

24. Trong số các nguồn dữ liệu trích dẫn nhiều nhất của Big data là dữ liệu thuế và thuế giá trị gia tăng (VAT). Những nguồn khác như đăng ký kinh doanh, giao



dịch qua máy quét mã vạch ở siêu thị, thông tin liên lạc qua điện thoại, giao dịch thẻ tín dụng và dữ liệu thu phí. Một số quốc gia cũng đã công bố sử dụng cảm biến vệ tinh, hình ảnh vệ tinh để thiết kế điều tra. Ảnh vệ tinh cũng đã được nêu ra như là một nguồn để nghiên cứu thống kê năng suất mùa vụ. Cảm biến đường được sử dụng cho thống kê giao thông vận tải (các chỉ số lưu lượng giao thông ở Hà Lan) và thống kê môi trường (đo chất lượng không khí ở Anh và Bắc Ireland).

25. Một số ít quốc gia sử dụng dữ liệu theo dõi GPS. Ví dụ, theo dõi điện thoại di động được sử dụng để cải thiện thống kê du lịch ở Estonia là một ví dụ điển hình. Tại Canada, dữ liệu GPS từ xe tải, đường sắt, đường biển và vận tải hàng không được sử dụng để bổ sung số liệu thống kê vận tải. Trong khi nhiều quốc gia không sử dụng dữ liệu hành vi, thì ở một số quốc gia đã có kế hoạch để sử dụng nguồn dữ liệu này. Ví dụ, ở Ý, truy vấn Internet có thể sẽ được sử dụng (a) là thông tin hỗ trợ để giảm chi phí các cuộc điều tra, (b) cho mục đích dự báo và (c) để tăng chi tiết của các ước tính theo khu vực không gian địa lý (đặc biệt là đối với các chỉ số thị trường lao động).

26. Tương tự như vậy, rất ít quốc gia sử dụng các dữ liệu quan điểm hoặc các phương tiện truyền thông xã hội làm nguồn Big data. Ở Hà Lan một số nghiên cứu công bố công khai dựa trên các thông điệp truyền thông xã hội (xem trong phần IV). Tại Singapore, các nghiên cứu đang được tiến hành sử dụng nguồn Big data để phát hiện sớm các vấn đề về an toàn thực phẩm.

### **B. Những thách thức trong việc sử dụng Big data**

27. Các câu hỏi chính xác cho phần này là “Đây có phải là [vấn đề] một thách thức lớn đối với hệ thống thống kê quốc gia ở nước bạn?”. Trả lời không (không phải là một thách thức), không có ý kiến (sẽ không thảo luận chi tiết) hoặc có (là một thách thức). Đối với câu trả lời của các quốc gia là có thì hầu hết đều liên quan đến cả 6 thách thức khi sử dụng Big data, đối với trường hợp là không có ý kiến (chỉ ra rằng những vấn đề đó vẫn chưa được thảo luận kỹ lưỡng) và có một số lượng nhỏ các kết quả là không (đây không phải là một thách thức). Nhìn chung, các thách thức về phương pháp, công nghệ thông tin và quản lý được cho là phổ biến nhất, và những thách thức được quan tâm theo dõi chặt chẽ là những thách thức về pháp lý và sự riêng tư.

28. Người trả lời đồng ý rằng một trong những thách thức lớn sẽ được giải quyết là vấn đề phương pháp, và gợi ý rằng cần phải nghiên cứu nhiều hơn nữa trong lĩnh vực này. Vấn đề công nghệ thông tin cũng đã được chỉ ra như là một thách thức lớn ở nhiều nước bởi vì các hệ thống xử lý và phân tích hiện tại không đáp ứng được cho Big data. Bản chất của Big data (đặc điểm không chỉ khối lượng lớn mà còn vận tốc lớn và tính đa dạng) dẫn đến những thách thức lớn đối với sự

phát triển của các công cụ công nghệ thông tin cần thiết và kiến trúc công nghệ thông tin, đặc biệt là trong bối cảnh thế giới đang phát triển. Vấn đề quản lý như tôn trọng các chính sách và các quy định mới, và phát triển nguồn nhân lực với các thiết lập cần thiết về kỹ năng và chuyên môn nghiệp vụ được xem như là một thách thức lớn ở hầu hết các nước.

29. Việc sử dụng hệ thống các nguồn dữ liệu phi Chính phủ đặt ra một thách thức pháp lý cho Chính phủ trong bối cảnh những cơ hội sử dụng Big data. Trong khi một số quốc gia đã có luật để đảm bảo quyền truy cập vào cả hai nguồn dữ liệu Chính phủ và phi Chính phủ, thì việc thiếu các luật đảm bảo quyền truy cập vào các nguồn dữ liệu phi Chính phủ đã được trích dẫn là một thách thức ở một số nước. Vấn đề bảo mật cũng là mối quan tâm lớn ở nhiều nước liên quan đến việc sử dụng Big data. Việc bảo mật các dữ liệu cá nhân trong lĩnh vực này là rất quan trọng; do đó, đây là điều cần thiết để đạt được sự cân bằng tối ưu giữa việc bảo mật dữ liệu cá nhân và sử dụng cho mục đích thống kê của chúng ta.

30. Các vấn đề tài chính cũng được coi là thách thức lớn đối với một số quốc gia, đặc biệt là chi phí có được nguồn dữ liệu được coi là một trong những thách thức khi sử dụng Big data. Ngoài ra, các nguồn lực tài chính cần thiết để có được các trang thiết bị và công nghệ để xử lý được Big data có thể không đầy đủ hoặc thiếu. Cùng với đó, hiện nay các công ty tư nhân đang ngày càng nhận ra giá trị của dữ liệu mà họ nắm giữ rất có giá trị.

31. Thách thức lớn khác liên quan đến việc sử dụng các Big data của hệ thống thống kê quốc gia đã được đề xuất. Đặc biệt, vấn đề được nhiều người đưa ra là làm thế nào để tích hợp Big data vào hệ thống thống kê nhà nước. Làm thế nào nguồn dữ liệu mới này có thể được gắn vào Chương trình điều tra thống kê hiện hành? Hầu hết đều đồng ý rằng Big data sẽ không thay thế các số liệu thống kê nhà nước, nhưng chính xác hơn là nó sẽ đóng vai trò như là sự bổ sung, ví dụ, bằng cách cung cấp số liệu thống kê có sẵn rất nhanh chóng và các chỉ số thử nghiệm cho thống kê. Khi Big data tạo thành một khái niệm tương đối mới trong các lĩnh vực thống kê nhà nước, thì các nhà thống kê sẽ có nhu cầu để nghiên cứu.

### **C. Các lĩnh vực và phạm vi sử dụng**

32. Trong Phần 3 của cuộc điều tra toàn cầu, những người tham gia được hỏi về các lĩnh vực sử dụng (hoặc nghiên cứu và sử dụng) Big data trong 12 tháng. Các phạm vi của cuộc điều tra bao gồm “số liệu thống kê xã hội và nhân khẩu học”, “thống kê sinh tử và đăng ký hộ tịch”, “thống kê kinh tế và tài chính”, “thống kê giá”, “thống kê giao thông vận tải”, “thống kê môi trường” và “lĩnh vực khác của thống kê nhà nước”. Những câu trả lời có thể là không hoặc có, đối với câu trả lời là có thì cần được giải thích. Các phạm vi được sử dụng thường xuyên nhất là lĩnh vực

sử dụng là số liệu thống kê nhân khẩu học và xã hội và thống kê kinh tế và tài chính. Tuy nhiên, như trong trường hợp sử dụng các nguồn Big data, các câu trả lời có trong phần này của điều tra toàn cầu cần phải được đánh giá lại bằng cách xác định "các hoạt động có liên quan". Sau khi đánh giá lại, nó chỉ ra rằng các hoạt động có liên quan của việc sử dụng Big data trong lĩnh vực thống kê giá và thống kê kinh tế và tài chính. Đặc biệt, một số quốc gia trích dẫn việc sử dụng dữ liệu bằng máy quét nhờ kỹ thuật rút trích nội dung thông tin từ Web để sản xuất các chỉ số giá, thường xuyên cập nhật bổ sung chỉ số giá tiêu dùng tiêu chuẩn. Nhìn chung, tỷ lệ hoạt động có liên quan cho các phạm vi sử dụng là hơn 10% cho thống kê giá, thống kê kinh tế và tài chính và thống kê nhân khẩu học và xã hội, và khoảng 5% cho mỗi lĩnh vực khác.

33. Trong lĩnh vực thống kê nhân khẩu học và xã hội đã có đề cập đến cách mà dữ liệu của điện thoại di động có thể được sử dụng để theo dõi các hành vi di chuyển và du lịch, dữ liệu phương tiện truyền thông xã hội để đo lường mức độ niềm tin tiêu dùng, và việc sử dụng tiềm năng của Google phân tích xu hướng, v. v... Các ví dụ khác như hợp tác với các nhà khai thác điện thoại di động để tính toán thống kê di cư. Trong lĩnh vực thống kê kinh tế và tài chính, nghiên cứu việc sử dụng dữ liệu ảnh vệ tinh và bộ cảm biến mặt đất để ước lượng năng suất mùa vụ. Một số quốc gia đã giải thích cách sử dụng Big data trong lĩnh vực thống kê giao thông vận tải. Như dữ liệu chi tiết từ thuế đường bộ và cảm biến đường được sử dụng để thống kê lưu lượng giao thông.

### **V. Dữ liệu lớn và hiện đại hóa hệ thống thống kê**

34. Để đối phó với những thách thức được đề cập trong các câu trả lời cuộc khảo sát toàn cầu, hệ thống thống kê sẽ cần phải được hiện đại hóa. Có ý kiến cho rằng cần nghiên cứu thêm để khắc phục những khó khăn cản trở việc khai thác nguồn Big data. Do bản chất của Big data, các vấn đề về tính đại diện và sự bao phủ dân số cũng có thể là một vấn đề. Hơn nữa, sự đa dạng và tính tiềm năng tạm thời của dữ liệu (ví dụ, mọi người sẽ tiếp tục sử dụng Facebook trong bao lâu?). Vấn đề đặt ra từ quan điểm thống kê nhà nước, giả định về khả năng so sánh, tính liên tục, và tính chặt chẽ của số liệu thống kê đều nằm trong các yếu tố chất lượng chính. Những thay đổi trong phương pháp có thể bao gồm việc sử dụng các mô hình thường xuyên hơn và có thể đòi hỏi sự hợp tác chặt chẽ hơn với các học viện. Nhiều nghiên cứu và nghiên cứu thực nghiệm cần phải được tiến hành để khám phá những ứng dụng tiềm năng của Big data trong thống kê nhà nước, và nghiên cứu có thể cần phải trở thành một phần của quá trình sản xuất thống kê.

35. Các cơ sở hạ tầng công nghệ thông tin hiện có và khả năng cần phải được tăng cường để có thể hỗ trợ việc xử lý dữ liệu với khối lượng lớn và tốc độ cao từ nhiều nguồn khác nhau. Ngoài ra, chuẩn hóa các phương pháp và công cụ giữa các

Cơ quan thống kê quốc gia nên được xem xét. Có một nhu cầu để có được các công nghệ mới nhất (phần mềm, phần cứng và kỹ năng con người) để bám sát các nhu cầu công nghệ trong việc thu thập, xử lý và sử dụng Big data. Các cơ quan có thể phải xem xét công nghệ điện toán đám mây thay vì cố gắng để di chuyển một khối lượng lớn dữ liệu đến các máy chủ của riêng họ. Ngoài ra, giám sát các nguồn Big data để thực hiện các tính toán ban đầu và cung cấp dữ liệu tổng hợp hơn và có cấu trúc tốt hơn cho Cơ quan thống kê. Nguồn lực đầy đủ là rất cần thiết để xây dựng các cơ sở hạ tầng công nghệ thông tin nhằm có dữ liệu một cách thường xuyên, liên kết và phân tích và trình bày dữ liệu ở các định dạng thích hợp cho người sử dụng.

36. Cả sự phát triển và duy trì các kỹ năng cần thiết của nhân viên trong phân tích dữ liệu tiên tiến và tin học hóa sẽ là một thách thức lớn. Các nguồn dữ liệu lớn đòi hỏi những thay đổi trong việc tổ chức thu thập dữ liệu và xử lý dữ liệu. Hầu hết các Cơ quan thống kê quốc gia không có đội ngũ nhân viên có các kỹ năng thích hợp, do đó đòi hỏi phải xây dựng năng lực và đào tạo chuyên sâu nguồn nhân lực để có thể khai thác các nguồn dữ liệu. Các chuyên gia hiện nay, cơ bản dựa trên thống kê cổ điển, điều này là không đủ để đảm bảo xử lý đúng cách các loại dữ liệu mới này, vì thế cần phối hợp hay đào tạo các chuyên gia mới (ví dụ, các nhà khoa học dữ liệu).

37. Đưa ra khung pháp lý bổ sung để kích cầu sản xuất số liệu thống kê nhà nước sử dụng nguồn Big data là cần thiết. Khung pháp lý hiện hành về thống kê ở nhiều nước không bao gồm truy cập và sử dụng Big data, ở cả trong Chính phủ và khu vực tư nhân. Vì vậy, sẽ rất khó khăn để có quyền truy cập vào nguồn Big data do các tổ chức khác thu thập và lưu giữ. Hơn nữa, xây dựng khung bảo mật cùng bộ các nguyên tắc trong việc tập hợp Big data để kết hợp, bảo vệ, chia sẻ, công bố, phân tích và sở hữu các dữ liệu lớn như thế nào. Điều này sẽ giải quyết các vấn đề quan trọng của công chúng về mức độ tin tưởng khi Chính phủ sử dụng dữ liệu cá nhân. Điều quan trọng là duy trì niềm tin của công chúng: Các tổ chức phải chắc chắn rằng thông tin cá nhân của họ sẽ được bảo vệ tốt - không được tiết lộ hoặc sử dụng sai. Ví dụ, trong tích dữ liệu vị trí điện thoại di động, đó là một trong những nguồn dữ liệu tiềm năng lớn, ngay cả khi nhận dạng được ngăn cản, mọi người vẫn sẽ quan tâm đến việc chuyển các thông tin như vậy từ các nhà cung cấp viễn thông di động cho các bên khác bởi bất kỳ mục đích nào. Tương tự như vậy, các nhà cung cấp thiết bị di động cần đảm bảo rằng các quyền riêng tư sẽ không bị vi phạm khi họ lần lượt chuyển các dữ liệu sang cho Chính phủ.

38. Phản ứng của cộng đồng về việc sử dụng Big data cho thống kê nhà nước là một điều cần thiết đối với việc sử dụng của chúng ta trong tương lai. Niềm tin công đồng và sự tin cậy đối với việc sử dụng dữ liệu cá nhân cần được xây dựng (không

phải là đe dọa) bởi Chính phủ, ngoài ra cả khu vực tư nhân. Khu vực Chính phủ cần đầu tư những nỗ lực cân nhắc trong việc thành lập khung ủy quyền về việc thiết lập các quy tắc đối với việc truy cập và sử dụng các dữ liệu kỹ thuật số được sinh ra một cách tự động đối với nội dung cá nhân, và cung cấp chỉ dẫn, về mặt quy tắc đối với người nắm giữ dữ liệu và các cơ quan Chính phủ liên quan. Chỉ trong một môi trường như vậy thì quyền riêng tư và sự tin cậy về thông tin cá nhân mới được đảm bảo và niềm tin cộng đồng về việc sử dụng Big data trong thống kê nhà nước mới có thể được thiết lập.

## **VI. Các trường hợp nghiên cứu**

### **A. Úc: Sử dụng hình ảnh vệ tinh trong thống kê nông nghiệp**

39. Dữ liệu vệ tinh được dùng đối với diện tích đất và chương trình nghiên cứu về các đặc điểm đất canh tác nông nghiệp để ước lượng các loại cây trồng mùa vụ cụ thể. Các đặc điểm về không gian được sử dụng như là đầu vào cho các mô hình thống kê được tạo ra để thực hiện các phép tính trong việc biên soạn các số liệu thống kê nông nghiệp. Có khả năng, chúng cũng được áp dụng để tính cho ngư nghiệp và lâm nghiệp cũng như các số liệu thống kê môi trường. Dữ liệu thô được truy cập và xử lý trên một đám mây máy chủ của hệ thống người cung cấp bên ngoài môi trường tính toán của các tổ chức thống kê. Các thuật toán phân tích hình ảnh cho việc trích xuất ra các đặc điểm không gian về diện tích đất được phát triển nhờ sự tham gia hợp tác với các phòng thí nghiệm ở trường đại học. Điều này liên quan đến sự phát triển các thuật toán, mô hình chuỗi thời gian và không gian, hệ thống mã địa lý và quy trình dự báo thống kê.

### **B. Brazil: Dữ liệu giám sát tần số của tài khoản nước**

40. Nhờ có thỏa thuận ký kết năm 2012 giữa Viện Địa lý và thống kê quốc gia (ủy quyền bên thống kê) với cơ quan tài nguyên nước (ủy quyền Liên bang về tài nguyên nước), Ban thư ký tài nguyên nước và môi trường đô thị của Bộ Môi trường (với việc phối hợp chính sách nguồn nước), một Ủy ban đã được thành lập được ủy quyền đối với việc phát triển tài khoản nước ở Brazil. Cơ quan nước quốc gia giám sát số lượng cũng như chất lượng các nguồn nước thông qua mạng lưới khí tượng thủy văn quốc gia bao gồm 22.000 trạm quan sát, trong đó thu thập các dữ liệu tần số cao liên quan tới dòng chảy, lượng mưa, chất lượng nước, sự bốc hơi và lắng đọng. Nói cách khác, Cơ quan nước quốc gia thu thập một lượng lớn các dữ liệu về nước hàng ngày, xử lý các dữ liệu này và cho phép tự do truy cập chúng trên trang web của cơ quan<sup>37</sup> hoặc từ hệ thống thông tin tài nguyên nước quốc gia<sup>38</sup>. Hệ thống

<sup>37</sup> Xem <http://www2.ana.gov.br>.

<sup>38</sup> <Http://www.snirh.gov.br/telemetria/consultardadoshidrovisualizacao/inicial.action>.

thông tin tạo thành một hệ thống toàn diện về thu thập, xử lý, lưu trữ và cung cấp các dữ liệu và thông tin liên quan tới sự quản lý nguồn nước. Viện quốc gia kết hợp các dữ liệu cung cấp thông qua hệ thống thông tin với các dữ liệu có được từ sự thu thập số liệu thống kê về kinh tế và môi trường của mình để biên soạn các tài khoản nước của Brazil.

### **C. Bhutan: Công nghệ thông tin di động trong chỉ số giá tiêu dùng**

41. Ở Bhutan, thu thập dữ liệu giá tiêu dùng được hỗ trợ bởi công nghệ truyền thông hiện đại. Chỉ số giá tiêu dùng được sản xuất trên cơ sở hàng tháng, dựa trên dữ liệu được thu thập từ các huyện khác nhau bởi cán bộ thống kê địa phương sử dụng ứng dụng điện thoại di động Android. Các điều tra viên thu thập dữ liệu tất cả các mặt hàng trong giỏ hàng hóa từ các cửa hàng trên địa bàn huyện và tự động gửi dữ liệu đến máy chủ tại trụ sở của Văn phòng Thống kê Quốc gia ở Thimphu. Trụ sở kết hợp các dữ liệu từ các huyện và tính chỉ số giá tiêu dùng cấp quốc gia và giá trị lạm phát ở mức hàng tháng.

42. Mặc dù các ứng dụng hiện tại ở Bhutan không liên quan đến một lượng lớn dữ liệu, việc sử dụng các thiết bị di động và truyền dẫn không dây có thể được tiếp tục phát triển và áp dụng cho các bộ sưu tập giá với một số lượng lớn hơn nhiều dữ liệu. Phương pháp sử dụng ở Bhutan, bắt đầu với việc sử dụng thu thập dữ liệu điện tử quy mô nhỏ, cũng có thể được xem như là bước đệm cho các nước đang phát triển khác. Tự động hoá các nguồn đầu vào của dữ liệu số và đường truyền không dây đến một máy chủ trung tâm là bước quan trọng đầu tiên hướng tới khai thác tiềm năng của các nguồn Big data.

### **D. Colombia: Hình ảnh vệ tinh trong thống kê nông nghiệp và giao thông vận tải**

43. Colombia đã sử dụng hình ảnh vệ tinh như một nguồn dữ liệu lớn trong một số lĩnh vực. Ví dụ, một thử nghiệm sử dụng hình ảnh vệ tinh để bổ sung cho công việc của các điều tra viên trong cuộc điều tra nông nghiệp quốc gia. Giai đoạn chẩn đoán của thử nghiệm nhằm xác định loại biến có thể được ghi lại bằng hình ảnh vệ tinh và có thể hỗ trợ các thông tin được thu thập bởi hoạt động điều tra dân số. Một ví dụ khác do Văn phòng Liên hợp quốc về Ma túy và Tội phạm sử dụng hình ảnh vệ tinh để đo lường và giám sát cây coca ở Colombia thông qua các hệ thống tích hợp bất hợp pháp giám sát cây trồng.

44. Viện nghiên cứu đường bộ của Colombia sử dụng dữ liệu GPS để cải thiện lưu thông và là đầu vào cho các số liệu thống kê vận tải. Với phương pháp này, các xe không phải dừng lại ở trạm thu phí; thay vào đó, một thiết bị theo dõi điện tử được cài đặt trong các xe này được hiển thị khi xe đi qua trạm thu phí. Phí sử dụng tự động

như vậy đảm bảo sự thoải mái, an toàn và chất lượng dịch vụ ở mức độ cao. Các thiết bị theo dõi cũng có tất cả các thông tin liên quan đến chiếc xe. Cho đến nay phương pháp mới này đã được thử nghiệm tại 10 trạm thu phí ở Colombia và đã tăng cường kiểm soát các luồng giao thông và số liệu thống kê vận tải.

#### **E. Estonia: Dữ liệu định vị điện thoại di động trong thống kê du lịch quốc tế**

45. Phương pháp luận cho việc biên soạn số liệu thống kê du lịch quốc tế, trong đó sử dụng định vị di động, được phát triển bởi Ngân hàng Trung ương Estonia, Đại học Tartu và Phòng thí nghiệm Positium. Positium là một công ty thu thập và xử lý dữ liệu ẩn danh từ các nhà khai thác di động với mô hình thống kê và đồng thời mở rộng các dữ liệu này thành tập dữ liệu chung. Phương pháp này cung cấp cái nhìn tổng quan đáng tin cậy liên quan đến những người vượt biên giới Estonia để di chuyển ra nước ngoài (đi ra nước ngoài) hoặc để nhập cư vào Estonia (đi về nước). Các dữ liệu định vị di động là ẩn danh. Chúng được xử lý, phân tích và trình bày bằng cách sử dụng một phương pháp không cho phép bất kỳ loại nhận dạng số điện thoại hoặc chủ sở hữu điện thoại.

46. Ngân hàng Trung ương sử dụng số liệu thống kê du lịch quốc tế để tính toán giá trị xuất nhập khẩu dịch vụ du lịch cho tài khoản vãng lai của cán cân thanh toán. Để phù hợp với phương pháp cán cân thanh toán, định vị điện thoại di động sẽ xác định nơi cư trú của một hành khách bằng các tiêu chí thời gian cư trú lâu dài, bất kể quốc tịch của người cư trú. Có thể ước chừng thời gian cư trú của người dùng điện thoại bằng cách sử dụng chuyển vùng điện thoại và mạng di động tại địa phương. Số lượng, độ dài và tính chất của các chuyến đi của người dân và người không cư trú của Estonia được xác định bằng cách sử dụng mô hình dựa trên địa điểm của điện thoại di động, phụ thuộc vào quốc gia và khu vực.

47. Thống kê về du lịch nước ngoài phản ánh các lần trong ngày và đêm khách cư trú tại các quốc gia khác. Di chuyển nước ngoài của người dân Estonia có thể bao gồm việc tham quan một số quốc gia. Các nước đến là đất nước xa nhất từ Estonia, hoặc các nước có chi tiêu ít nhất trong hai ngày liên tiếp. Ở các quốc gia khác, được đặt chân đến mà không ở lại qua đêm và không thỏa mãn các tiêu chí quốc gia đến, được phân loại như các quốc gia quá cảnh. Số lượng và thời gian các chuyến đi của người dân và người không cư trú được tính toán với mô hình thống kê và mở rộng đến các bộ dữ liệu nói chung, trong khi cũng có tính đến các nguồn dữ liệu chính thức khác (số lượng hành khách tại cảng Tallinn và sân bay Tallinn, giao cắt biên giới phía đông của Estonia, thống kê các nơi lưu trú chính thức).

48. Các dữ liệu định vị điện thoại di động ẩn danh được sử dụng phù hợp với pháp luật của Estonia và Ủy ban châu Âu về bảo mật dữ liệu và dữ liệu cá nhân (các hành động bảo vệ dữ liệu cá nhân của Estonia; Chỉ thị 95/46/EC của Nghị viện Châu

Âu và của Hội đồng ngày 24/10/1995 về bảo hộ các cá nhân liên quan đến việc xử lý các dữ liệu cá nhân và các hoạt động về dữ liệu đó; và Chỉ thị 2002/58/EC của Nghị viện châu Âu và của Hội đồng ngày 12/7/2002 liên quan đến việc xử lý các dữ liệu cá nhân và bảo vệ sự riêng tư trong lĩnh vực truyền thông điện tử). Các phương pháp định vị điện thoại di động đã được phát triển và sử dụng trong các lĩnh vực khác nhau ở Estonia kể từ năm 2006 với sự hợp tác thanh tra liên quan đến bảo vệ dữ liệu và dữ liệu cá nhân.

#### **F. Hà Lan: Phương tiện truyền thông xã hội như một nguồn dữ liệu tiềm năng cho thống kê nhà nước**

49. Ngày càng có nhiều người dân đang hoạt động trên phương tiện truyền thông xã hội. Ở đây, mọi người tự nguyện chia sẻ thông tin, thảo luận về chủ đề quan tâm và mối liên hệ gia đình, bạn bè. Tại Hà Lan, tỷ lệ lớn dân số sử dụng phương tiện truyền thông xã hội; khoảng 70% báo cáo có những hoạt động trên một hoặc nhiều phương tiện truyền thông xã hội. Kể từ khi việc trả lời các câu hỏi của Cơ quan thống kê Hà Lan tiếp tục suy giảm, tiềm năng sử dụng các tin nhắn được tạo ra và chia sẻ tự nguyện trên phương tiện truyền thông xã hội như là một nguồn dữ liệu cho thống kê nhà nước đã được điều tra. Với mục đích này, các thông điệp công khai trên phương tiện truyền thông xã hội, chẳng hạn như Twitter và Facebook, cũng như các tin nhắn công khai đăng trên trang tin, diễn đàn web và blog, đã được nghiên cứu. Các tin nhắn thu được từ một công ty thương mại thường thu thập tất cả các thông báo công khai bằng văn bản tiếng Hà Lan trên lập ngôn ngữ của web-tiếng Hà Lan.

50. Cả nội dung và ý nghĩa của các thông điệp đã được nghiên cứu. Các nghiên cứu nội dung các tin nhắn ở Hà Lan trên Twitter, các phương tiện truyền thông xã hội sử dụng tiếng Hà Lan, chỉ ra rằng gần 50% tin nhắn là "lảm nhảm và vô nghĩa". Phần còn lại chủ yếu thảo luận về các hoạt động trong thời gian rảnh rỗi (10%), công việc (7%), truyền thông (truyền hình và đài phát thanh) (5%) và chính trị (3%). Việc sử dụng những thông điệp mang nhiều ý nghĩa cũng bị cản trở bởi các thông điệp có nội dung "lảm nhảm". Xác định ý nghĩa trong tất cả các tin nhắn trên tất cả các nền tảng có sẵn đưa ra một tiềm năng sử dụng rất thú vị của các số liệu thống kê. Ý nghĩa trong các tin nhắn này được phát hiện là có liên quan đến sự tin cậy của người tiêu dùng Hà Lan, đặc biệt là liên quan đến tình hình kinh tế. Các mối quan hệ sau này là ổn định ở hàng tháng dựa trên cơ sở hàng tuần. Số liệu biến động cao hàng ngày cho thấy chúng có thể sản xuất các chỉ số tâm lý hàng tháng và hàng tuần có thể so sánh với độ tin cậy của người tiêu dùng. Các chỉ số sau có thể được sản xuất vào ngày làm việc đầu tiên sau tuần nghiên cứu, chứng minh khả năng để cung cấp kết quả nhanh chóng.



## **VII. Kết luận và con đường phía trước**

51. Cộng đồng thống kê đã nhận ra tiềm năng của Big data cho thống kê nhà nước. Big data có thể giúp thực hiện tốt hơn nhiệm vụ cung cấp số liệu thống kê kịp thời và chặt chẽ về kinh tế, xã hội và môi trường phục vụ cho việc ra quyết định, nghiên cứu và tranh luận công khai. Hơn nữa, cùng với sự tiến bộ của công nghệ trong việc mã hóa các đơn vị của các nền kinh tế, xã hội và môi trường và sự sẵn có ngày càng tăng của thông tin ở cấp độ thấp, sự phát triển đầu tiên vào việc sử dụng Big data tại các nước phát triển và đang phát triển. Vì vậy, phát triển các phương pháp luận, các thực hành tốt và cơ hội đào tạo cần phải được chia sẻ trong cộng đồng thống kê toàn cầu thông qua các sáng kiến khu vực, nếu Big data là để lợi dụng, kể cả trong việc giải quyết các vấn đề liên quan đến pháp luật, sự riêng tư, tài chính, quản lý, phương pháp và công nghệ. Trong khi giải quyết những vấn đề Big data, Chính phủ và khu vực doanh nghiệp sẽ được kêu gọi để tạo ra một môi trường đảm bảo sự riêng tư và bảo mật thông tin cá nhân cho mục đích thống kê nhà nước, và công chúng có thể tin tưởng việc sử dụng Big data cho thống kê nhà nước.

52. Các đánh giá toàn cầu về việc sử dụng Big data cho thống kê nhà nước mang lại một số ví dụ tốt, đặc biệt là từ các giao dịch, theo dõi, cảm biến và các nguồn dữ liệu vệ tinh. Thống kê giá cả đã được biên soạn từ các thông tin máy quét, chi tiêu của người tiêu dùng từ các thẻ giao dịch điện tử, số liệu thống kê đi lại và vận chuyển từ các thiết bị theo dõi GPS, các số liệu thống kê tốt từ phương tiện truyền thông xã hội và dữ liệu thống kê nông nghiệp từ hình ảnh vệ tinh. Các cuộc khảo sát toàn cầu cũng cho thấy nhiều thách thức, đặc biệt là đối với những khó khăn về phương pháp trong việc biên soạn Big data, nhu cầu công nghệ thông tin và sự quan tâm các vấn đề riêng tư.

53. Rõ ràng là phải đối mặt với những thách thức của Big data trong hiện đại hóa hệ thống thống kê. Ví dụ, các bộ phận nghiên cứu của hệ thống thống kê sẽ cần phải được tăng cường, năng lực công nghệ thông tin cũng cần phải nâng cao, nguồn nhân lực sẽ phải được tái cấu trúc để bổ sung thêm các nhà khoa học dữ liệu, và các đối tác sẽ cần phải phối hợp với khu vực tư nhân trong việc chia sẻ thông tin kỹ thuật số tự động tạo ra trong một môi trường nơi mà những mối quan tâm riêng tư được giải quyết và các quy tắc bảo mật được quy định.

54. Dữ liệu lớn và hiện đại hóa hệ thống thống kê giới thiệu những thách thức và cơ hội đối với hầu hết các nước. Kinh nghiệm thực hành và các giải pháp đó có thể được chia giữa các Viện Thống kê Quốc gia. Một số sáng kiến theo hướng đó đã được bắt đầu ở cấp khu vực. Trong bối cảnh của Hội nghị các nhà thống kê châu Âu, một đề xuất dự án ECE đã được phê duyệt để phát huy vai trò của Big data trong việc hiện đại hóa sản xuất thống kê. Sáng kiến tương tự đang được ESCAP thông

qua trong khu vực châu Á-Thái Bình Dương. Hơn nữa, sự ra đời của dự án phản ánh nguyện vọng tương tự đang được xem xét trong các khu vực khác.

55. Nhóm làm việc toàn cầu về Big data được hình dung là để bổ sung các nguyện vọng và sáng kiến của khu vực trong Chương trình toàn cầu về công việc. Chương trình này cho phép chia sẻ kinh nghiệm giữa các Ủy ban trong khu vực. Trường hợp các Ủy ban khu vực có thể làm việc cũng với một số nước trong khu vực ở các Dự án cụ thể, nhóm làm việc toàn cầu sẽ được chia sẻ và quản lý để đạt được thỏa thuận các nội dung liên quan đến phương pháp luận, kinh nghiệm thực hành tốt nhất, đào tạo và sử dụng các nền tảng phổ biến. Là một phần trong chương trình làm việc của mình, nhóm làm việc toàn cầu cần đưa ra các trường hợp mà hệ thống thống kê quốc gia có lợi thế so sánh so với các bên khác trong chứng minh vai trò sử dụng chính sách dữ liệu lớn hoặc kết hợp các nguồn dữ liệu lớn với số liệu thống kê nhà nước hiện có.

56. Nhóm làm việc toàn cầu còn có thể giải quyết các vấn đề cụ thể liên quan đến vai trò của Big data cho các Chương trình nghị sự phát triển 2015 và nên tạo điều kiện hợp tác với các tổ chức khác của Liên hợp quốc và quốc tế, các tổ chức xã hội dân sự và các thành viên khác của các đối tác toàn cầu cho các sáng kiến cách mạng dữ liệu mới nổi. Một trong những khiếm khuyết liên quan đến các chỉ số Mục tiêu Phát triển Thiên niên là tính kịp thời kém. Big data thường được tạo ra trong thời gian thực và có khả năng cung cấp các chỉ số một cách kịp thời.

57. Để khởi động nhóm làm việc toàn cầu, một Hội nghị quốc tế sẽ được tổ chức trong Quý 2 hoặc Quý 3/2014 với mục đích xây dựng chương trình làm việc dựa trên phương pháp luận chung, đào tạo, nền tảng và các vấn đề khác liên quan đến việc sử dụng Big data cho thống kê nhà nước.

### **VIII. Các điểm thảo luận**

58. UNSD được mời bày tỏ quan điểm của mình về:

(a) Sự cần thiết phải điều tra thêm các nguồn tin, những thách thức và lĩnh vực sử dụng Big data cho thống kê nhà nước ở cấp độ toàn cầu, đặc biệt là đối với với hoàn cảnh của các nước đang phát triển và liên kết các Chương trình nghị sự phát triển 2015 và chủ động cách mạng dữ liệu;

(b) Việc xây dựng một nhóm làm việc toàn cầu về việc sử dụng Big data cho thống kê nhà nước mà hoạt động này sẽ bổ sung cho các công việc được thực hiện bởi các Ủy ban trong khu vực và quản lý các vấn đề có liên quan trên toàn cầu.