

## THỐNG KÊ NHÀ NƯỚC VỚI BIG DATA: KINH NGHIỆM QUỐC TẾ VÀ ĐỊNH HƯỚNG CỦA TỔNG CỤC THỐNG KÊ

ThS.Nguyễn Văn Đoàn  
Viện trưởng Viện KHTK

Dữ liệu lớn (Big data) là chủ đề đang được thế giới quan tâm nghiên cứu, ứng dụng vào nhiều lĩnh vực như: Hoạt động chính trị; giao thông; y tế; thể thao; tài chính; thương mại; thống kê... Bài viết này tập trung trình bày một số nghiên cứu ứng dụng Big data trong thống kê Nhà nước của cộng đồng thống kê thế giới và một số kiến nghị đối với thống kê Việt Nam. Báo cáo này gồm 5 nội dung: 1) Một vài số liệu thống kê được sản xuất từ Big data; 2) Dự án "Vai trò của Big data trong hiện đại hóa sản xuất thống kê"; 3) Nhóm Công tác toàn cầu về Big data của Ủy ban Thống kê Liên hợp quốc; 4) Cơ quan thống kê quốc gia với Big data; 5) Định hướng của Tổng cục Thống kê.

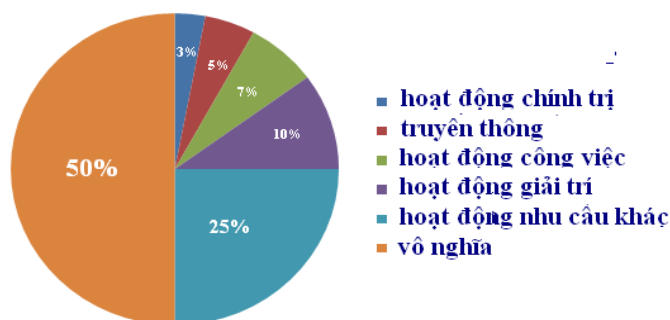
### 1. Một vài số liệu thống kê được sản xuất từ nguồn Big data

- Thống kê về số tin nhắn của Hà Lan

Số lượng các tin nhắn được sản xuất từ nguồn Big data của Thống kê Hà Lan. Cơ quan thống kê Hà Lan đã thống kê số lượng các tin nhắn thu thập được từ các phương tiện truyền thông ở Hà lan cho thấy có đến 50% số lượng các tin nhắn vô nghĩa (xem Đồ thị 1).

Số liệu này nói rằng người dân Hà Lan đã phải chi trả một khoản tiền khá lớn, nhưng không mang lại lợi ích gì. Một kiến nghị với Chính phủ là cần có những điều chỉnh về hoạt động này để mang lại lợi ích cho người dân cũng như xã hội Hà Lan.

**Hình 1:** Số lượng các tin nhắn phân theo nội dung tin nhắn



Nghiên cứu nội dung tin nhắn tại HÀ LAN

Nguồn: Báo cáo kết quả khảo sát dữ liệu lớn của UNSD/UNECE

- Một số lĩnh vực thống kê đang sử dụng nguồn Big data

Theo kết quả điều tra của Cơ quan thống kê Liên hợp quốc (UNSD) và Ủy ban Kinh tế châu Âu của Liên hợp quốc (UNECE)<sup>11</sup>, cho thấy có 10 lĩnh vực thống kê Nhà nước đang được sản xuất từ nguồn Big data: thống kê kinh tế và tài chính (48.1%); thống kê xã hội và nhân khẩu học (44.2%); thống kê giá (38.5%); thống kê...

**Hình 2:** Các lĩnh vực tiềm năng sử dụng Big data

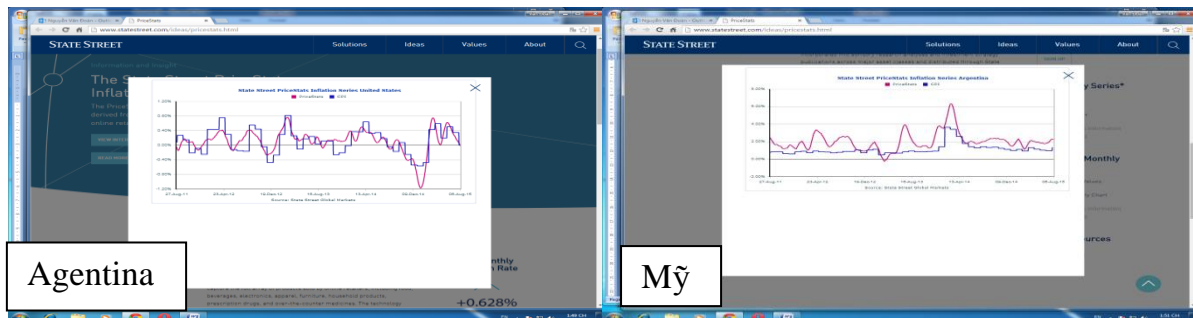


Nguồn: Báo cáo kết quả khảo sát dữ liệu lớn của UNSD/UNECE

#### - Chỉ số lạm phát toàn cầu của **State Street PriceStats**

Nhà cung cấp chỉ số lạm phát toàn cầu sử dụng phần mềm quét dữ liệu các hàng hóa được bán trên các trang web (khoảng 5 triệu mặt hàng, hơn 70 nước). Hiện nay thường xuyên cung cấp chỉ số lạm phát cho 22 quốc gia, trong đó có Mỹ và Argentina (Hình 3).

**Hình 3:** CPI của Mỹ, Argentina



Màu đỏ là CPI được biên soạn từ Big data; Màu xanh là CPI truyền thống

## 2. Dự án “Vai trò của Big data trong hiện đại hóa sản xuất thống kê”

Nhận thức được những lợi ích to lớn của Big data cũng như những thách thức của Big data trong hiện đại hóa sản xuất thống kê, năm 2010 Ủy ban Kinh tế châu

<sup>11</sup> Results of the UNSD/UNECE Survey on Organizational context and individual projects of Big data

Âu của Liên hợp quốc (UNECE) đã hình thành Nhóm Cấp cao về Hiện đại hóa thống kê Nhà nước (HLG-MOS)<sup>12</sup> để giám sát và điều phối công việc quốc tế liên quan đến hiện đại hóa thống kê. Theo đó, năm 2014, Dự án “Vai trò của Big data trong hiện đại hóa sản xuất thống kê được vận hành với 3 mục tiêu chính: (1) Hướng dẫn các Cơ quan thống kê quốc gia các vấn đề về chiến lược và phương pháp luận chính mà Big data đặt ra đối với thống kê Nhà nước; (2) Chứng minh tính khả thi, hiệu quả của việc sản xuất số liệu thống kê Nhà nước sử dụng nguồn Big data so với thống kê truyền thống; (3) Chia sẻ kiến thức, kinh nghiệm, công cụ và phương pháp để sản xuất số liệu thống kê Nhà nước từ các nguồn Big data. Để đạt được các mục tiêu trên, Dự án hình thành 4 nhóm công việc sau:

*Nhóm Chất lượng số liệu* gồm 13 thành viên đến từ 9 Cơ quan thống kê quốc gia và quốc tế. Nhóm này chịu trách nhiệm thiết kế khung chất lượng cho Big data;

*Nhóm Bảo mật thông tin riêng tư* gồm 11 thành viên đến từ 9 Cơ quan thống kê quốc gia, quốc tế, Viện, Trường). Nhóm này chịu trách nhiệm xem xét các công cụ sẵn có để quản lý rủi ro trong quan điểm về vấn đề riêng tư, tập trung vào các đặc điểm của Big data và tác động của chúng đối với thông tin riêng tư;

*Nhóm Đối tác* gồm 12 thành viên đến từ 8 Cơ quan thống kê quốc gia, quốc tế. Nhiệm vụ của nhóm này là xây dựng mối quan hệ đối tác với các cơ quan, tổ chức có nguồn dữ liệu lớn, với cộng đồng khoa học, nhà cung cấp công nghệ;

*Nhóm Công nghệ* gồm 38 thành viên đến từ 18 Cơ quan thống kê quốc gia, quốc tế. Nhóm này chịu trách nhiệm thiết kế, cài đặt và sử dụng môi trường máy tính truy cập web, các công cụ và phương pháp sản xuất số liệu thống kê Nhà nước từ nguồn Big data.

Kết quả của Dự án đã trình bày tại Hội nghị quốc tế về Big data trong thống kê Nhà nước được tổ chức năm 2014, tại Bắc Kinh và sẽ tiếp tục bổ sung kết quả thực hiện năm 2015 để báo cáo tại Hội thảo về “Big data in secure societies” được tổ chức tại Brussels (Bỉ) vào 30/9/2015<sup>13</sup> và Hội nghị toàn cầu lần thứ hai về Big data đối với thống kê Nhà nước sẽ được tổ chức tại Abu Dhabi, United Arab Emirates, từ ngày 20-22/10/2015<sup>14</sup>.

<sup>12</sup> UNECE (2013) what does “big data” mean for official statistic? Report of the High-level Group for the Modernisation of Statistical Production and Services (HLG). <http://www1.unece.org/stat/platform/display/hlgbas>

<sup>13</sup> <http://www.big-data-europe.eu/event/sc7-brussels-2015/>

<sup>14</sup> Chủ đề cuộc hội thảo này là “Moving from examples to guidelines” (tạm dịch là “Chuyển từ các ví dụ đến hướng dẫn thực hành”), nhằm thúc đẩy sự phát triển của Big data thêm một bước nữa.

### **(3) Ủy ban Thống kê Liên hợp quốc với Big data**

Tại Phiên họp lần thứ 45 của Ủy ban thống kê Liên hợp quốc (UNSC) được tổ chức vào tháng 3/2014, tại New York (Mỹ), UNSC đã thành lập Nhóm Công tác toàn cầu (Global Working Group - GWG) về Big data với 28 thành viên (18 quốc gia và 10 tổ chức quốc tế, tổ chức đào tạo, khu vực tư nhân<sup>15</sup>), nhằm nghiên cứu giải quyết các vấn đề về phương pháp luận; chất lượng; công nghệ; truy cập dữ liệu; pháp luật; bảo mật; quản lý và tài chính; phân tích chi phí-lợi ích về Big data.

Phiên làm việc đầu tiên của GWG (31/10/2014) đã khẳng định việc sử dụng Big data cho thống kê Nhà nước là một nghĩa vụ đối với cộng đồng thống kê quốc tế dựa trên các nguyên tắc cơ bản để đáp ứng sự kỳ vọng của xã hội đối với các sản phẩm thống kê và cải thiện cách thức làm việc đạt hiệu quả hơn. Tại cuộc họp này, GWG đã hình thành 8 tổ công tác (task teams) theo các chủ đề: Vận động và truyền thông; liên kết Big data và các Mục tiêu Phát triển bền vững; tiếp cận và quan hệ đối tác; đào tạo, xây dựng kỹ năng và năng lực thống kê; dữ liệu từ điện thoại di động; ảnh vệ tinh; dữ liệu từ các phương tiện truyền thông xã hội.

Một số kết quả nghiên cứu bước đầu về Big data đối với thống kê Nhà nước đã được công bố: Các Cơ quan thống kê Quốc gia (NSOs) có thể tận dụng nguồn dữ liệu mới và phương thức thu thập như thế nào? Chính phủ điện tử - Thống kê điện tử - Làm thế nào để các nhà thống kê đưa chúng vào trung tâm của chuyển đổi chính quyền thông qua khoa học dữ liệu; suy luận Big data và mô hình cho thống kê Nhà nước, với ứng dụng của phân tích hình ảnh vệ tinh; Sử dụng Big data trong thống kê Nhà nước; Big data và thống kê Nhà nước - Thách thức của Big data trong Hệ thống Thống kê quốc gia; quyền riêng tư, bảo mật và Big data; thống kê trong thời đại thách thức hay cơ hội của Big data; sự hợp nhất dữ liệu không gian và thời gian cho Big data và ứng dụng của nó trong vệ tinh viễn thám (Chi tiết, xem tại Report\_GWG2014\_E.pdf).

Hoạt động tiếp theo: Nhóm Công tác toàn cầu về Big data của Ủy ban Thống kê Liên hợp quốc tiếp tục nghiên cứu và ứng dụng Big data trong thống kê Nhà nước.

<sup>15</sup> 18 quốc gia (Australia, Bangladesh, Cameroon, China, Colombia, Denmark, Egypt, Indonesia, Italy, Mexico, Morocco, Netherlands, Oman, Pakistan, Philippines, United Arab Emirates, United Republic of Tanzania, United States of America) và 10 tổ chức quốc tế, đào tạo, khu vực tư nhân (Economic and Social Commission for Asia and the Pacific/Statistical Institute for Asia and the Pacific, Economic Commission for Europe, Eurostat, International Telecommunication Union, Organization for Economic Cooperation and Development, Statistical Centre for the Cooperation Council for the Arab Countries of the Gulf, Statistics Division of the Department of Economic and Social Affairs of the, United Nations, United Nations Global Pulse, Universal Postal Union, World Bank.

Đặc biệt tập trung vào các chủ đề như: vận động và truyền thông về Big data; liên kết Big data với các Mục tiêu Phát triển bền vững; đào tạo, xây dựng các kỹ năng và năng lực; các thách thức đối với Big data... Tổ chức một số cuộc hội thảo quốc tế và khu vực về Big data với thống kê Nhà nước sẽ được tổ chức vào cuối năm 2015, trong đó có Hội nghị toàn cầu lần thứ hai về Big data đối với thống kê Nhà nước sẽ được tổ chức tại Abu Dhabi, United Arab Emirates, từ ngày 20-22/10/2015<sup>16</sup>; Hội nghị về "Toán thống kê: Cơ hội và thách thức với Big data" do Hiệp hội Toán thống kê quốc tế (IASC-ARS) tổ chức tại Singapore, từ 17-19/12/2015.

#### **4. Cơ quan thống kê quốc gia với Big data**

Cơ quan thống kê quốc gia của một số nước đang triển khai nghiên cứu sử dụng nguồn Big data để sản xuất số liệu thống kê Nhà nước. Một số nước đã xây dựng hẳn một chiến lược về Big data (Big data strategy) như: China, Denmark, Finland, Italy, Japan, Portugal, Romania, Serbia, Sweden, UK.

Một số dự án về Big data đang được thống kê các nước triển khai thực hiện như sau:

Thống kê Cameroon (NIS) thực hiện 02 dự án: Xây dựng năng lực trong sử dụng Big data như nguồn số liệu thống kê Nhà nước; xây dựng năng lực cho việc sử dụng Big data cho mục đích thống kê<sup>17</sup>.

Thống kê Anh đang thực hiện 4 dự án: Khai thác cơ sở dữ liệu về thị trường thương mại để ước tính số liệu điều tra dân số; Kiểu dữ liệu Smartmeter cho cấu trúc hộ gia đình/quy mô và nghề nghiệp; tiềm năng dữ liệu Smartmeter để phát hiện ngôi nhà vắng chủ; Aggregated Mobile Phone data to identify commuting patterns<sup>18</sup>.

Thống kê Mexico thực hiện 2 dự án: Tweet analysis; Thống kê Nhà nước sử dụng dữ liệu định vị của điện thoại di động với một ứng dụng cụ thể để xây dựng lưới dân số.

Thống kê Trung quốc thực hiện 2 dự án: Big Data Enterprise Statistical Indicator Ten-day Report; Online Price Changes of Means of Production in Circulation Area in Shandong Zhuochuang;

<sup>16</sup> Chủ đề cuộc hội thảo này là "Moving from examples to guidelines" (tạm dịch là "Chuyển từ các ví dụ đến hướng dẫn thực hành"), nhằm thúc đẩy sự phát triển của Big data thêm một bước nữa.

<sup>17</sup> Capacity building in using Big data as sources for public statistics; Capacity building for the use of Big data for statistical Cameroon.

<sup>18</sup> Counts of individuals held on a major commercial marketing database comparison to Census data estimates; Smartmeter type data for household structure/size and occupancy; Smartmeter data potential for detecting Unoccupied dwellings



Thống kê Canada thực hiện 2 dự án: Non-Residential Buildings Inventory: Feasibility Study; House market indicators (based on website information)

Thống kê Romania thực hiện 1 dự án: Using scanner data

Thống kê Nam Phi thực hiện 1 dự án: Assessing use of scanner data for compiling the Consumer Price Index

Thống kê Italy (ISTAT) thực hiện 3 dự án: Persons and Places: Mobility Estimates based on Mobile Phone Data; Use of scanner data for consumer price index; Internet as a Data Source for ICT Usage by Enterprises and Public Institutions.

**Tóm lại:** Big data có vai trò trong hiện đại hóa sản xuất thông tin thống kê Nhà nước, nên đang được cộng đồng thống kê thế giới quan tâm. Ủy ban Thống kê (UNSC), Ủy ban Kinh tế châu Âu (UNECE) của Liên hợp quốc đều thành lập nhóm công tác chuyên về Big data. Cơ quan thống kê quốc gia của một số nước, tổ chức quốc tế đang triển khai thực hiện các dự án sử dụng nguồn Big data như nguồn dữ liệu thay thế hoặc bổ sung cho các nguồn dữ liệu truyền thống để sản xuất thông tin thống kê Nhà nước.

Sử dụng nguồn Big data không chỉ làm giảm thời gian và chi phí sản xuất thông tin thống kê, mà còn làm tăng sức hút của những nguồn dữ liệu này. Tuy nhiên, những thách thức cũng không hề nhỏ, đó là vấn đề hành lang pháp lý, phương pháp luận, công nghệ thông tin, kỹ năng, bảo mật thông tin riêng tư, đặc biệt là thách thức trong việc tiếp cận nguồn Big data do các tổ chức, khu vực tư nhân nắm giữ.

Đối với Thống kê Việt Nam nói chung, Tổng cục Thống kê nói riêng, Big data là vấn đề rất mới và đầy tiềm năng. Sử dụng nguồn dữ liệu Big data để bổ sung hoặc thay thế nguồn dữ liệu truyền thống để sản xuất số liệu thống kê Nhà nước không chỉ giải quyết vấn đề chất lượng và hiệu quả của các hoạt động thống kê ở Việt Nam, mà còn đóng góp vào các hoạt động của thống kê thế giới. Những thách thức khi sử dụng nguồn Big data sẽ phức tạp hơn nhiều so với các nước có nền thống kê phát triển. Tổng cục Thống kê cần có cách tiếp cận phù hợp để sử dụng hiệu quả nguồn dữ liệu Big data cho thống kê Nhà nước.

### **5) Định hướng của Tổng cục Thống kê**

(1) Lãnh đạo Tổng cục cần có quyết tâm cao và cam kết chính trị trong việc sử dụng nguồn dữ liệu Big data đầy tiềm năng với thống kê Nhà nước theo phương thức vừa học, vừa làm.

(2) Tổng cục giao một đơn vị chủ trì phối hợp với các đơn vị liên quan thành lập Tổ Công tác Big data (Tổ trưởng là một lãnh đạo cấp Vụ/tương đương trong danh

sách quy hoạch lãnh đạo Tổng cục). Tổ này được chia thành các bộ phận: Phương pháp luận; công nghệ; đối tác...). Tổ Công tác có các nhiệm vụ chính sau:

- Xác định tầm nhìn đến năm 2030 về sử dụng nguồn Big data cho hiện đại hóa sản xuất thông tin thống kê Nhà nước;

- Xây dựng và triển khai thực hiện kế hoạch về sử dụng nguồn Big data cho thống kê Nhà nước. Trước mắt, cần nghiên cứu, học hỏi kinh nghiệm của các nước và tổ chức quốc tế (Nghiên cứu tài liệu; tham dự hội thảo và các lớp đào tạo Big data của khu vực, quốc tế; tham gia vào dự án quốc tế...).

Tiếp theo, cần chọn một hoặc một số lĩnh vực thống kê cụ thể (ví dụ, thống kê giá (CPI), thống kê du lịch, thống kê công nghệ thông tin và truyền thông (ICT), thống kê tổng mức bán lẻ) để thử nghiệm và so sánh với số liệu thống kê được sản xuất theo phương pháp truyền thống.

(2) Thực hiện việc truyền thông về vai trò của Big data trong hiện đại hóa sản xuất thông tin thống kê Nhà nước; lợi ích và thách thức của Big data; xu hướng sử dụng Big data của cộng đồng thống kê thế giới.

(3) Xây dựng các đối tác chiến lược về nguồn dữ liệu Big data

(4) Bố trí kinh phí hàng năm từ nguồn ngân sách nhà nước và hỗ trợ kỹ thuật của các dự án, hợp tác song phương...

#### **Tài liệu tham khảo:**

[1] Statistical Commission, Report of the Global Working Group on Big Data for Official Statistics, Forty-sixth session, March 2015;

[2] Statistical Commission, Results of the UNSD/UNECE Survey on organizational context and individual projects of Big data, Forty-sixth session, March 2015;

[3] Tổng cục Thống kê, Báo cáo kết quả chuyên công tác nước ngoài, 8/2015;

[4] Capacity building in using Big data as sources for public statistics; Capacity building for the use of Big data for statistical Cameroon;

[5] Counts of individuals held on a major commercial marketing database comparison to Census data estimates; Smartmeter type data for household structure/size and occupancy; Smartmeter data potential for detecting Unoccupied dwellings;

[6] <http://www.big-data-europe.eu/event/sc7-brussels-2015/>.