

ĐẢM BẢO CHẤT LƯỢNG TRONG HỆ SINH THÁI DỮ LIỆU MỚI: Lưu ý khoảng cách giữa dữ liệu và số liệu thống kê

Matthias Reister,

United Nations Statistics Division, New York, NY, USA

E-mail: reister@un.org

Tóm tắt

Dựa trên công việc gần đây nhằm phát triển Sổ tay Khung Đảm bảo Chất lượng Quốc gia về Thống kê Chính thức của Liên hợp quốc nhằm đáp ứng hệ sinh thái dữ liệu mới, bài viết này giải quyết ba câu hỏi quan trọng mà cộng đồng thống kê hiện đang phải đối mặt: (1) Làm thế nào số liệu thống kê chính thức có thể đảm bảo chất lượng dữ liệu từ nguồn hành chính và các nguồn khác? (2) Khung đảm bảo chất lượng cho số liệu thống kê chính thức có thể được áp dụng cho dữ liệu thay vì số liệu thống kê không? (3) Sự khác biệt giữa dữ liệu và số liệu thống kê có ý nghĩa gì khác đối với vai trò của số liệu thống kê chính thức trong hệ sinh thái dữ liệu mới? Bài viết lập luận rằng các cơ quan thống kê nên ủng hộ mạnh mẽ việc thành lập cơ quan quản lý dữ liệu quốc gia nhưng bản thân họ không nên đảm nhận vai trò đó. Việc lẫn lộn trách nhiệm về dữ liệu và số liệu thống kê chính thức có nguy cơ làm suy yếu số liệu thống kê chính thức và không đáp ứng

nhu cầu phát triển dữ liệu như một tài sản một cách có trách nhiệm.

1. Nhận thức được tác động của hệ sinh thái dữ liệu mới đối với số liệu thống kê chính thức

Việc công bố báo cáo “Cách mạng dữ liệu”, “A World That Counts” vào tháng 11 năm 2014 bởi Nhóm chuyên gia tư vấn độc lập của Tổng thư ký Liên hợp quốc về Cách mạng dữ liệu vì sự phát triển bền vững đã đánh dấu một cột mốc và bước ngoặt cho sự phát triển của thống kê chính thức¹. Nó nêu rõ nhu cầu thúc đẩy đổi mới để lấp đầy khoảng trống dữ liệu bằng cách kết hợp các

¹ Số liệu thống kê chính thức mô tả, trên cơ sở đại diện, các hiện tượng kinh tế, nhân khẩu học, xã hội và môi trường được công chúng quan tâm. Số liệu thống kê chính thức được các thành viên của hệ thống thống kê quốc gia phát triển, sản xuất và phổ biến như một loại hàng hóa công, tuân thủ các Nguyên tắc cơ bản của Thống kê chính thức và các khung chất lượng được chấp nhận cũng như các tiêu chuẩn và khuyến nghị thống kê được quốc tế công nhận khác.

nguồn dữ liệu truyền thống và mới (bao gồm cả Dữ liệu lớn) và tạo cơ sở hạ tầng mới để phát triển và chia sẻ dữ liệu, đồng thời phát triển các tiêu chuẩn đạo đức, pháp lý và thống kê toàn cầu để cải thiện chất lượng dữ liệu và bảo vệ mọi người khỏi bị lạm dụng trong hệ sinh thái dữ liệu đang thay đổi nhanh chóng.

Báo cáo và các cuộc thảo luận tiếp theo đã nêu bật mức độ tăng trưởng chưa từng có về dữ liệu và số lượng khía cạnh trong cuộc sống của chúng ta đang được ghi lại trong dữ liệu. Cụ thể, các công nghệ mới cho phép ghi lại, phân tích và tích hợp dữ liệu ở mức độ không thể tưởng tượng được trước đây về hành vi của con người và xu hướng xã hội tổng thể. Dữ liệu là thành phần chính của nhiều mô hình kinh doanh và nhiều doanh nghiệp đang quản lý và sử dụng lượng dữ liệu rất lớn, bao gồm cả dữ liệu có thể liên kết với các cá nhân. Nhận thức được tầm quan trọng của dữ liệu và nguy cơ lạm dụng dữ liệu, nhiều quốc gia, bao gồm Vương quốc Anh, Hoa Kỳ, Trung Quốc và Liên minh Châu Âu đã áp dụng các chiến lược và chiến lược dữ liệu quốc gia đã thông qua luật để quản lý việc thu thập, lưu trữ và sử dụng dữ liệu ở cấp quốc gia và xuyên biên giới.

Sự xuất hiện của hệ sinh thái dữ liệu mới này, có thể được định nghĩa là một hệ thống trong đó một số tác nhân tương tác để trao đổi, sản xuất và sử dụng dữ liệu, đưa ra cả cơ hội và thách thức đối với số liệu thống kê chính thức. Các nguồn dữ liệu mới có thể nâng cao hiệu quả hoặc tính kịp thời của số liệu thống kê chính thức bằng cách sử dụng dữ liệu sẵn có. Các cơ quan thống kê có thể tự mình tạo ra ít dữ liệu hơn và có thể thường xuyên trở thành người quản lý các số liệu thống kê chính thức do người khác tạo ra. Đồng thời, vai trò của Cơ quan Thống kê

Quốc gia (NSO)² và các nhà sản xuất số liệu thống kê chính thức khác với tư cách là nhà cung cấp dữ liệu và số liệu thống kê chủ yếu cho các nhà hoạch định chính sách và công chúng có thể bị thách thức vì người dùng có thể sử dụng dữ liệu và số liệu thống kê được tạo ra bên ngoài hệ thống thống kê quốc gia hoặc thậm chí do chính họ tạo ra. Các cơ quan thống kê đã ứng phó với thách thức này bằng cách nhấn mạnh giá trị của thống kê chính thức như một nguồn thông tin chất lượng đáng tin cậy và các giá trị cốt lõi của thống kê chính thức, bằng cách hiện đại hóa quy trình sản xuất thông tin thống kê và sử dụng dữ liệu, đồng thời bằng cách mở rộng việc cung cấp dữ liệu, phân tích và thông tin. Đồng thời, nhiều nhà lãnh đạo của NSO đang xem xét lại vai trò tổng thể của số liệu thống kê chính thức trong hệ sinh thái dữ liệu mới và thảo luận về vai trò có thể có trong tương lai của tổ chức của họ trong việc quản lý dữ liệu quốc gia³.

2. Hướng dẫn của Liên hợp quốc về đảm bảo chất lượng số liệu thống kê chính thức đáp ứng hệ sinh thái dữ liệu mới như thế nào

Tại phiên họp thứ 50 vào tháng 3 năm 2019, Ủy ban Thống kê Liên hợp quốc (LHQ) đã thông qua Sổ tay Khung đảm bảo chất lượng quốc gia của Liên hợp quốc về Thống kê chính thức (Sổ tay). Sổ tay đã cập nhật và thay thế mẫu và hướng dẫn ban hành năm 2012. Các khuyến nghị về đảm bảo chất lượng thống kê, khung đảm bảo chất lượng

² Ở đây được hiểu là có ý nghĩa tương tự như viện thống kê quốc gia

³ Phiên họp toàn thể lần thứ 67 của Hội nghị các nhà thống kê châu Âu tại Paris vào ngày 26-28 tháng 6 năm 2019 có thể được coi là sự khởi đầu của cuộc thảo luận rộng hơn về quản lý dữ liệu trong cộng đồng thống kê chính thức vẫn tiếp tục kể từ đó.

➤ ➤ ➤ THỐNG KÊ QUỐC TẾ VÀ HỘI NHẬP

thống kê được cập nhật và các hướng dẫn triển khai đều nhằm mục đích giải quyết các yêu cầu của hệ sinh thái dữ liệu mới.

2.1. Hai khuyến nghị cốt lõi đề cập đến phạm vi đảm bảo chất lượng của số liệu thống kê chính thức

Sổ tay quy định rằng tất cả các thành viên của hệ thống thống kê quốc gia phải áp dụng khung đảm bảo chất lượng quốc gia cho tất cả dữ liệu và số liệu thống kê được sử dụng cho việc ra quyết định của chính phủ. Đặc biệt,

1. “Khuyến nghị các nước thiết lập khung đảm bảo chất lượng quốc gia về thống kê chính thức và tất cả các thành viên của hệ thống thống kê quốc gia cam kết liên tục đánh giá, cải thiện và báo cáo về chất lượng thống kê chính thức, cũng như chất lượng dữ liệu và số liệu thống kê được sử dụng trong việc sản xuất số liệu thống kê chính thức theo yêu cầu” (Khuyến nghị 3).

2. “Khuyến nghị thực hiện khung đảm bảo chất lượng quốc gia tại cơ quan thống kê quốc gia và trong toàn bộ hệ thống thống kê quốc gia. Hơn nữa, nên áp dụng khung đảm bảo chất lượng quốc gia cho tất cả dữ liệu và số liệu thống kê được tạo ra bên ngoài hệ thống thống kê quốc gia được phổ biến với sự giúp đỡ và hỗ trợ của một thành viên trong hệ thống thống kê quốc gia hoặc được sử dụng để ra quyết định của chính phủ, nếu thấy phù hợp và cần thiết” (Khuyến nghị 5).

2.2 Những cập nhật quan trọng đối với khung đảm bảo chất lượng của Liên hợp quốc

Chương 3 và Phụ lục của Sổ tay cập nhật Khung đảm bảo chất lượng quốc gia của Liên hợp quốc (UN-NQAF) từ phiên bản trước năm

2012 để phản ánh nhu cầu tích cực sử dụng các nguồn dữ liệu mới:

1. Theo Nguyên tắc 1 “Điều phối hệ thống thống kê quốc gia”, đề nghị 1.3 yêu cầu “có cơ chế xem xét số liệu thống kê được tạo ra bên ngoài hệ thống thống kê quốc gia và nếu phù hợp, để những số liệu thống kê này trở thành chính thức”.

2. Theo Nguyên tắc 2 “Quản lý mối quan hệ với người dùng dữ liệu, nhà cung cấp dữ liệu và các bên liên quan khác”, ba yêu cầu giải quyết nhu cầu có được quyền truy cập và sử dụng đầy đủ dữ liệu và dữ liệu hành chính (bao gồm cả “Dữ liệu lớn”) do các công ty tư nhân duy trì hoặc các tổ chức phi chính phủ khác nhằm mục đích thống kê một cách thường xuyên, bao gồm cả việc thử nghiệm và thử nghiệm⁴. NSO cũng nên hợp tác và cung cấp hỗ trợ, hướng dẫn cho các nhà cung cấp dữ liệu.

3. Các nguyên tắc liên quan đến quản lý quy trình thống kê kêu gọi thúc đẩy đổi mới trong phát triển, sản xuất và phổ biến số liệu thống kê (yêu cầu 10.5), nỗ lực cải thiện tiềm năng thống kê của dữ liệu hành chính và các nguồn dữ liệu khác (yêu cầu 12.3) và chia sẻ dữ liệu, liên kết dữ liệu và sử dụng các nguồn dữ liệu hành chính và dữ liệu khác để giảm thiểu gánh nặng cho người trả lời (yêu cầu 13.4).

4. Theo Nguyên tắc 14 “Đảm bảo tính phù hợp”, yêu cầu 14.3 là “số liệu thống kê dựa trên các nguồn dữ liệu mới và hiện có đang được phát triển để đáp ứng nhu cầu thông tin mới nổi của xã hội”.

⁴ Các khung đảm bảo chất lượng khác cho số liệu thống kê chính thức như Quy tắc thực hành thống kê châu Âu cũng giải quyết nhu cầu truy cập dữ liệu.

2.3. Thảo luận trong Sổ tay về việc sử dụng các nguồn dữ liệu khác

Sổ tay bao gồm một số chương về việc thực hiện khung đảm bảo chất lượng quốc gia trong toàn bộ hệ thống thống kê quốc gia. Cụ thể, Chương 7 đề cập đến vấn đề đảm bảo chất lượng khi sử dụng các nguồn dữ liệu khác để đưa ra số liệu thống kê chính thức. Nó phân biệt giữa các nguồn dữ liệu thống kê, nguồn dữ liệu hành chính và các nguồn dữ liệu khác theo mục đích của chúng và đơn vị chịu trách nhiệm tổng hợp dữ liệu và thảo luận về việc sử dụng nhiều nguồn dữ liệu. Hướng dẫn xác định ba nguồn dữ liệu khác nhau này như sau:

1. Nguồn dữ liệu thống kê là các bộ sưu tập dữ liệu được tạo ra chủ yếu cho mục đích thống kê chính thức của các cơ quan Chính phủ hoặc các đơn vị khác làm việc thay mặt Chính phủ. Nguồn dữ liệu thống kê bao gồm điều tra thống kê, tổng điều tra và đăng ký thống kê. Các nguồn dữ liệu thống kê thường được gọi là nguồn dữ liệu sơ cấp trong khi các nguồn hành chính và các nguồn khác là nguồn thứ cấp.

2. Nguồn dữ liệu hành chính là các bộ dữ liệu được tạo ra chủ yếu cho mục đích hành chính của các cơ quan Chính phủ hoặc các đơn vị khác làm việc thay mặt Chính phủ. Nguồn dữ liệu hành chính bao gồm đăng ký hành chính của cá nhân, pháp nhân và hồ sơ của các bộ, ngành, cơ quan chuyên môn như tờ khai thuế, hồ sơ dịch vụ xã hội và dữ liệu hải quan hoặc dữ liệu của chính quyền khu vực hoặc địa phương.

3. Các nguồn dữ liệu khác bao gồm tất cả các bộ dữ liệu không được tạo ra chủ yếu cho mục đích thống kê hoặc hành chính chính thức mà nhằm mục đích thương mại hoặc mục đích riêng tư khác. Chúng bao gồm các

bộ dữ liệu được tạo bởi các nhà cung cấp dịch vụ truyền thông, phương tiện truyền thông và thương mại điện tử, nhà cung cấp dịch vụ dựa trên quan sát Trái đất và viễn thám, và các công ty bảo hiểm tư nhân, nhưng cũng thông qua các điều tra mẫu truyền thống do các công ty thực hiện cho mục đích riêng của họ, chẳng hạn như khảo sát thị trường, nghiên cứu.

Sổ tay đưa ra quan điểm rằng UN-NQAF và các khung đảm bảo chất lượng quốc gia chung khác áp dụng cho việc sản xuất số liệu thống kê chính thức bất kể nguồn dữ liệu. Tuy nhiên, người ta nhận thấy rằng những thách thức để đạt được sự tuân thủ có thể khác nhau tùy thuộc vào nguồn dữ liệu.

3. Các phương pháp đánh giá chất lượng nguồn dữ liệu để đưa ra số liệu thống kê chính thức

Nhu cầu đánh giá chất lượng của các nguồn dữ liệu hành chính và các nguồn dữ liệu khác để đưa ra số liệu thống kê chính thức ngày càng trở nên cần thiết và ngày càng tăng. Tuy nhiên, việc đánh giá chất lượng của nguồn dữ liệu là một phần nhưng cũng khác với việc đánh giá chất lượng của một sản phẩm thống kê. Người ta có thể coi việc đánh giá nguồn dữ liệu là một bước riêng biệt có thể cần được chú ý đặc biệt khi sử dụng các nguồn dữ liệu hành chính và các nguồn dữ liệu khác.

3.1. Những thách thức chất lượng điển hình trong việc sử dụng các nguồn dữ liệu hành chính và các nguồn dữ liệu khác

Sổ tay xác định nhiều thách thức trong việc sử dụng các nguồn dữ liệu hành chính hoặc các nguồn dữ liệu khác để tạo ra số liệu thống kê chính thức, bao gồm các vấn đề về truy cập và phối hợp, sử dụng không đúng

➤ ➤ ➤ THỐNG KÊ QUỐC TẾ VÀ HỘI NHẬP

các khái niệm thống kê, thiếu minh bạch, nhu cầu đảm bảo các thủ tục thống kê phù hợp và tôn trọng các nguyên tắc bảo mật, mức độ liên quan, độ chính xác và độ tin cậy (đặc biệt là về mức độ bao phủ), khả năng so sánh và siêu dữ liệu. Ví dụ: dữ liệu viễn thông có thể chỉ ra sự di chuyển của con người, nhưng chúng ta thường không biết mức độ đại diện của chúng đối với tổng thể và do đó phạm vi bao phủ của dữ liệu cũng như các vấn đề so sánh có thể phát sinh từ việc thay đổi mô hình sử dụng dịch vụ viễn thông. Cốt lõi của những thách thức này thực tế là các nguồn dữ liệu hành chính và các nguồn dữ liệu khác không hướng tới việc tạo ra số liệu thống kê chính thức. Các cơ quan thống kê chỉ có ảnh hưởng hạn chế hoặc không có ảnh hưởng gì đến dữ liệu nào đang được tổng hợp, không giống như trường hợp các nguồn dữ liệu thống kê được định hướng cụ thể cho việc sản xuất số liệu thống kê chính thức. Chương 7 của Sổ tay cung cấp danh sách các ví dụ về các yếu tố cụ thể cần được đảm bảo khi sử dụng nguồn dữ liệu thống kê, hành chính, dữ liệu khác hoặc nhiều nguồn dữ liệu.

3.2. Các phương pháp đánh giá chất lượng nguồn dữ liệu

Phần này giới thiệu ngắn gọn một số cách tiếp cận để đánh giá chất lượng của nguồn dữ liệu nhằm sử dụng chúng để tạo ra số liệu thống kê chính thức⁵. Những cách tiếp cận này cũng có thể phù hợp để đánh giá chất lượng dữ liệu nói chung - mỗi quan tâm thường được người sử dụng và nhà sản xuất

⁵ Phần này được phỏng theo một phần từ công trình của Hans Viggo Saeboe với sự hỗ trợ của Nhóm chuyên gia Liên hợp quốc về Khung đảm bảo chất lượng quốc gia và được trình bày tại một sự kiện bên lề hội nghị lần thứ 53 phiên họp của Ủy ban Thống kê Liên Hợp Quốc vào ngày 10 tháng 2 năm 2022.

dữ liệu trong chính phủ, xã hội dân sự và khu vực tư nhân nêu ra.

Dự án BLUE-ETS Châu Âu đã cung cấp cái nhìn tổng quan về các phương pháp đo lường để đánh giá chất lượng dữ liệu hành chính khi được sử dụng làm đầu vào cho số liệu thống kê chính thức. Các chỉ tiêu của nó được nhóm theo năm khía cạnh chất lượng: kiểm tra kỹ thuật (khả năng sử dụng kỹ thuật của tệp và dữ liệu của nó), độ chính xác (dữ liệu chính xác, đáng tin cậy và được chứng nhận), tính đầy đủ (quá phạm vi hay thiếu phạm vi), các khía cạnh về thời gian (bao gồm cả tính kịp thời) và khả năng tích hợp (mức độ mà dữ liệu nguồn có thể được liên kết với dữ liệu hành chính và dữ liệu khác). Ngoài ra, dự án đã xác định các điều kiện tiên quyết để sử dụng như tiếp cận pháp lý, sự chấp nhận của công chúng, sự sẵn có của hệ thống nhận dạng thống nhất, hệ thống hành chính công toàn diện và đáng tin cậy cũng như sự hợp tác giữa các cơ quan chức năng phải được đáp ứng.

Tiếp nối công việc trước đó, Ủy ban Kinh tế Liên hợp quốc về Châu Âu đã công bố hướng dẫn đánh giá chất lượng dữ liệu hành chính để sử dụng trong các cuộc tổng điều tra. Chúng xem xét chất lượng ở bốn giai đoạn: ở nguồn, ở giai đoạn NSO nhận được dữ liệu, ở giai đoạn xử lý và ở giai đoạn đầu ra. Các khía cạnh chất lượng tại nguồn bao gồm mức độ phù hợp và độ chính xác (để sử dụng trong tổng điều tra), tính kịp thời, tính minh bạch và khả năng so sánh, khả năng tiếp cận (sự dễ dàng mà NSO có thể lấy được dữ liệu) và khả năng diễn giải (tính sẵn có của siêu dữ liệu). Môi trường thể chế của chủ sở hữu dữ liệu (bao gồm cả năng lực của nó) cũng được đưa vào. Các khía cạnh chất lượng ở giai đoạn nhận dữ liệu là xác thực và hài

hòa (ví dụ: định dạng tệp có thể đọc được), độ chính xác và độ tin cậy (đối với các biến và phạm vi tổng thể), tính kịp thời và đúng giờ cũng như khả năng liên kết (ví dụ: mã định danh duy nhất chung). Nguồn và các bước tiếp nhận dữ liệu cùng nhau đưa ra đánh giá về chất lượng đầu vào.

Những nỗ lực quan trọng khác là phát triển danh sách kiểm tra để đánh giá chất lượng dữ liệu đầu vào như một phần của chương trình Eurostat ESS Vision 2020 ADMIN. Những danh sách kiểm tra này xác định một danh sách tổng hợp gồm sáu khía cạnh chất lượng và 17 chỉ tiêu liên quan. Ngoài ra, vào năm 2014, Nhóm đặc nhiệm chất lượng dữ liệu lớn của UNECE đã đề xuất Khung chất lượng của dữ liệu lớn sử dụng cấu trúc phân cấp bao gồm ba siêu chiều: nguồn, siêu dữ liệu và dữ liệu. Gần đây hơn, De Broe et al đã thảo luận về các tiêu chí chất lượng để tích hợp dữ liệu và phương pháp mới trong thống kê chính thức và kết luận rằng các nguyên tắc chất lượng của Hệ thống Thống kê Châu Âu như được quy định trong Quy tắc Thực hành Thống kê Châu Âu và do đó cả trong UN-NQAF, không cần phải điều chỉnh đến sự xuất hiện của các nguồn dữ liệu mới. De Broe và cộng sự lập luận rằng các khía cạnh chất lượng mới liên quan đến nguồn dữ liệu lớn có thể được tích hợp vào các khung chất lượng hiện có ở mức dưới các nguyên tắc chất lượng tương ứng.

3.3. Kết luận cho công việc tiếp theo

Tóm lại, công việc quan trọng đã được thực hiện nhằm hướng dẫn đánh giá chất lượng của nguồn dữ liệu nhằm đưa ra số liệu thống kê chính thức. Có những điểm tương đồng giữa các phương pháp này như việc sử dụng các nguyên tắc chất lượng của thống kê chính thức liên quan đến kết quả thống kê và

tập trung vào các đặc điểm của tập dữ liệu. Tuy nhiên, tại thời điểm này, chưa có câu trả lời duy nhất hoặc thống nhất về cách đảm bảo chất lượng dữ liệu từ các nguồn hành chính và các nguồn khác để đưa ra số liệu thống kê chính thức. Có thể điều này là không thể tránh khỏi, nhưng sẽ có lợi nếu có một số danh sách các tiêu chí và/hoặc chỉ tiêu và hướng dẫn cung cấp ít nhất một điểm khởi đầu hoặc điểm tham chiếu chung cho tất cả những người quan tâm đến chất lượng của dữ liệu nguồn từ các nguồn dữ liệu hành chính hoặc các nguồn dữ liệu khác.

4. Tại sao khung đảm bảo chất lượng cho số liệu thống kê chính thức không thể áp dụng trực tiếp cho dữ liệu

4.1. Sự khác biệt giữa dữ liệu và số liệu thống kê

Thông thường, có những lời kêu gọi giải quyết "chất lượng dữ liệu và số liệu thống kê" và chỉ sử dụng "dữ liệu chất lượng" để ra quyết định. Bài viết này lập luận rằng việc phân biệt giữa dữ liệu và số liệu thống kê là rất quan trọng và hữu ích khi thảo luận về việc đảm bảo chất lượng số liệu thống kê chính thức. Thông thường, thuật ngữ "dữ liệu" và "số liệu thống kê" được sử dụng đồng nghĩa hoặc số liệu thống kê được xem như một phần của tập hợp dữ liệu lớn hơn. Hướng dẫn sử dụng chính xác hơn. Nó định nghĩa số liệu thống kê là thông tin số liên quan đến tổng hợp dữ liệu về các đơn vị hoặc quan sát. Nó sử dụng thuật ngữ "số liệu thống kê" khi đề cập đến đầu ra của một quy trình sản xuất thông tin thống kê và thuật ngữ "dữ liệu" khi đề cập đến đầu vào hoặc có thể là thông lượng trong quy trình đó. Thuật ngữ "microdata" là một trường hợp đặc biệt vì tùy theo ngữ cảnh, nó cũng có thể là đầu ra. Thuật ngữ "dữ liệu" không chỉ bao gồm

➤ ➤ ➤ THỐNG KÊ QUỐC TẾ VÀ HỘI NHẬP

dữ liệu được tạo ra và sử dụng cho mục đích thống kê mà còn bao gồm vô số dữ liệu lớn hơn vô cùng do khu vực công và tư nhân tạo ra thông qua việc sử dụng công nghệ cho nhiều mục đích khác nhau và có khả năng sẵn có cho nhiều mục đích sử dụng.

4.2. Đảm bảo chất lượng được thúc đẩy bởi nhu cầu của người dùng

Chúng ta hãy nhớ lại định nghĩa tiêu chuẩn về chất lượng. Chất lượng là mức độ mà một tập hợp các đặc tính vốn có của một đối tượng đáp ứng được các yêu cầu. Một định nghĩa đơn giản về chất lượng là “phù hợp để sử dụng” hoặc “phù hợp với mục đích”. Chính nhu cầu của người dùng quyết định chất lượng. Nhu cầu của người dùng có thể khác nhau và phải được dung hòa.

UN-NQAF hướng tới việc đảm bảo chất lượng số liệu thống kê chính thức - tổng hợp là sản phẩm cuối cùng của quá trình sản xuất số liệu thống kê. Nhu cầu của người dùng liên quan đến số liệu thống kê chính thức phần lớn được gói gọn trong Nguyên tắc 1 của Nguyên tắc cơ bản của Thống kê chính thức (FPOS), nói về nhu cầu đáp ứng quyền của công dân đối với thông tin công cộng liên quan đến tình hình kinh tế, nhân khẩu học, xã hội và môi trường. Tất cả số liệu thống kê chính thức được tạo ra nhằm mục đích cung cấp cho người dùng thông tin thống kê cụ thể. Các khung đảm bảo chất lượng cho số liệu thống kê chính thức như UN-NQAF chỉ đánh giá chất lượng của dữ liệu đầu vào hoặc dữ liệu nguồn từ góc độ mục đích cuối cùng là tạo ra các kết quả thống kê cụ thể để đáp ứng nhu cầu của người dùng.

Nhu cầu của người dùng liên quan đến dữ liệu rộng hơn nhiều và tất cả dữ liệu đều được *ưu tiên* sử dụng nhiều lần. Dữ liệu có

thể hỗ trợ mọi hành động và tương tác có thể tưởng tượng được của con người, bao gồm sản xuất, trao đổi và tiêu thụ hàng hóa và dịch vụ tư nhân và công cộng, bất kể dữ liệu ban đầu có được dự định cho các mục đích này hay không. Do đó, các yêu cầu hoặc sự phù hợp để sử dụng dữ liệu cho những mục đích khác nhau đó có thể rất khác với các yêu cầu để đưa ra số liệu thống kê chính thức.

4.3. Khung đảm bảo chất lượng cho số liệu thống kê chính thức không thể được áp dụng trực tiếp vào dữ liệu

Do tính chất khác nhau của nhu cầu của người dùng, các khung đảm bảo chất lượng thống kê như UN-NQAF không hướng tới việc đánh giá và cải thiện chất lượng dữ liệu được sử dụng cho các mục đích phi thống kê. Chúng không thể được áp dụng trực tiếp vào dữ liệu nói chung. Điều này được minh họa trong các ví dụ sau:

1. Nhiều yêu cầu được phản ánh trong các nguyên tắc 14-18 của UN-NQAF liên quan đến mức độ phù hợp, độ chính xác và độ tin cậy, tính kịp thời và đúng hạn, khả năng tiếp cận và tính rõ ràng, tính mạch lạc và khả năng so sánh của các kết quả thống kê cũng liên quan đến dữ liệu. Tuy nhiên, chúng phải được điều chỉnh hoặc áp dụng khác nhau tùy thuộc vào mục đích sử dụng dữ liệu. Ví dụ, nguyên tắc liên quan áp dụng cho số liệu thống kê chính thức giải quyết câu hỏi liệu kết quả thống kê có đáp ứng được nhu cầu thông tin hiện tại của người sử dụng hay không. Ngược lại, nhu cầu của người sử dụng *dữ liệu* có thể khác nhau, rất khác nhau và phần lớn chúng ta, những nhà thống kê, không biết đến. Chúng ta có thể biết một số nhưng không phải tất cả người dùng. Do đó, có thể khó hoặc không thể xác định và cân bằng nhu cầu dữ liệu của người dùng, trong

khi việc sử dụng và sử dụng số liệu thống kê chính thức thường đã được xác định trong thiết kế quy trình sản xuất thông tin thống kê. Một ví dụ khác là yêu cầu về số liệu thống kê chính thức để tiến hành các nghiên cứu sửa đổi nhằm đảm bảo tính chính xác và độ tin cậy. Rõ ràng các nghiên cứu sửa đổi có ý nghĩa đối với kết quả thống kê, nhưng lại không rõ ràng lắm về ý nghĩa của chúng đối với dữ liệu, trong nhiều trường hợp chỉ được tạo ra một lần.

2. Tương tự, một số yêu cầu chung và các phương pháp hay nhất để quản lý quy trình thống kê cũng có thể áp dụng cho dữ liệu. Ví dụ, nguyên tắc 10 của UN-NQAF chỉ ra sự cần thiết phải đánh giá dữ liệu nguồn và cải thiện các phương pháp cũng như thúc đẩy đổi mới để đảm bảo tính đúng đắn về phương pháp luận và nguyên tắc 12 thừa nhận nhu cầu tạo điều kiện thuận lợi cho việc liên kết dữ liệu và sử dụng mã định danh duy nhất cho các đơn vị thống kê. Tuy nhiên, nhiều yêu cầu của thống kê chính thức như việc sử dụng các tiêu chuẩn quốc tế hoặc quy trình thống kê không có hoặc không có khả năng áp dụng rõ ràng đối với dữ liệu.

3. Một trong những nguyên tắc chính của thống kê chính thức (và một trong những giá trị cốt lõi của nó) là bảo mật thống kê, dựa trên Nguyên tắc 6 của FPOS trong đó nêu rõ rằng dữ liệu chỉ được sử dụng cho mục đích thống kê. Trong thực tế, luật thống kê có thể cấp quyền truy cập vào dữ liệu cấp độ cá nhân cho mục đích nghiên cứu nếu đáp ứng các điều kiện cụ thể. Ngược lại, mục tiêu liên quan đến dữ liệu được phản ánh trong các chiến lược dữ liệu quốc gia khác nhau không phải là hạn chế và ngăn chặn mà là khuyến khích các cách sử dụng khác nhau đồng thời ngăn chặn việc lạm dụng. Thông thường,

chính thông tin về các cá nhân hoặc đơn vị thống kê riêng lẻ mới làm cho dữ liệu trở nên hữu ích và có giá trị. Do đó, bảo mật dữ liệu thống kê có thể và thường gây bất lợi cho việc sử dụng dữ liệu.

4.4. Kết luận cho công việc tiếp theo ngoài số liệu thống kê chính thức

Mặc dù các khung đảm bảo chất lượng cho số liệu thống kê chính thức không thể áp dụng trực tiếp vào dữ liệu nhưng nỗ lực đánh giá chất lượng của dữ liệu nguồn để tạo ra số liệu thống kê chính thức như được mô tả trong Phần 3 có thể tạo cơ sở cho việc phân tích các yêu cầu về chất lượng của dữ liệu nói chung. Tuy nhiên, công việc như vậy cần phải bắt đầu bằng việc nghiên cứu những gì đã được thực hiện trong khu vực công và tư và được thực hiện với sự hợp tác của những người nắm giữ và sử dụng dữ liệu tương ứng. Trong bối cảnh này, điều quan trọng cần đề cập là Vương quốc Anh (Anh) đã phát triển Khung chất lượng dữ liệu của chính phủ không tập trung vào số liệu thống kê chính thức mà tập trung vào dữ liệu để hỗ trợ tham vọng của chính phủ liên quan đến chuyển đổi kỹ thuật số các dịch vụ công và Vương quốc Anh trở thành quốc gia dẫn đầu thế giới về sử dụng trí tuệ nhân tạo Khung này dựa trên năm nguyên tắc sau: (1) Cam kết chất lượng dữ liệu, (2) Biết người dùng của bạn và nhu cầu của họ, (3) Đánh giá chất lượng trong suốt vòng đời dữ liệu, (4) Truyền đạt chất lượng dữ liệu rõ ràng và hiệu quả và (5) Dự đoán những thay đổi ảnh hưởng đến chất lượng dữ liệu. Cách tiếp cận của Vương quốc Anh trái ngược đáng kể với báo cáo của OECD về con đường trở thành khu vực công dựa trên dữ liệu, vốn coi sự phù hợp với mục đích của dữ liệu phần lớn là đương nhiên và thay vào đó tập trung vào quản trị dữ liệu và

➤ ➤ ➤ THỐNG KÊ QUỐC TẾ VÀ HỘI NHẬP

quản lý dữ liệu. Những quan điểm khác nhau này cho thấy, ít nhất là theo quan điểm của tác giả này, nhu cầu cấp thiết phải đào sâu và thúc đẩy cuộc thảo luận về đảm bảo chất lượng dữ liệu trong khu vực công.

5. Ý nghĩa của sự khác biệt giữa dữ liệu và số liệu thống kê đối với cuộc thảo luận về quản lý dữ liệu và các vấn đề khác

Như đã trình bày ở trên, dữ liệu và số liệu thống kê là khác nhau và khung đảm bảo chất lượng cho số liệu thống kê chính thức không thể được áp dụng trực tiếp cho dữ liệu nói chung, ngay cả khi một số tiêu chí chất lượng áp dụng cho cả hai. Hai kết luận này có ý nghĩa mạnh mẽ đối với các cuộc thảo luận hiện nay trong giới thống kê chính thức về quản lý dữ liệu. Các vấn đề họ nêu ra có tầm quan trọng đặc biệt vì các khung đảm bảo chất lượng cho số liệu thống kê chính thức đã gói gọn mục đích và giá trị cốt lõi của chúng. Việc xem xét cẩn thận sự khác biệt giữa dữ liệu và số liệu thống kê cũng có thể mang lại sự rõ ràng hơn cho một số cuộc thảo luận quan trọng khác có ảnh hưởng đến số liệu thống kê chính thức.

5.1. Các cuộc thảo luận hiện tại về quản lý dữ liệu và các chủ đề khác kết hợp dữ liệu và thống kê

Nhiều cuộc thảo luận gần đây xoay quanh vấn đề vai trò quản lý của NSO trong hệ sinh thái dữ liệu mới. Hội nghị các nhà thống kê châu Âu kết luận rằng vai trò của NSO đang thay đổi để đáp ứng nhu cầu và cơ hội mới của hệ sinh thái, được đẩy nhanh bởi cuộc khủng hoảng COVID-19. Nó chỉ ra rằng các NSO có thể định vị mình trong hệ sinh thái dữ liệu mới theo nhiều cách khác nhau, từ thay đổi tối thiểu đến quản lý dữ liệu đầy đủ.

Nhóm làm việc của Liên hợp quốc về quản lý dữ liệu đã đưa ra kết luận tương tự, lưu ý rằng có nhiều cách hiểu khác nhau về quản lý dữ liệu và định nghĩa cũng như ứng dụng của nó phải phù hợp với từng ngữ cảnh cụ thể và sẽ phụ thuộc vào vai trò có thể có của NSO. Nhóm công tác chỉ ra rằng việc quản lý dữ liệu hiệu quả nhằm mục đích tăng cường niềm tin vào dữ liệu cũng như giá trị, cách sử dụng và tác động của nó. Một lưu ý trước đó của Estonia (với sự đóng góp của những người khác) tại Hội nghị các nhà thống kê châu Âu đã gợi ý rằng quản lý dữ liệu, trong số những việc khác, nên bao gồm việc hỗ trợ chất lượng cao và sử dụng dữ liệu được tối ưu hóa.

Các cuộc thảo luận về quản lý dữ liệu tại Ủy ban Thống kê Liên hợp quốc và Hội nghị các nhà thống kê châu Âu nhìn chung không tạo ra sự phân biệt rõ ràng giữa thống kê và dữ liệu. Ghi chú của Estonia về vai trò mới của NSO nói về vai trò của NSO và số liệu thống kê chính thức đồng thời đề cập đến việc cho phép sử dụng dữ liệu. Tương tự, báo cáo của Nhóm công tác Liên Hợp Quốc đề cập đến “dữ liệu và số liệu thống kê” hoặc chỉ “dữ liệu” khi thảo luận về vai trò của các NSO trong quản lý dữ liệu.

Xu hướng nói về dữ liệu và số liệu thống kê một cách dễ dàng mà không phân biệt giữa chúng, không chỉ giới hạn ở việc thảo luận về quản lý dữ liệu. Các điều khoản tham chiếu mới được thông qua của Ủy ban Thống kê Liên Hợp Quốc (Nghị quyết E/RES/2022/3) đề cập đến trách nhiệm của Ủy ban đối với các hệ thống thống kê và liên quan đến dữ liệu. Ngoài ra, việc thảo luận về việc tích hợp thông tin thống kê và không gian địa lý thường xuyên không phân biệt các thuật ngữ thống kê và dữ liệu. Điều tương tự cũng áp

dụng cho các cuộc thảo luận về dữ liệu mở trong cộng đồng thống kê hoặc khi thảo luận về nguồn tài trợ của nhà tài trợ quốc tế cho “dữ liệu và thống kê”. Trong một số trường hợp, sự kết hợp có thể hợp lý, nhưng cần xem xét một số lĩnh vực trong đó NSO có thể quan trọng để quan sát cẩn thận ý nghĩa của sự khác biệt giữa dữ liệu và số liệu thống kê.

5.2. Ý nghĩa của việc phân biệt giữa dữ liệu và thống kê trong cuộc thảo luận về quản lý dữ liệu

Ghi chú của Estonia gửi Hội nghị các nhà thống kê châu Âu vào tháng 6 năm 2020 đã cung cấp danh sách đầu tiên về các hoạt động và kỹ năng cần được đưa vào quản lý dữ liệu trong quản trị dữ liệu công. Chúng bao gồm hỗ trợ chất lượng cao và sử dụng dữ liệu tối ưu, tạo điều kiện truy cập dữ liệu, thúc đẩy chuyên môn, đạo đức, kỹ năng và hiểu biết về dữ liệu, thúc đẩy các tiêu chuẩn, khung và chính sách dữ liệu chung cũng như xây dựng chiến lược dữ liệu. Công việc tiếp theo dưới sự bảo trợ của Hội nghị và công việc đang diễn ra của Ủy ban Thống kê Liên Hợp Quốc đã mở rộng và đào sâu cuộc thảo luận về quản lý dữ liệu, đặc biệt tập trung vào quản trị dữ liệu và các vấn đề về niềm tin, công bằng và hòa nhập.

Tuy nhiên, vẫn chưa có định nghĩa và danh sách thống nhất về chức năng của người quản lý dữ liệu quốc gia, có thể là do cách tiếp cận quản lý dữ liệu của các quốc gia rất đa dạng. Trong trường hợp không có danh sách quy định đã được thống nhất, vì mục đích thảo luận của chúng tôi, chúng tôi có thể sử dụng định nghĩa quy chuẩn về chức năng cơ bản của người quản lý dữ liệu mà không đi sâu vào chi tiết về cách thực hiện điều này trong các tình huống khác nhau của quốc gia. Do đó, dựa trên các mục tiêu đã

nêu của các chiến lược dữ liệu quốc gia khác nhau, chức năng chính của người quản lý dữ liệu sẽ là thúc đẩy việc sử dụng dữ liệu có nguồn gốc từ nguồn chính phủ hoặc tư nhân bởi bất kỳ ai và cho bất kỳ mục đích nào trừ khi nó vi phạm luật và quy định hiện hành.

Dựa trên khái niệm quản lý dữ liệu này, ý thức về sự khác biệt giữa thống kê và dữ liệu, đồng thời thừa nhận rằng các khung đảm bảo chất lượng thể hiện mục đích và giá trị cốt lõi của thống kê chính thức, bài viết này đưa ra các kết luận sơ bộ sau đây.

1. Vai trò của người quản lý dữ liệu đặt ra xung đột cố hữu đối với NSO.

NSO chịu trách nhiệm sản xuất số liệu thống kê chính thức theo FPOS để đáp ứng nhu cầu thông tin cụ thể. Các giá trị và hoạt động cốt lõi của nó như được phản ánh trong các khung đảm bảo chất lượng quốc gia về thống kê chính thức đều hướng tới nhiệm vụ này chứ không phải nhiệm vụ thúc đẩy việc sử dụng dữ liệu của bất kỳ ai cho bất kỳ mục đích nào. Xung đột này được minh họa rõ ràng bởi Nguyên tắc 6 của FPOS. Như đã đề cập, điều này nêu rõ rằng dữ liệu cá nhân do cơ quan thống kê thu thập để tổng hợp thống kê được bảo mật nghiêm ngặt và chỉ được sử dụng cho mục đích thống kê. Hiểu theo nghĩa đen, điều này có nghĩa là tất cả dữ liệu được NSO thu thập thông qua điều tra thống kê, tổng điều tra và đăng ký hành chính cũng như có thể tất cả dữ liệu khác do NSO biên soạn từ các nguồn hành chính và các nguồn khác chỉ có thể được sử dụng cho mục đích thống kê. Trong thực tế, người ta chấp nhận rằng luật thống kê xác định các trường hợp ngoại lệ khỏi nguyên tắc này, chẳng hạn như cho mục đích nghiên cứu hoặc chia sẻ với các nhà sản xuất số liệu thống kê chính thức khác trong những điều kiện hạn chế nhất

➤➤➤ THỐNG KÊ QUỐC TẾ VÀ HỘI NHẬP

định. Tuy nhiên, điều này rõ ràng là không đủ đối với người quản lý dữ liệu được giao nhiệm vụ cung cấp dữ liệu cho tất cả các mục đích sử dụng được chấp nhận. Các NSO có thể đã giảm thiểu tác động của các hạn chế bảo mật bằng cách cung cấp các bộ dữ liệu vi mô và luật pháp để cho phép tái sử dụng dữ liệu cho các mục đích khác. Tuy nhiên, việc vượt quá giới hạn chặt chẽ có thể làm suy yếu lòng tin của các nhà cung cấp dữ liệu, bao gồm cả những người trả lời khảo sát thống kê và/hoặc yêu cầu các biện pháp bảo vệ quyền riêng tư phức tạp. Việc sử dụng dữ liệu sẽ yêu cầu một cách tiếp cận khác để bảo vệ quyền riêng tư so với cách được áp dụng cho số liệu thống kê chính thức.

2. Có nhiều khả năng khác nhau để dung hòa vai trò của người quản lý dữ liệu và nhà sản xuất số liệu thống kê chính thức.

- Đầu tiên, có lựa chọn hiện trạng hoặc chờ xem, trong đó một quốc gia quyết định ngầm hoặc rõ ràng rằng vẫn chưa cần người quản lý dữ liệu.

- Lựa chọn thứ hai là tạo ra chức năng quản lý dữ liệu tại NSO với phạm vi hạn chế mà không ảnh hưởng đến vai trò chính của NSO là cơ quan thống kê.

- Thứ ba, vai trò người quản lý dữ liệu có thể được giao cho một tổ chức không phải NSO, với NSO đóng vai trò hỗ trợ.

- Khả năng thứ tư là mở rộng quyền hạn của NSO để bao gồm các yếu tố chính của quản lý dữ liệu, nhưng tách biệt chức năng và trách nhiệm này với vai trò của NSO trong việc sản xuất số liệu thống kê chính thức.

- Khả năng thứ năm là hợp nhất trách nhiệm sản xuất số liệu thống kê chính thức với trách nhiệm quản lý dữ liệu. Điều này về

cơ bản sẽ làm thay đổi khái niệm cơ bản của NSO với tư cách là một cơ quan thống kê và có thể yêu cầu sửa đổi luật thống kê quốc gia để dung hòa các yêu cầu và nguyên tắc của thống kê chính thức với yêu cầu về vai trò của người quản lý dữ liệu. Ở cấp độ toàn cầu, điều này có thể hàm ý việc sửa đổi hoặc thay thế FPOS nếu nhiều NSO đi theo con đường này.

Hoàn cảnh quốc gia khác nhau có thể quyết định lựa chọn nào trong số này là khả thi. Tuy nhiên, theo nguyên tắc chung, có vẻ như các NSO không nên cố gắng đơn giản đảm nhận vai trò quản lý dữ liệu theo các nhiệm vụ hiện có của họ. Như bài viết này đã chỉ ra, có những căng thẳng giữa các mục tiêu cơ bản của chính sách thống kê chính thức và chính sách dữ liệu/Dữ liệu lớn, do đó, việc hợp nhất trách nhiệm đối với chúng trong một cơ quan duy nhất có nguy cơ làm suy yếu số liệu thống kê chính thức và không đáp ứng nhu cầu phát triển dữ liệu như một tài sản một cách có trách nhiệm.

3. Quản lý dữ liệu là "không thể bỏ lỡ" – cơ hội để có số liệu thống kê chính thức.

Truy cập dữ liệu và nguồn lực đầy đủ là những yêu cầu then chốt để tạo ra số liệu thống kê chính thức như được phản ánh trong các khung chất lượng hiện có cho số liệu thống kê chính thức. Sự tồn tại của người quản lý dữ liệu hứa hẹn sẽ cải thiện đáng kể quyền truy cập vào dữ liệu hành chính và dữ liệu riêng tư cho mọi người và thậm chí có thể còn hơn thế nữa đối với số liệu thống kê chính thức vì cơ quan này đã có các biện pháp bảo vệ nội bộ mạnh mẽ để bảo vệ quyền riêng tư của dữ liệu. Điều này rất quan trọng vì các nguồn mới sẽ thay thế những nguồn cũ. Hơn nữa, có thể kỳ vọng rằng mọi nguồn lực công và tư bổ sung cho dữ liệu và

thống kê sẽ chủ yếu tập trung vào việc phát triển và sử dụng tài sản dữ liệu quốc gia hơn là phát triển số liệu thống kê chính thức bổ sung. Các nguồn lực hiện có dành cho thống kê chính thức thậm chí có thể được lập trình lại theo dữ liệu, điều dường như đã xảy ra trong trường hợp hỗ trợ phát triển thống kê⁶. Ngoài ra, có lẽ sẽ hiệu quả hơn nếu thiết lập các trung tâm dữ liệu cho phép toàn bộ chính phủ sử dụng AI hơn là thực hiện các thỏa thuận riêng biệt về số liệu thống kê chính thức. Do đó, điều quan trọng đối với các NSO là hỗ trợ thiết lập chức năng quản lý dữ liệu quốc gia vì nó sẽ cho phép truy cập dữ liệu tốt hơn và giúp khắc phục những hạn chế về tài nguyên như những hạn chế liên quan đến khảo sát tốn kém hoặc sử dụng các công nghệ mới như AI. Ngoài ra, các NSO có thể sử dụng việc thành lập cơ quan quản lý dữ liệu quốc gia như một cơ hội để xác định rõ hơn vai trò của chính họ trong hệ sinh thái dữ liệu mới, bất kể vai trò này có thể là gì.

5.3. Ý nghĩa của việc phân biệt giữa dữ liệu và thống kê trong việc thảo luận các vấn đề quan trọng khác trong thống kê chính thức

Việc phân biệt giữa dữ liệu và số liệu thống kê cũng mang lại sự rõ ràng hơn về các vấn đề khác:

1. Đã có nỗ lực tích cực về dữ liệu mở trong cộng đồng thống kê. Tầm quan trọng của các nguyên tắc dữ liệu mở như được nêu rõ trong Hiến chương Dữ liệu Mở Quốc tế đối

⁶ Hỗ trợ phát triển chính thức (ODA) cho các nước đang phát triển cho các dự án tập trung chủ yếu vào dữ liệu và thống kê đã bị đình trệ trong 10 năm qua mặc dù yêu cầu giám sát SDG ngày càng tăng và bất chấp mức tăng ODA tổng thể. Đồng thời, số lượng và kinh phí của các dự án có thành phần dữ liệu và thống kê đã tăng lên đáng kể.

với số liệu thống kê chính thức không thể bị phóng đại. Tuy nhiên, có nhiều điểm tương đồng giữa các nguyên tắc của Hiến chương và các nguyên tắc chất lượng thống kê của UN-NQAF, và hầu hết các số liệu thống kê chính thức đều đã được công khai. Do đó, việc thảo luận về dữ liệu mở đối với dữ liệu quan trọng hơn nhiều so với số liệu thống kê chính thức và nên được theo đuổi chủ yếu trong các diễn đàn về dữ liệu.

2. Có nhiều cuộc thảo luận và hướng dẫn về việc tích hợp thông tin không gian địa lý và thống kê. Một số phân biệt rõ ràng số liệu thống kê và dữ liệu, số khác thì không. Điều quan trọng cần lưu ý là kiểu tích hợp tự nhiên của cả hai là ở cấp độ dữ liệu, tức là tích hợp dữ liệu thống kê và không gian địa lý ở cấp độ đầu vào hoặc sản xuất của số liệu thống kê chính thức. Đầu ra thống kê là những tổng hợp có khía cạnh không gian địa lý, nhưng giá trị tích hợp ở cấp độ đó thường bị hạn chế hơn.

3. Cần phân biệt rõ kinh phí cho dữ liệu và kinh phí cho thống kê chính thức. Ví dụ, trong nhiều năm qua đã có nhiều lời kêu gọi tăng cường hỗ trợ của các nhà tài trợ quốc tế cho “dữ liệu và thống kê” và thông tin thường được trình bày về mức tài trợ không phân biệt giữa hỗ trợ cho thống kê chính thức và hệ thống dữ liệu. Điều này ở một số khía cạnh có thể hiểu được, tạo ra mối liên kết chặt chẽ giữa các hệ thống hành chính nhất định như đăng ký dân sự và thống kê quan trọng với việc sản xuất số liệu thống kê chính thức. Tuy nhiên, nếu hệ thống thống kê quốc gia của một quốc gia đang phát triển phải chịu trách nhiệm về khả năng cung cấp số liệu thống kê chính thức như các chỉ số SDG, thì cần có thông tin về sự hỗ trợ mà hệ thống

➤ ➤ ➤ THỐNG KÊ QUỐC TẾ VÀ HỘI NHẬP

đó nhận được. Ngoài ra, khi tầm quan trọng của dữ liệu tăng lên, việc chuyển nguồn tài trợ từ số liệu thống kê chính thức sang dữ liệu có thể không được chú ý trừ khi tài trợ cho dữ liệu được phân biệt rõ ràng với tài trợ cho số liệu thống kê chính thức. Duy trì sự khác biệt này cũng có thể mang lại những hiểu biết sâu sắc hơn về hiệu quả hỗ trợ. Ngoài ra, vì liên quan đến nguồn lực trong nước, việc tăng tổng kinh phí cho dữ liệu và thống kê ban đầu có thể mang lại lợi ích cho số liệu thống kê chính thức. Tuy nhiên, do tính hấp dẫn và sự phát triển năng động của việc sử dụng dữ liệu, số liệu thống kê chính thức theo thời gian có thể chỉ trở thành một trong nhiều chức năng của dữ liệu. Cả hai đều khác nhau và quan trọng theo cách riêng của chúng và do đó cần có nguồn tài trợ riêng.

6. Tóm tắt các phát hiện

Sở tay Khung Đảm bảo Chất lượng Quốc gia của Liên hợp quốc về Thống kê Chính thức phân biệt rõ ràng giữa dữ liệu là đầu vào và số liệu thống kê là đầu ra của quá trình sản xuất số liệu thống kê chính thức. Việc quan sát cẩn thận sự khác biệt này dẫn đến kết luận rằng các khung đảm bảo chất lượng cho số liệu thống kê chính thức không thể được áp dụng trực tiếp cho dữ liệu nói chung hoặc chỉ ở một mức độ hạn chế. Đã có những nỗ lực quan trọng nhằm hệ thống hóa việc đánh giá chất lượng hành chính và các nguồn khác để đưa ra số liệu thống kê chính thức, nhưng những nỗ lực này có sự khác biệt đáng kể trong cách tiếp cận. Có thể có lợi khi có một danh sách chung các tiêu chí hoặc chỉ số và hướng dẫn cơ bản để cung cấp ít nhất một điểm khởi đầu hoặc tài liệu tham khảo thống nhất cho những

người quan tâm đến chất lượng của nguồn dữ liệu để đưa ra số liệu thống kê chính thức. Điều này cũng có thể hỗ trợ việc xác định các tiêu chí chất lượng cho dữ liệu nói chung, những tiêu chí này cần được phát triển với sự hợp tác của các nhà sản xuất và người sử dụng dữ liệu thuộc chính phủ và tư nhân.

Việc phân biệt dữ liệu với số liệu thống kê có thể giúp mang lại sự rõ ràng cho cuộc thảo luận về vai trò của số liệu thống kê chính thức trong hệ sinh thái dữ liệu mới. Đặc biệt, sự khác biệt rất quan trọng đối với cuộc thảo luận về quản lý dữ liệu và liệu NSO có nên đảm nhận vai trò đó hay không. Bài viết này lập luận rằng vai trò kép của người giám sát số liệu thống kê chính thức và vai trò của người quản lý dữ liệu quốc gia không thể dễ dàng dung hòa được. Đồng thời, các NSO nên hỗ trợ việc thành lập cơ quan quản lý dữ liệu quốc gia vì điều này có thể tạo điều kiện thuận lợi cho việc tạo ra số liệu thống kê chính thức⁷. Việc chú ý hơn đến sự khác biệt giữa dữ liệu và thống kê chính thức cũng sẽ giúp mở rộng các khía cạnh khác về tương lai của thống kê chính thức trong hệ sinh thái dữ liệu mới, bao gồm các câu hỏi liên quan đến dữ liệu mở, tích hợp thông tin không gian địa lý và hỗ trợ phát triển thống kê.

Đỗ Ngát (dịch)

Nguồn: <https://content.iospress.com/download/statistical-journal-of-the-iaos/sji230008?id=statistical-journal-of-the-iaos%2Fsjj230008>

⁷ Mặc dù đây không phải là chủ đề của bài viết này nhưng tác giả có quan điểm rằng cần có một tổ chức hoặc chức năng quốc gia trong chính phủ thúc đẩy và giải quyết việc sử dụng dữ liệu được cung cấp nguồn lực hợp lý cho nhiệm vụ này.