

TỔNG QUAN VỀ DỮ LIỆU LỚN

ThS.Nguyễn Văn Đoàn¹

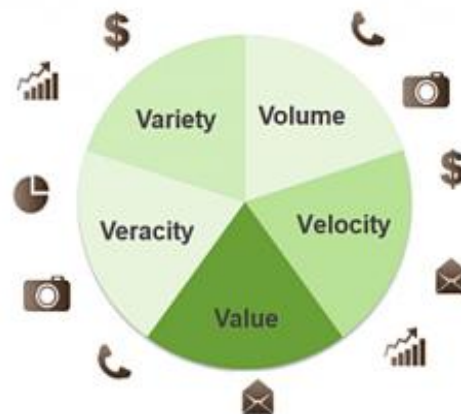
KS.Nguyễn Công Hoar²

Trước đây, chúng ta mới chỉ biết đến dữ liệu có cấu trúc (structure data), ngày nay, với sự kết hợp của dữ liệu và internet, đã xuất hiện một dạng khác của dữ liệu - Big data (dịch là "dữ liệu lớn"). Dữ liệu này có từ các nguồn như: hồ sơ hành chính, giao dịch điện tử, dòng trạng thái (status), chia sẻ hình ảnh, bình luận, nhắn tin... của chính chúng ta, nói cách khác chúng là dữ liệu được sản sinh qua quá trình chia sẻ thông tin trực tuyến liên tục của người sử dụng. Nhận thấy vai trò và lợi ích mà dữ liệu lớn đem lại, cộng đồng thống kê quốc tế đã và đang có rất nhiều dự án nghiên cứu, ứng dụng dữ liệu lớn. Để cung cấp cái nhìn tổng quan về dữ liệu lớn cho quý độc giả, bài viết này giới thiệu tóm tắt những nét chính cũng như những cơ hội và thách thức mà dữ liệu lớn mang lại.

Khái niệm, đặc trưng của dữ liệu lớn và sự khác biệt với dữ liệu truyền thống

- Dữ liệu lớn (Big data) là một thuật ngữ chỉ bộ dữ liệu lớn hoặc phức tạp mà các phương pháp truyền thống không đủ các ứng dụng để xử lý dữ liệu này [1].

Dữ liệu lớn được hình thành chủ yếu từ các nguồn: (1) Dữ liệu hành chính (phát sinh từ chương trình của một tổ chức, có thể là chính phủ hay phi chính phủ). Ví dụ, hồ sơ y tế điện tử ở bệnh viện, hồ sơ bảo hiểm, hồ sơ ngân hàng...; (2) Dữ liệu từ hoạt động thương mại (phát sinh từ các giao dịch giữa hai thực thể). Ví dụ, các giao dịch thẻ tín dụng, giao dịch trên mạng, bao gồm cả từ các thiết bị di động; (3) Dữ liệu từ các thiết bị cảm biến như hình ảnh vệ tinh, cảm biến đường, cảm biến khí hậu; (4) Dữ liệu từ các thiết bị theo dõi, ví



dự theo dõi dữ liệu từ điện thoại di động, GPS; (5) Dữ liệu từ các hành vi, ví dụ như tìm kiếm trực tuyến về một sản phẩm, một dịch vụ hay bất kỳ loại thông tin khác, trang xem trực tuyến; (6) Dữ liệu từ các thông tin ý kiến trên các phương tiện thông tin xã hội.

Dữ liệu lớn có 5 đặc trưng cơ bản như sau (mô hình 5Vs về dữ liệu lớn)³:

¹ Viện trưởng Viện Khoa học Thống kê

² Trung tâm Thông tin khoa học thống kê, Viện Khoa học Thống kê

³ Mô hình 5Vs do Gartner xây dựng. Gartner là công ty nghiên cứu và tư vấn về công nghệ thông tin hàng đầu thế giới do một người Mỹ tên là Gideon Gartner sáng lập năm 1979. 5Vs: Khối lượng dữ liệu (Volume); Tốc độ (Velocity); Giá trị (Value); Độ tin cậy/chính xác (Veracity); Đa dạng (Variety).

(1) Khối lượng dữ liệu (Volume): Đây là đặc điểm tiêu biểu nhất của dữ liệu lớn, khối lượng dữ liệu rất lớn. Kích cỡ của Big Data đang từng ngày tăng lên, và tính đến năm 2012 thì nó có thể nằm trong khoảng vài chục terabyte cho đến nhiều petabyte (1 petabyte = 1024 terabyte) chỉ cho một tập hợp dữ liệu. Dữ liệu truyền thống chúng ta có thể lưu trữ trên các thiết bị đĩa mềm, đĩa cứng. Nhưng với dữ liệu lớn sẽ sử dụng công nghệ “đám mây” mới có khả năng lưu trữ được dữ liệu lớn.

(2) Tốc độ (Velocity): Tốc độ có thể hiểu theo 2 khía cạnh: (a) Khối lượng dữ liệu gia tăng rất nhanh (mỗi giây có tới 72.9 triệu các yêu cầu truy cập tìm kiếm trên web bán hàng của Amazon); (b) Xử lý dữ liệu nhanh ở mức thời gian thực (real-time), có nghĩa dữ liệu được xử lý ngay tức thời ngay sau khi chúng phát sinh (tính đến bằng mili giây). Các ứng dụng phổ biến trên lĩnh vực Internet, Tài chính, Ngân hàng, Hàng không, Quân sự, Y tế – Sức khỏe như hiện nay phần lớn dữ liệu lớn được xử lý real-time. Công nghệ xử lý dữ liệu lớn ngày một tiên tiến cho phép chúng ta xử lý tức thì trước khi chúng được lưu trữ vào cơ sở dữ liệu.

(3) Đa dạng (Variety): Đối với dữ liệu truyền thống chúng ta hay nói đến dữ liệu có cấu trúc, thì ngày nay hơn 80% dữ liệu được sinh ra là phi cấu trúc (tài liệu, blog, hình ảnh, vi deo, bài hát, dữ liệu từ thiết bị cảm biến vật lý, thiết bị chăm sóc sức khỏe...). Big Data cho phép liên kết và phân tích nhiều dạng dữ liệu khác nhau. Ví dụ, với các comments/post của một nhóm người dùng nào đó trên Facebook với thông tin video được chia sẻ từ Youtube và Twitter.

(4) Độ tin cậy/chính xác (Veracity): Một trong những tính chất phức tạp nhất của BigData là độ tin cậy/chính xác của dữ liệu. Với xu hướng phương tiện truyền thông xã hội

(Social Media) và mạng xã hội (Social Network) ngày nay và sự gia tăng mạnh mẽ tính tương tác và chia sẻ của người dùng Mobile làm cho bức tranh xác định về độ tin cậy và chính xác của dữ liệu ngày một khó khăn hơn. Bài toán phân tích và loại bỏ dữ liệu thiếu chính xác và nhiễu đang là tính chất quan trọng của BigData.

(5) Giá trị (Value): Giá trị là đặc điểm quan trọng nhất của dữ liệu lớn, vì khi bắt đầu triển khai xây dựng dữ liệu lớn thì việc đầu tiên chúng ta cần phải làm đó là xác định được giá trị của thông tin mang lại như thế nào, khi đó chúng ta mới có quyết định nên triển khai dữ liệu lớn hay không. Nếu chúng ta có dữ liệu lớn mà chỉ nhận được 1% lợi ích từ nó, thì không nên đầu tư dữ liệu lớn. Kết quả dự báo chính xác thể hiện rõ nét nhất về giá trị của dữ liệu lớn mang lại. Ví dụ, từ khối dữ liệu phát sinh trong quá trình khám, chữa bệnh sẽ giúp dự báo về sức khỏe được chính xác hơn, sẽ giảm được chi phí điều trị và các chi phí liên quan đến y tế.

Dữ liệu lớn khác với dữ liệu truyền thống (ví dụ, kho dữ liệu - Data Warehouse) ở 4 điểm: Dữ liệu đa dạng hơn; lưu trữ dữ liệu lớn hơn; truy vấn nhanh hơn; độ chính xác cao hơn.

(1) Dữ liệu đa dạng hơn: Khi khai thác dữ liệu truyền thống (thường là dữ liệu có cấu trúc), chúng ta thường phải trả lời các câu hỏi: Dữ liệu lấy ra kiểu gì? định dạng dữ liệu như thế nào? Đối với dữ liệu lớn, không phải trả lời các câu hỏi trên. Hay nói khác, khi khai thác, phân tích dữ liệu lớn chúng ta không cần quan tâm đến kiểu dữ liệu và định dạng của chúng; điều quan tâm là giá trị mà dữ liệu mang lại có đáp ứng được cho công việc hiện tại và tương lai hay không.

(2) Lưu trữ dữ liệu lớn: Lưu trữ dữ liệu truyền thống vô cùng phức tạp và luôn đặt ra

câu hỏi lưu trữ nào: dung lượng kho lưu trữ bao nhiêu là đủ? gắn kèm với câu hỏi đó thì chi phí đầu tư sẽ là rất lớn. Công nghệ lưu trữ dữ liệu lớn hiện nay đã phần nào có thể giải quyết được vấn đề trên nhờ những công nghệ lưu trữ đám mây, phân phối lưu trữ dữ liệu phân tán và có thể kết hợp các dữ liệu phân tán lại với nhau một cách chính xác và xử lý nhanh.

(3) Truy vấn dữ liệu nhanh hơn: dữ liệu lớn được cập nhật liên tục, trong khi đó kho dữ liệu truyền thống thì lâu lâu mới được cập nhật và trong tình trạng không theo dõi thường xuyên, gây ra tình trạng lỗi cấu trúc truy vấn dẫn đến không tìm kiếm được thông tin đáp ứng theo yêu cầu.

(4) Độ chính xác cao hơn: dữ liệu lớn khi đưa vào sử dụng thường được kiểm định lại dữ liệu với những điều kiện chặt chẽ, số lượng thông tin được kiểm tra thông thường rất lớn, và đảm bảo về nguồn lấy dữ liệu không có sự tác động của con người vào thay đổi số liệu thu thập.

Ứng dụng dữ liệu lớn trong các lĩnh vực

Dữ liệu lớn đã được ứng dụng trong nhiều lĩnh vực như: hoạt động chính trị; giao thông; y tế; thể thao; tài chính; thương mại; thống kê... dưới đây là một số ví dụ về ứng dụng dữ liệu lớn.

(1) Ứng dụng dữ liệu lớn trong hoạt động chính trị



Hình trên cho thấy Tổng thống Mỹ Obama đã sử dụng dữ liệu dữ liệu lớn để phục vụ cho cuộc tranh cử Tổng thống của mình. Ông xây dựng một đội ngũ nhân viên chuyên đi thu thập thông tin và phân tích dữ liệu thu được trong dự án triển khai về dữ liệu lớn. Đội ngũ nhân viên này thu thập tất cả thông tin về người dân ở các khu vực, sau đó phân tích và chỉ ra một số thông tin quan trọng về người dân Mỹ như: Thích đọc sách gì, thích mua loại thuốc gì, thích sử dụng phương tiện gì... Thậm chí còn biết được cả thông tin về người đó đã bỏ phiếu tín nhiệm ai ở lần bầu cử trước. Trên cơ sở những thông tin này, Obama đã đưa ra kế hoạch vận động phù hợp, giúp ông tái đắc cử Tổng thống lần 2 của nước Mỹ.

Ngoài ra một số ứng dụng khác trong lĩnh vực chính trị mà dữ liệu lớn được áp dụng như: Hệ thống chính phủ điện tử; phân tích quy định và việc tuân thủ quy định; phân tích, giám sát, theo dõi và phát hiện gian lận, mối đe dọa, an ninh mạng.

(2) Ứng dụng dữ liệu lớn trong giao thông



Sử dụng số liệu CDR trong quá khứ để ước lượng các dòng giao thông trong thành phố vào các giờ cao điểm, từ đó có những kế hoạch phân luồng giao thông chi tiết, hợp lý giúp giảm thiểu kẹt xe. Ngoài ra còn đưa ra thông tin cho người tham gia giao thông được biết nếu muốn đi từ nơi này đến nơi khác thì nên đi vào giờ nào để tránh kẹt xe, hoặc đi đường nào là ngắn nhất v.v... Ngoài ra dữ liệu

lớn còn giúp phân tích định vị người dùng thiết bị di động, ghi nhận chi tiết cuộc gọi trong thời gian thực; và giảm thiểu tình trạng ùn tắc giao thông.

(3) Ứng dụng dữ liệu lớn trong y tế



Trong y học các bác sĩ dựa vào số liệu trong các bệnh án để đưa ra dự đoán về nguy cơ mắc bệnh. Đồng thời cũng đưa ra được xu hướng lây lan của bệnh. Ví dụ, ứng dụng Google Flu Trend là một trong những ứng dụng thành công của Google ứng dụng này dựa trên từ khóa tìm kiếm ở một khu vực nào đó, sau đó bộ máy phân tích của google sẽ phân tích và đối chiếu kết quả tìm kiếm đó, sau cùng là đưa ra dự báo về xu hướng dịch cúm tại khu vực đó. Qua đó cho biết tình hình cúm tại khu vực đó sẽ diễn ra như thế nào để đưa ra các giải pháp phòng tránh. Những kết quả mà Google Flu Trend đưa ra, hoàn toàn phù hợp với báo cáo của Tổ chức y tế thế giới WHO về tình hình bệnh cúm tại các khu vực đó.

(4) Ứng dụng dữ liệu lớn trong thể thao

Phân tích mô hình hệ thống cấu trúc sơ đồ chiến thuật của đội tuyển Đức (hình dưới)



đã đưa ra những điểm bất hợp lý trong cấu trúc của đội tuyển Đức, từ đó giúp cho đội tuyển Đức khắc phục được điểm yếu và đã dành được World cup 2014.

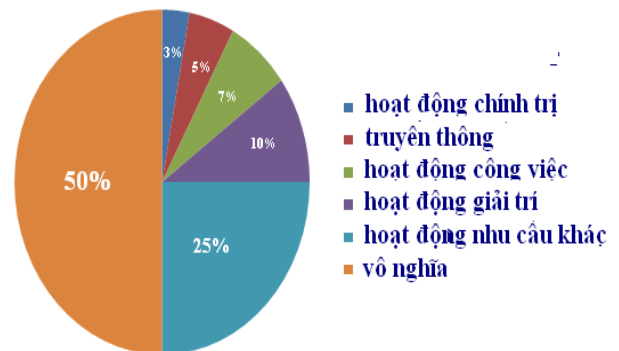
(5) Ứng dụng dữ liệu lớn trong tài chính

Từ những dữ liệu chính xác, kịp thời thu thập được thông qua các giao dịch của khách hàng, tiến hành phân tích, xếp hạng và quản lý các rủi ro trong đầu tư tài chính, tín dụng.

(6) Ứng dụng dữ liệu lớn trong thương mại

Trong thương mại dữ liệu lớn giúp cho chúng ta thực hiện được một số công việc sau: Phân khúc thị trường và khách hàng; phân tích hành vi khách hàng tại cửa hàng; tiếp thị trên nền tảng định vị; phân tích tiếp thị chéo kênh, tiếp thị đa kênh; quản lý các chiến dịch tiếp thị và khách hàng thân thiết; So sánh giá; Phân tích và quản lý chuỗi cung ứng; Phân tích hành vi, thói quen người tiêu dùng.

(7) Ứng dụng dữ liệu lớn trong thống kê



Nghiên cứu nội dung tin nhắn tại HÀ LAN

Một nghiên cứu thu thập thông tin qua các phương tiện truyền thông ở Hà Lan đã thống kê được nội dung các tin nhắn của người dân Hà Lan (hình trên), trong đó có đến 50% số tin nhắn vô nghĩa. Số liệu này cho thấy kinh phí đã chi ra không đem lại lợi ích cho người dân cũng như xã hội Hà Lan, Chính phủ cần phải có những điều chỉnh về hoạt động này.

Ủy ban Thống kê Liên hợp quốc đang triển khai nhiều hoạt động liên quan đến dữ liệu lớn với thống kê chính thức.

Dữ liệu lớn và thống kê chính thức

Mối quan tâm của thống kê chính thức đối với dữ liệu lớn: Nhận thấy những lợi ích (giảm thời gian và chi phí sản xuất thông tin thống kê) và những thách thức của dữ liệu lớn đối với thống kê chính thức, Ủy ban Thống kê Liên hợp quốc cũng như các tổ chức thống kê khu vực và cơ quan thống kê quốc gia của một số nước đã triển khai hàng loạt các hoạt động về dữ liệu lớn đối với thống kê chính thức. Cụ thể, Hội nghị cấp cao về Hợp lý hóa các dịch vụ và sản phẩm thống kê được tổ chức năm 2012⁴, đã chỉ ra những cơ hội và thách thức của dữ liệu lớn với thống kê chính thức. Dự án dữ liệu lớn⁵ được hình thành, một trong các sản phẩm của Dự án là Báo cáo về “Dữ liệu lớn như thế nào?” đã cho thấy vai trò tiềm năng của dữ liệu lớn đối với thống kê chính thức, đặc biệt là những thách thức và yêu cầu xét trên phương diện các phương pháp thống kê bao gồm chất lượng, công nghệ thông tin, năng lực và kỹ năng của đội ngũ nhân viên. Hội thảo về dữ liệu lớn của Thống kê châu Âu tại Rome (2014); Hội nghị quốc tế về dữ liệu lớn trong Thống kê chính thức tại Bắc Kinh (2014); Hội thảo vệ tinh UNECE NTTS 2015 về dữ liệu lớn ở Brussels (2015).

Một số báo cáo về dữ liệu lớn được trình bày trong các sự kiện quốc tế khác đang thu

⁴ 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>

⁵ UNECE (2014a) final project proposal: The role of Big data in the Modernisation of Statistical Production. <http://www1.unece.org/stat/platform/display/hlgbas>.

hút được sự quan tâm của nhà tổ chức và đại biểu tham dự, như: Hội nghị của Giám đốc các Cơ quan thống kê quốc gia DGINS 213 ở The Hague; Hội nghị Chất lượng Thống kê châu Âu 2014 (Q2014) ở Vienna; Hội nghị của Hiệp hội Quốc tế về Thống kê chính thức 2014 (IAOS 2014) ở Đà Nẵng; Đại hội Thống kê thế giới lần thứ 60 tại Rio de Janeiro. Đặc biệt, năm 2014, Ủy ban Thống kê Liên hợp quốc thành lập Nhóm công tác toàn cầu về dữ liệu lớn.

Nhóm công tác toàn cầu về dữ liệu lớn: Nhận thấy những lợi ích và thách thức của dữ liệu lớn đối với thống kê chính thức, tại cuộc họp lần thứ 45 (tháng 3/2014), Ủy ban thống kê Liên hợp quốc đã thành lập Nhóm công tác toàn cầu (Global Working Group - GWG) về dữ liệu lớn, nhằm nghiên cứu, giải quyết các vấn đề liên quan đến dữ liệu lớn như: 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 dữ liệu lớn...

Ngày 31/10/2014, phiên họp đầu tiên của GWG đã khẳng định việc sử dụng dữ liệu lớn cho các thống kê chính thức là một nghĩa vụ đối với cộng đồng thống kê 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 dữ liệu lớn 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. Các Tổ công tác đã khẩn trương triển khai thực hiện nhiệm vụ và đã đưa ra một số kết quả nghiên cứu bước đầu.

Một số kết quả nghiên cứu bước đầu về dữ liệu lớn đối với thống kê chính thức: Mặc

dù mới được hình thành trong vòng một năm nay, nhưng Nhóm công tác toàn cầu về dữ liệu lớn của Ủy ban Thống kê Liên hợp quốc đã công bố một số kết quả nghiên cứu về dữ liệu lớn đối với thống kê chính thức như: 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 dữ liệu lớn và mô hình cho thống kê chính thức, với ứng dụng của phân tích hình ảnh vệ tinh; Sử dụng dữ liệu lớn trong thống kê chính; dữ liệu lớn và thống kê chính thức – Thách thức của dữ liệu lớn trong Hệ thống Thống kê quốc gia; Quyền riêng tư, bảo mật và dữ liệu lớn; Thống kê trong thời đại thách thức hay cơ hội của dữ liệu lớn; Sự hợp nhất dữ liệu không gian và thời gian cho dữ liệu lớn và Ứng dụng của nó trong vệ tinh viễn thám.

Một số ứng dụng dữ liệu lớn trong thống kê chính thức như: Thống kê du lịch (Dữ liệu định vị của thiết bị di động trong công tác thống kê du lịch); thống kê sử dụng công nghệ thông tin và truyền thông; thống kê giá; thống kê thị trường lao động; thống kê vận tải; tổng điều tra nông nghiệp và các cuộc điều tra ở một số lĩnh vực khác.

Một số hoạt động tiếp theo:

- Nhóm công tác toàn cầu về dữ liệu lớn của Ủy ban Thống kê Liên hợp quốc tiếp tục nghiên cứu và ứng dụng dữ liệu lớn trong thống kê chính thức. Đặc biệt tập trung vào các chủ đề như: Vận động và truyền thông về dữ liệu lớn; liên kết dữ liệu lớn với các Mục tiêu Phát triển bền vững; đào tạo, xây dựng các kỹ năng và năng lực; các thách thức đối với dữ liệu lớn...

- Tổ chức một số cuộc hội thảo quốc tế và khu vực về dữ liệu lớn với thống kê chính thức sẽ được tổ chức vào cuối năm 2015, như: Hội nghị toàn cầu lần thứ hai về dữ liệu lớn đối với thống kê chính thức do Ủy ban Thống kê Liên hợp quốc sẽ tổ chức tại Abu Dhabi, Tiểu vương quốc Ả rập Thống nhất, từ ngày 20-22/10/2015⁶; Hội nghị IASC-ARS⁷ được tổ chức tại Singapore từ 17-19/12/2015 với chủ đề “Toán Thống kê: Cơ hội và thách thức với Dữ liệu lớn”.

- Cơ quan thống kê quốc gia của một số nước đang tập trung nghiên cứu dữ liệu lớn với thống kê chính thức như: Australia, Bangladesh, Cameroon, China, Colombia, Denmark, Egypt, Indonesia, Italy, Mexico, Morocco, Netherlands, Oman, Pakistan, Philippines, Tanzania, USA. Một số tổ chức quốc tế như: UNSD, UNECE, UNESCAP, UN Global Pulse, ITU, OECD, World Bank, Eurostat, GCC-stat. Đặc biệt Google đã và đang nghiên cứu mỗi liên hệ và điều tra thông qua việc sử dụng dữ liệu lớn trong hệ thống của mình.

Tóm lại

Dữ liệu lớn là một trong các chủ đề được quan tâm tại WSC lần thứ 60. Dữ liệu lớn mang lại nhiều lợi ích, nhưng cũng có nhiều thách thức khi triển khai, sử dụng dữ liệu lớn nói chung và dữ liệu lớn với thống kê chính thức nói riêng.

(Xem tiếp trang 6)

⁶ 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.

⁷ Cuộc Hội thảo này do Hiệp hội quốc tế về Toán thống kê tổ chức.

liên quan nghiên cứu, tìm hiểu về dữ liệu lớn và dữ liệu mở để có đề xuất nội dung, phạm vi, lộ trình sử dụng các nguồn dữ liệu này phục vụ công tác thống kê khi điều kiện kinh tế - xã hội ở nước ta cho phép.

Sáu là, Đề nghị Tổng cục giao Vụ Phương pháp chế độ Thống kê và Công nghệ thông tin chủ trì, phối hợp Trung tâm Tin học Thống kê khu vực I, Vụ Thống kê Tổng hợp và Trung tâm Tư liệu và Dịch vụ Thống kê nghiên cứu, vận dụng hệ thống thông tin địa lý vào tổng hợp và phổ biến thông tin thống kê của một số chỉ tiêu thống kê trong Niên giám thống kê hàng năm để đưa lên Trang thông tin điện tử của Tổng cục. Trên cơ sở đó có kế hoạch đẩy mạnh áp dụng kỹ thuật này ở các

ngành vụ thống kê chuyên ngành.

Bảy là, Đề nghị Tổng cục giao Vụ Thống kê nước ngoài và Hợp tác quốc tế liên hệ với Cơ quan Thống kê Trung ương Palestine để bàn việc hợp tác song phương giữa hai cơ quan thống kê.

Tám là, Đề nghị Tổng cục giao Vụ Thống kê Nông, Lâm nghiệp và Thủy sản làm việc với Văn phòng FAO tại Hà Nội về việc hỗ trợ kỹ thuật trong Tổng điều tra nông thôn, nông nghiệp và thủy sản năm 2016

Chín là, Đề nghị Tổng cục tham khảo kinh nghiệm của New Zealand xây dựng các phần mềm đào tạo thống kê trực tuyến để thí điểm tại Tổng cục.

(Tiếp theo trang 12)

Tại kỳ họp lần thứ 44 của Ủy ban Thống kê Liên hợp quốc được tổ chức vào tháng 2/2013 tại New York, phiên họp với tiêu đề "Big Data for Policy, Development and Official Statistics" (tạm dịch là "Big Data đối với chính sách, phát triển và thống kê nhà nước"). Tiếp theo đó Ủy ban Thống kê Liên hợp quốc đã hình thành Nhóm công tác toàn cầu về dữ liệu lớn. Nhóm này đã công bố một số kết quả ban đầu về dữ liệu lớn với thống kê chính thức. Hoạt động tiếp theo của Nhóm này sẽ đưa ra các hướng dẫn về dữ liệu lớn đối với thống kê chính thức. Một số cơ quan thống kê quốc gia của một số nước và tổ chức quốc tế cũng đang tập trung nghiên cứu, sử dụng dữ liệu lớn.

Đối với Việt Nam nói chung, Tổng cục Thống kê nói riêng, dữ liệu lớn là vấn đề mới. Tổng cục Thống kê cần giao một đơn vị của Tổng cục chủ trì tiếp cận các nghiên cứu của Nhóm công tác toàn cầu về dữ liệu lớn và triển khai áp dụng một cách bài bản ở Việt Nam. Dữ liệu lớn đối với thống kê chính thức được triển khai ở nước ta, không chỉ giải quyết được vấn đề chất lượng và hiệu quả của hoạt động thống kê, mà còn thể hiện mức độ hội nhập với thống kê quốc tế.

Tài liệu tham khảo:

[1] Wikipedia;

https://en.wikipedia.org/wiki/Big_data