

XÂY DỰNG CƠ SỞ DỮ LIỆU KHOA HỌC XÃ HỘI: HỆ SINH THÁI SỐ CHO NGHIÊN CỨU VÀ HOẠCH ĐỊNH CHÍNH SÁCH TRONG KỶ NGUYÊN MỚI

ĐẶNG NGUYỄN ANH*

Tóm tắt: Bài viết phân tích vai trò của cơ sở dữ liệu khoa học xã hội trong đổi mới hoạt động nghiên cứu và nâng cao chất lượng tư vấn, hoạch định chính sách trong bối cảnh chuyển đổi số và phát triển kinh tế dữ liệu. Trên cơ sở phân tích tài liệu và tổng quan nghiên cứu, bài viết làm rõ vai trò của dữ liệu như một nền tảng của quản trị quốc gia hiện đại. Kết quả nghiên cứu cho thấy chuyển đổi số trong khoa học xã hội cần hướng đến xây dựng hạ tầng dữ liệu xã hội có khả năng tích lũy, liên thông và tái sử dụng để phục vụ nghiên cứu, dự báo và tư vấn chính sách, thay vì chỉ tập trung vào số hóa tài liệu và ứng dụng công nghệ số. Trong đó, các cuộc điều tra cơ bản giữ vai trò nền tảng trong việc hình thành hệ sinh thái dữ liệu có tính kế thừa và khả năng so sánh theo thời gian. Bài viết cũng chỉ ra những thách thức nổi bật hiện nay, bao gồm tình trạng dữ liệu phân tán, thiếu liên thông, hạn chế về nguồn nhân lực và năng lực quản trị dữ liệu. Trên cơ sở đó, bài viết đề xuất định hướng xây dựng cơ sở dữ liệu khoa học xã hội như một hạ tầng chiến lược, góp phần nâng cao chất lượng quản trị quốc gia và thúc đẩy hoạch định chính sách dựa trên bằng chứng trong kỷ nguyên phát triển mới.

Từ khóa: khoa học xã hội, cơ sở dữ liệu, hệ sinh thái dữ liệu, quản trị dữ liệu, chuyển đổi số.

Nhận bài: 20/4/2026

Gửi phản biện: 23/4/2026

Duyệt đăng: 25/6/2026

1. Đặt vấn đề

Nghị quyết số 57-NQ/TW của Bộ Chính trị về đột phá phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia đã xác lập định hướng chiến lược và tạo động lực cho phát triển đất nước trong giai đoạn mới. Nghị quyết nhấn mạnh yêu cầu “khai thác tối đa tiềm năng của dữ liệu, đưa dữ liệu thành tư liệu sản xuất chính, thúc đẩy phát triển nhanh cơ sở dữ liệu lớn, công nghiệp dữ liệu, kinh tế dữ liệu” (Bộ Chính trị, 2024). Định hướng này đặt ra những yêu cầu mới đối với các lĩnh vực kinh tế, quản trị quốc gia và khoa học xã hội. Với chức năng nghiên cứu các quá trình xã hội, con người và thể chế, khoa học

* Viện Xã hội học và Tâm lý học, Viện Hàn lâm Khoa học xã hội Việt Nam.

xã hội có vai trò cung cấp luận cứ khoa học phục vụ tư vấn và điều hành chính sách phát triển.

Trong bối cảnh chuyển đổi số, khối lượng dữ liệu xã hội được tạo ra ngày càng lớn và đa dạng, bao gồm dữ liệu hành chính, dữ liệu điều tra khảo sát, dữ liệu từ các nền tảng dịch vụ công, cũng như dữ liệu lớn phát sinh từ các hoạt động trên không gian số. Nếu được tổ chức, chuẩn hóa, liên thông và khai thác hiệu quả, các nguồn dữ liệu này sẽ trở thành hạ tầng tri thức quan trọng của quốc gia, góp phần đổi mới phương pháp nghiên cứu khoa học xã hội, nâng cao chất lượng phân tích, dự báo và đánh giá chính sách, đồng thời tăng cường vai trò tư vấn của khoa học xã hội trong phục vụ phát triển kinh tế-xã hội.

Tuy nhiên, khoa học xã hội Việt Nam hiện vẫn chưa có một hạ tầng dữ liệu đủ mạnh, đồng bộ và có khả năng liên thông để đáp ứng yêu cầu của chuyển đổi số và quản trị dựa trên dữ liệu. Dữ liệu hiện phân tán theo từng đề tài, dự án và chương trình nghiên cứu, thiếu các tiêu chuẩn thống nhất về thu thập, lưu trữ và chia sẻ, đồng thời chưa hình thành cơ chế khai thác và tái sử dụng hiệu quả. Đặc biệt, các cuộc điều tra cơ bản quy mô lớn, được thực hiện định kỳ nhằm xây dựng chuỗi dữ liệu dài hạn phục vụ phân tích xu hướng và dự báo, còn hết sức hạn chế. Hơn nữa, năng lực khoa học dữ liệu cũng chưa đáp ứng yêu cầu mới, từ việc xác định các lĩnh vực ưu tiên, thiết kế điều tra quy mô quốc gia đến quản trị, phân tích và khai thác dữ liệu phục vụ nghiên cứu và hoạch định chính sách. Trong khi đó, các vấn đề xã hội ngày càng phức tạp, biến động nhanh và chịu tác động đồng thời của các yếu tố kinh tế, công nghệ, môi trường, nhân khẩu học và hội nhập quốc tế. Điều này đặt ra yêu cầu chuyển đổi mạnh mẽ đối với khoa học xã hội trên các phương diện: phương pháp nghiên cứu, hạ tầng dữ liệu và năng lực phân tích. Phương thức nghiên cứu truyền thống, chủ yếu dựa trên các khảo sát đơn lẻ, phân tán và quy mô nhỏ, ngày càng bộc lộ những hạn chế trong việc cung cấp bằng chứng khoa học kịp thời, có hệ thống và đủ độ tin cậy để phục vụ điều hành, đánh giá và điều chỉnh chính sách.

Trong bối cảnh Nghị quyết số 57-NQ/TW xác định dữ liệu là tư liệu sản xuất chủ yếu của nền kinh tế số và quản trị hiện đại, việc xây dựng cơ sở dữ liệu khoa học xã hội không còn là một lựa chọn mà đã trở thành nhiệm vụ chiến lược. Một hạ tầng dữ liệu xã hội được xây dựng bài bản, có khả năng tích lũy, liên thông và tái sử dụng sẽ tạo nền tảng cho nghiên cứu dựa trên bằng chứng, nâng cao chất lượng tư vấn chính sách và góp phần tăng cường hiệu quả quản trị quốc gia trong kỷ nguyên phát triển mới.

Bài viết phân tích vai trò của dữ liệu khoa học xã hội trong đổi mới hoạt động nghiên cứu và nâng cao chất lượng tư vấn, hoạch định chính sách trong bối cảnh chuyển đổi số quốc gia. Trên cơ sở đó, nghiên cứu làm rõ yêu cầu xây dựng cơ sở dữ liệu khoa học xã hội theo hướng chuẩn hóa, liên thông và chia sẻ, đồng thời đề xuất định hướng phát triển hệ sinh thái dữ liệu phục vụ nghiên cứu và hoạch định chính sách dựa trên bằng chứng ở Việt Nam. Bài viết khẳng định dữ liệu khoa học xã hội là hạ tầng chiến lược của hoạt động nghiên cứu và hoạch định chính sách dựa trên bằng chứng, đồng thời làm rõ sự khác biệt

giữa dữ liệu khoa học xã hội với dữ liệu hành chính và dữ liệu thống kê quốc gia, qua đó nhấn mạnh vai trò của các cuộc điều tra cơ bản dài hạn trong việc hình thành các bộ dữ liệu phục vụ nghiên cứu và dự báo. Trên nền tảng đó, bài viết đề xuất mô hình liên thông “Nghiên cứu - Dữ liệu - Chính sách”, trong đó dữ liệu giữ vai trò hạ tầng kết nối hoạt động nghiên cứu với quá trình hoạch định và đánh giá chính sách.

2. Tổng quan nghiên cứu

Trên thế giới, hoạch định chính sách dựa trên bằng chứng (Evidence-based Policymaking) ngày càng được thừa nhận là một cách tiếp cận quan trọng nhằm nâng cao chất lượng và hiệu quả quản trị công trong bối cảnh các vấn đề phát triển ngày càng phức tạp và biến động nhanh (Head, 2016). Hoạch định chính sách dựa trên bằng chứng không đơn thuần là sử dụng số liệu trong quản lý nhà nước, mà là quá trình tích hợp bằng chứng khoa học vào toàn bộ chu trình chính sách nhằm giảm tính chủ quan trong ra quyết định. Tuy nhiên, đối với các quốc gia đang phát triển, thiếu dữ liệu thực chứng và năng lực phân tích đáng tin cậy là một trong những nguyên nhân làm gia tăng rủi ro chính sách và hạn chế hiệu quả quản trị (Sutcliffe và Court, 2005).

Trong khoa học xã hội, nơi đối tượng nghiên cứu là con người, các thiết chế và quan hệ xã hội luôn vận động và biến đổi, hạ tầng dữ liệu ngày càng trở thành yếu tố quan trọng quyết định năng lực nghiên cứu và khả năng cung cấp bằng chứng khoa học phục vụ tư vấn, hoạch định và đánh giá chính sách. Cuộc cách mạng số đang thúc đẩy sự chuyển đổi phương pháp luận từ các nghiên cứu dựa trên những bộ dữ liệu đơn lẻ sang mô hình nghiên cứu tích lũy dựa trên các hệ thống dữ liệu có quy mô lớn, được liên thông và cập nhật thường xuyên. Sự chuyển đổi này mở ra điều kiện để ứng dụng các phương pháp phân tích hiện đại, qua đó nâng cao chất lượng nghiên cứu, năng lực dự báo và hiệu quả tư vấn chính sách. Việc xây dựng cơ sở dữ liệu dùng chung góp phần tối ưu hóa nguồn lực nghiên cứu thông qua cơ chế tái sử dụng dữ liệu, đồng thời tạo nền tảng cho việc triển khai các phương pháp mô hình hóa, phân tích dữ liệu lớn và trí tuệ nhân tạo. Trong nền kinh tế số, dữ liệu ngày càng được xem là một tư liệu sản xuất mới và là nguồn lực chiến lược, tạo ra giá trị thông qua quá trình kết nối, phân tích và sử dụng, từ đó thúc đẩy đổi mới sáng tạo và phát triển kinh tế-xã hội (OECD, 2019; World Bank, 2021).

Ở Việt Nam, chuyển đổi số và xây dựng cơ sở dữ liệu đã được xác định là định hướng chiến lược trong nhiều chủ trương, chính sách của Đảng và Nhà nước. Tuy nhiên, do đây là lĩnh vực còn tương đối mới, các nghiên cứu trong nước thời gian qua chủ yếu tập trung vào chuyển đổi số trong quản trị công, phát triển chính phủ số, xã hội số và quản trị dữ liệu (Bộ Thông tin và Truyền thông, 2024; Lê Thị Hoàng Minh và Vũ Công Giao, 2025, 2026). Trong khi đó, các nghiên cứu về cơ sở dữ liệu khoa học xã hội vẫn còn hạn chế cả về số lượng và chiều sâu. Đến nay, chưa có nhiều công trình tiếp cận cơ sở dữ liệu khoa học xã hội như một hạ tầng nghiên cứu quốc gia, tích hợp các cuộc điều tra cơ bản, dữ liệu hành chính và các nguồn dữ liệu số nhằm phục vụ nghiên cứu, dự báo và hoạch định chính sách

dựa trên bằng chứng. Phần lớn các nghiên cứu mới dừng ở số hóa dữ liệu, ứng dụng công nghệ thông tin hoặc chuyển đổi số trong quản lý, trong khi các vấn đề về thiết kế, tổ chức, quản trị, chia sẻ và khai thác cơ sở dữ liệu khoa học xã hội chưa được phân tích một cách hệ thống.

Khoảng trống nghiên cứu này cho thấy sự cần thiết phải tiếp cận cơ sở dữ liệu khoa học xã hội như một hạ tầng tri thức quốc gia, vượt ra ngoài chức năng lưu trữ thông tin để trở thành nền tảng cho nghiên cứu, dự báo và hoạch định chính sách trong bối cảnh chuyển đổi số. Để làm rõ vấn đề này, bài viết sử dụng phương pháp phân tích tài liệu, tổng quan nghiên cứu và phân tích chính sách nhằm hệ thống hóa cơ sở lý luận và thực tiễn về dữ liệu trong khoa học xã hội; đồng thời phân tích vai trò của cơ sở dữ liệu khoa học xã hội đối với đổi mới hoạt động nghiên cứu và nâng cao chất lượng hoạch định chính sách. Trên cơ sở đó, bài viết đề xuất định hướng phát triển hệ sinh thái dữ liệu khoa học xã hội phục vụ quản trị quốc gia dựa trên bằng chứng.

3. Vai trò của dữ liệu khoa học xã hội trong tư vấn và xây dựng chính sách

Cơ sở dữ liệu khoa học xã hội là hệ thống dữ liệu được thu thập, chuẩn hóa, lưu trữ và quản lý một cách có hệ thống nhằm phản ánh cấu trúc, động thái và các quá trình xã hội, phục vụ nghiên cứu khoa học, phân tích, dự báo và hoạch định chính sách. Cơ sở dữ liệu khoa học xã hội vừa là nguồn dữ liệu phục vụ nghiên cứu, vừa là nền tảng tri thức quan trọng cho phát triển đất nước. Khác với các loại tài nguyên vật chất truyền thống, dữ liệu có khả năng được tái sử dụng nhiều lần và liên tục tạo ra giá trị mới thông qua quá trình tích hợp, kết nối và phân tích bằng các công nghệ hiện đại. Do đó, xây dựng cơ sở dữ liệu khoa học xã hội là một nhiệm vụ chiến lược nhằm phát triển hạ tầng tri thức quốc gia, đáp ứng yêu cầu quản trị và phát triển dựa trên bằng chứng.

Khác với dữ liệu thống kê quốc gia, dữ liệu khoa học xã hội không được xây dựng để công bố các chỉ tiêu thống kê chính thức mà hướng tới phản ánh toàn diện cấu trúc, động thái và các quá trình xã hội, cũng như nhận thức, giá trị, niềm tin, hành vi và lối sống của con người. Nguồn dữ liệu này được hình thành từ sự tích hợp giữa các cuộc điều tra cơ bản định kỳ với dữ liệu hành chính và các nguồn dữ liệu số, qua đó tạo lập nền tảng thông tin phục vụ nghiên cứu, dự báo và đánh giá chính sách. Đây cũng là cơ sở thực chứng quan trọng để thiết kế, điều chỉnh và hoàn thiện chính sách theo hướng lấy con người làm trung tâm. Theo Sutcliffe và Court (2005), hoạch định chính sách dựa trên bằng chứng là quá trình xây dựng và điều chỉnh chính sách trên cơ sở dữ liệu thực chứng và các kết quả phân tích đáng tin cậy nhằm nâng cao hiệu quả và tính khả thi của chính sách công.

Một hệ thống dữ liệu khoa học được thiết kế bài bản cho phép theo dõi liên tục các xu hướng kinh tế-xã hội, nhận diện sớm những vấn đề mới phát sinh và đánh giá tác động của chính sách một cách khách quan. Thông qua việc tích hợp dữ liệu điều tra cơ bản với dữ liệu hành chính, các cơ quan quản lý có điều kiện phân tích đối tượng thụ hưởng, mức

độ và xu hướng tác động của chính sách; so sánh sự khác biệt giữa các vùng, nhóm dân cư và giai đoạn khác nhau; đồng thời đánh giá tính bền vững của các kết quả đạt được để kịp thời điều chỉnh chính sách phù hợp với mục tiêu phát triển nhanh và bền vững.

Trong khoa học xã hội, cơ sở dữ liệu là nền tảng cốt lõi nhằm tạo lập nguồn dữ liệu dùng chung cho nhiều đề tài, chương trình và lĩnh vực nghiên cứu. Khả năng so sánh theo thời gian, vùng lãnh thổ và nhóm xã hội giúp tích lũy, kiểm chứng và tái sử dụng kết quả nghiên cứu, qua đó nâng cao giá trị khoa học và khả năng ứng dụng vào thực tiễn. Trong bối cảnh nguồn lực nghiên cứu còn hạn chế, đầu tư xây dựng cơ sở dữ liệu khoa học xã hội góp phần tối ưu hóa chi tiêu công, giảm trùng lặp trong các hoạt động điều tra, khảo sát, đồng thời nâng cao chất lượng phân tích, dự báo và tư vấn chính sách dựa trên bằng chứng.

Bên cạnh giá trị đối với nghiên cứu, cơ sở dữ liệu khoa học xã hội còn là cầu nối giữa hoạt động nghiên cứu và quá trình hoạch định chính sách. Khi dữ liệu được chuẩn hóa, liên thông và bảo đảm khả năng tiếp cận, các kết quả nghiên cứu có thể được kiểm chứng, chia sẻ và tích hợp vào quá trình xây dựng, thực thi và đánh giá chính sách. Trong mô hình quản trị hiện đại, chính sách công cần được xây dựng trên nền tảng dữ liệu và bằng chứng khoa học nhằm nâng cao tính khách quan, hiệu quả và khả năng giải trình. Điều này đặc biệt có ý nghĩa đối với các lĩnh vực có tính biến động cao như lao động, di cư, an sinh xã hội và bất bình đẳng, nơi các quyết định chính sách đòi hỏi thông tin kịp thời, chính xác và có khả năng dự báo.

4. Xây dựng cơ sở dữ liệu khoa học xã hội: Nền tảng của hệ sinh thái số

Hệ sinh thái số trong khoa học xã hội là không gian kết nối giữa nghiên cứu, dữ liệu và chính sách trên nền tảng công nghệ số. Trong mô hình này, hạ tầng dữ liệu dùng chung giữ vai trò trung tâm, với dữ liệu được quản trị như một tài sản công, tạo điều kiện để các nhà khoa học, cơ quan nghiên cứu và cơ quan quản lý cùng khai thác, chia sẻ và bổ sung tri thức. Tính mở, khả năng liên thông và chia sẻ dữ liệu giúp khắc phục tình trạng phân tán, cục bộ, đồng thời tạo điều kiện ứng dụng dữ liệu lớn, trí tuệ nhân tạo và các phương pháp phân tích hiện đại, qua đó rút ngắn khoảng cách giữa nghiên cứu và hoạch định chính sách. Để hiện thực hóa mô hình này, việc xây dựng cơ sở dữ liệu khoa học xã hội cần được xác định là một nhiệm vụ chiến lược, có tầm nhìn dài hạn, thay vì triển khai theo từng đề tài hoặc nhiệm vụ nghiên cứu riêng lẻ. Trọng tâm là phát triển các bộ dữ liệu cốt lõi phản ánh toàn diện cấu trúc và động thái của xã hội, đồng thời bảo đảm tính chuẩn hóa, liên thông, khả năng tích hợp và mở rộng. Khi được đầu tư đồng bộ, hạ tầng dữ liệu sẽ trở thành trụ cột của hệ sinh thái số, tạo nền tảng cho nghiên cứu tích lũy, nâng cao chất lượng dự báo và hỗ trợ kịp thời quá trình xây dựng, thực thi và đánh giá chính sách.

Trong chiến lược đó, các cuộc điều tra cơ bản cần được đặt ở vị trí trung tâm. Đây là nguồn dữ liệu nền tảng phục vụ nghiên cứu, dự báo và hoạch định chính sách trên nhiều lĩnh vực. Tuy nhiên, khoa học xã hội ở Việt Nam hiện vẫn thiếu các chương trình khảo sát

được thực hiện định kỳ với bộ công cụ chuẩn hóa và khả năng so sánh theo thời gian. Trong khi đó, các khảo sát dọc và điều tra xã hội lặp lại là cơ sở quan trọng để theo dõi biến đổi xã hội, đánh giá tác động của chính sách và nâng cao năng lực dự báo (OECD, 2019). Thiếu các chuỗi dữ liệu dài hạn sẽ làm hạn chế khả năng nhận diện xu hướng phát triển, khiến nhiều quyết sách mang tính ngắn hạn và thiên về xử lý tình huống. Các cuộc điều tra cơ bản cần bao quát những lĩnh vực trọng yếu của đời sống kinh tế-xã hội. Dữ liệu về kinh tế và lao động giúp theo dõi việc làm, thu nhập, phân bổ nguồn lực và năng suất lao động; dữ liệu về văn hóa, dân tộc và tôn giáo phản ánh hệ giá trị, niềm tin, hành vi và mức độ gắn kết xã hội, tạo cơ sở cho các chính sách phát triển bao trùm và bảo tồn bản sắc văn hóa; trong khi dữ liệu về môi trường và phát triển bền vững hỗ trợ theo dõi chất lượng môi trường, sử dụng tài nguyên và tác động của biến đổi khí hậu, phục vụ xây dựng các chính sách quản lý và thích ứng hiệu quả.

Tuy nhiên, cần xác định rõ ranh giới và vai trò của cơ sở dữ liệu khoa học xã hội. Dữ liệu này không nhằm công bố các chỉ tiêu chính thức từ các cuộc điều tra thống kê quốc gia, mà nhằm cung cấp bằng chứng khoa học cho việc thiết kế, điều chỉnh và đánh giá chính sách theo hướng lấy con người làm trung tâm. Ví dụ, một cơ sở dữ liệu về giá trị, sức khỏe và lối sống của thanh niên không nhằm thay thế các chỉ tiêu thống kê về dân số trong tuổi thanh niên, mà nhằm giúp các nhà hoạch định chính sách hiểu rõ hơn về kỳ vọng, động lực và thách thức của thế hệ trẻ trong bối cảnh chuyển đổi kinh tế-xã hội. Tương tự, dữ liệu về thực hành văn hóa cộng đồng không nhằm xếp hạng hay xây dựng chuẩn mực văn hóa, mà nhằm hỗ trợ các chính sách bảo tồn, phát huy giá trị văn hóa theo hướng tôn trọng đa dạng và chủ thể văn hóa. Đây là những nội dung không có trong dữ liệu thống kê, nhưng lại là trọng tâm của khoa học xã hội.

Bên cạnh các cuộc điều tra cơ bản, việc xây dựng cơ sở dữ liệu khoa học xã hội cần chú trọng tích hợp nhiều nguồn dữ liệu, bao gồm dữ liệu hành chính, dữ liệu thống kê chính thức và dữ liệu số phát sinh từ các hệ thống thông tin. Sự kết hợp giữa dữ liệu điều tra cơ bản và dữ liệu hành chính tạo điều kiện để phản ánh toàn diện thực trạng xã hội, đồng thời theo dõi sự thay đổi của các nhóm dân cư, vùng lãnh thổ và lĩnh vực chính sách theo thời gian. Việc tích hợp các nguồn dữ liệu giúp nhận diện chính xác đối tượng chịu tác động, mức độ ảnh hưởng và xu hướng biến động của các vấn đề kinh tế-xã hội, qua đó cung cấp bằng chứng tin cậy và kịp thời cho quá trình xây dựng, điều chỉnh và đánh giá chính sách. Đây là nền tảng quan trọng để nâng cao chất lượng quản trị quốc gia, tăng cường năng lực dự báo và khả năng thích ứng của chính sách trước những biến động nhanh và phức tạp của bối cảnh phát triển hiện nay.

Một nguyên tắc quan trọng trong xây dựng cơ sở dữ liệu khoa học xã hội là thiết kế theo hướng mở, liên thông và dùng chung trong khuôn khổ pháp luật. Dữ liệu được hình thành từ ngân sách nhà nước cần được xem là tài sản công, được quản lý tập trung, bảo đảm an toàn và bảo mật, đồng thời tạo điều kiện cho các nhà khoa học, cơ quan nghiên cứu và

đơn vị tư vấn chính sách tiếp cận, khai thác phục vụ lợi ích công. Cách tiếp cận này phù hợp với tinh thần Nghị quyết số 57-NQ/TW, trong đó dữ liệu được xác định là nguồn lực chiến lược và tư liệu sản xuất quan trọng, cần được tổ chức, quản trị và khai thác hiệu quả nhằm phục vụ phát triển kinh tế-xã hội và nâng cao chất lượng hoạch định chính sách.

Một hạ tầng dữ liệu được xây dựng bài bản, bảo đảm khả năng chia sẻ và liên thông sẽ tạo điều kiện áp dụng các phương pháp phân tích hiện đại như phân tích định lượng, phân tích dữ liệu lớn, mô phỏng xã hội và trí tuệ nhân tạo, qua đó nâng cao năng lực nghiên cứu, dự báo và tư vấn chính sách. Ngược lại, khi dữ liệu còn phân tán, thiếu chuẩn hóa và khó tiếp cận, hiệu quả khai thác của các phương pháp này sẽ bị hạn chế. Vì vậy, hạ tầng dữ liệu xã hội giữ vai trò nền tảng trong quá trình chuyển đổi từ nghiên cứu dựa trên các bộ dữ liệu rời rạc của từng đề tài sang mô hình nghiên cứu tích lũy và chia sẻ dữ liệu. Trong mô hình này, mỗi nghiên cứu vừa bổ sung dữ liệu cho hệ thống dùng chung, vừa khai thác và gia tăng giá trị của nguồn dữ liệu đã được tích lũy.

5. Thuận lợi, thách thức và điều kiện hình thành cơ sở dữ liệu khoa học xã hội trong giai đoạn phát triển mới

Việc xây dựng cơ sở dữ liệu khoa học xã hội hiện nay đang hội tụ nhiều điều kiện thuận lợi. Trước hết, chủ trương phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số đã được xác lập rõ trong các văn kiện của Đảng và Nhà nước, tạo nền tảng chính trị và pháp lý để nhìn nhận dữ liệu như một nguồn lực chiến lược của phát triển. Trong bối cảnh đó, dữ liệu khoa học xã hội ngày càng được coi là hạ tầng tri thức phục vụ nghiên cứu, dự báo và hoạch định chính sách. Cùng với sự phát triển của công nghệ số, điện toán đám mây, dữ liệu lớn và trí tuệ nhân tạo, khả năng thu thập, lưu trữ, xử lý và chia sẻ dữ liệu ngày càng được nâng cao, mở ra cơ hội hình thành các hệ thống dữ liệu liên thông và dùng chung. Bên cạnh đó, đội ngũ nghiên cứu trẻ được đào tạo trong môi trường số, có khả năng tiếp cận các phương pháp phân tích dữ liệu hiện đại, đang trở thành nguồn nhân lực quan trọng cho quá trình phát triển cơ sở dữ liệu khoa học xã hội.

Tuy nhiên, quá trình này vẫn đang đối mặt với nhiều thách thức. Hoạt động nghiên cứu hiện nay chủ yếu vận hành theo từng đề tài hoặc nhiệm vụ riêng lẻ, chưa hình thành cơ chế tích lũy, chia sẻ và tái sử dụng dữ liệu một cách hệ thống. Dữ liệu vẫn chủ yếu được xem là phương tiện phục vụ một nghiên cứu cụ thể, thay vì một tài sản khoa học có giá trị lâu dài cần được chuẩn hóa, quản trị và khai thác chung. Do đó, dữ liệu thường được xem là sản phẩm phục vụ phụ trong nghiên cứu hơn là một tài sản khoa học có giá trị lâu dài cần được quản trị, lưu trữ và khai thác dùng chung. Tình trạng phân tán và thiếu liên thông giữa các nguồn dữ liệu làm giảm khả năng tích hợp, đồng thời hạn chế việc hình thành các bộ dữ liệu có quy mô lớn phục vụ nghiên cứu liên ngành. Bên cạnh đó, năng lực khoa học dữ liệu của đội ngũ nghiên cứu còn nhiều hạn chế, từ thiết kế điều tra, quản lý và chuẩn hóa dữ liệu đến phân tích, khai thác và ứng dụng các công cụ phân tích hiện đại. Trong khi đó, lực lượng chuyên gia quản trị và phân tích dữ liệu trong lĩnh vực khoa học xã hội vẫn còn

thiếu, cùng với sự hạn chế trong cơ chế phối hợp giữa các cơ quan nghiên cứu, làm ảnh hưởng đến chất lượng nghiên cứu và hiệu quả tư vấn chính sách trong bối cảnh chuyển đổi số.

Những hạn chế trên xuất phát từ các nguyên nhân mang tính cấu trúc. Mô hình quản lý nghiên cứu theo ngành dọc trong thời gian dài đã tạo ra tình trạng phân tán dữ liệu và tư duy quản lý cục bộ giữa các cơ quan, tổ chức và viện nghiên cứu. Cơ chế đánh giá kết quả nghiên cứu cũng chủ yếu tập trung vào sản phẩm khoa học cuối cùng mà chưa coi dữ liệu là một sản phẩm nghiên cứu có giá trị cần được đầu tư, lưu trữ và chia sẻ.

6. Kết luận và kiến nghị

Đại hội XIV của Đảng đã xác định phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số gắn với phát triển kinh tế số, xã hội số và chính phủ số là động lực quan trọng thúc đẩy đổi mới mô hình tăng trưởng và nâng cao hiệu quả quản trị quốc gia (Đảng Cộng sản Việt Nam, 2026). Trong bối cảnh đó, dữ liệu khoa học xã hội ngày càng khẳng định vai trò của một hạ tầng tri thức phục vụ nghiên cứu, dự báo và hoạch định chính sách dựa trên bằng chứng. Kết quả nghiên cứu cho thấy trọng tâm của chuyển đổi số trong khoa học xã hội là xây dựng hệ thống dữ liệu có khả năng tích lũy, liên thông, chia sẻ và tái sử dụng, trong đó các cuộc điều tra cơ bản giữ vai trò nền tảng để hình thành các bộ dữ liệu dài hạn, phục vụ phân tích xu hướng và đánh giá tác động chính sách.

Bài viết khẳng định việc xây dựng cơ sở dữ liệu khoa học xã hội cần được triển khai như một chương trình chiến lược ở tầm quốc gia, thay cho cách tiếp cận phân tán theo từng đề tài hoặc nhiệm vụ nghiên cứu. Quá trình này đòi hỏi sự đồng bộ về thể chế, hạ tầng dữ liệu, nguồn nhân lực, cơ chế phối hợp và năng lực quản trị dữ liệu, đồng thời thúc đẩy ứng dụng dữ liệu lớn, trí tuệ nhân tạo và các phương pháp phân tích hiện đại nhằm nâng cao chất lượng nghiên cứu, dự báo và tư vấn chính sách. Đây là những điều kiện nền tảng để chuyển từ mô hình nghiên cứu phân tán sang hệ sinh thái dữ liệu phục vụ nghiên cứu, dự báo và hoạch định chính sách dựa trên bằng chứng. Trên cơ sở đó, bài viết đề xuất một số kiến nghị sau:

Một là, xây dựng cơ sở dữ liệu khoa học xã hội ở tầm quốc gia theo hướng tích hợp, liên thông và phát triển dài hạn, lấy các cuộc điều tra cơ bản định kỳ làm nền tảng. Hệ thống điều tra cần được chuẩn hóa về nội dung, phương pháp và quy trình thu thập nhằm hình thành chuỗi dữ liệu có khả năng so sánh theo thời gian, đáp ứng yêu cầu nghiên cứu, dự báo và hoạch định chính sách.

Hai là, xây dựng chương trình quốc gia về phát triển cơ sở dữ liệu khoa học xã hội, xác định các bộ dữ liệu ưu tiên gắn với những lĩnh vực chính sách trọng yếu; đồng thời hoàn thiện cơ chế phối hợp liên ngành, tiêu chuẩn dữ liệu và quy trình chia sẻ, kết nối giữa dữ liệu điều tra cơ bản với dữ liệu hành chính và dữ liệu thống kê chính thức.

Ba là, phát triển nguồn nhân lực khoa học xã hội số thông qua đào tạo đội ngũ nghiên cứu, chuyên gia quản trị dữ liệu và phân tích dữ liệu; đồng thời tăng cường đầu tư hạ tầng số, nền tảng lưu trữ và các công cụ phân tích hiện đại cho các cơ quan nghiên cứu khoa học xã hội.

Bốn là, hình thành mô hình liên thông “Nghiên cứu - Dữ liệu - Chính sách”, trong đó dữ liệu đóng vai trò hạ tầng kết nối hoạt động nghiên cứu với quá trình xây dựng, thực thi và đánh giá chính sách. Trên cơ sở đó, từng bước phát triển hệ sinh thái dữ liệu khoa học xã hội theo hướng mở, liên thông và dùng chung, bảo đảm khả năng tích hợp và khai thác hiệu quả giữa các cơ quan, tổ chức nghiên cứu.

Năm là, hoàn thiện khung pháp lý và cơ chế quản trị dữ liệu trong nghiên cứu khoa học xã hội theo hướng hài hòa giữa chia sẻ, khai thác dữ liệu và bảo đảm an toàn thông tin và đạo đức nghiên cứu. Đồng thời, xây dựng cơ chế giám sát, phân quyền truy cập và quản lý dữ liệu minh bạch nhằm bảo đảm việc khai thác dữ liệu an toàn, hiệu quả và bền vững.

Trong kỷ nguyên số, năng lực nghiên cứu và chất lượng hoạch định chính sách ngày càng phụ thuộc vào khả năng tạo lập, quản trị và khai thác dữ liệu. Xây dựng cơ sở dữ liệu khoa học xã hội trở thành một nhiệm vụ quan trọng và cấp bách nhằm hiện đại hóa hạ tầng tri thức, nâng cao năng lực quản trị quốc gia dựa trên bằng chứng và tạo nền tảng cho phát triển nhanh, bền vững đất nước.

Tài liệu tham khảo

- Bộ Chính trị. (2024). *Nghị quyết số 57-NQ/TW ngày 22 tháng 12 năm 2024 về đột phá phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia*. Hà Nội.
- Bộ Thông tin và Truyền thông. (2024). *Báo cáo chuyển đổi số quốc gia năm 2024*. Hà Nội: Bộ Thông tin và Truyền thông.
- Đảng Cộng sản Việt Nam (2026). *Văn kiện Đại hội đại biểu toàn quốc lần thứ XIV*. Tập I & II. Hà Nội: Nxb Chính trị quốc gia Sự thật.
- Head, B. W. (2016). Toward more “evidence-informed” policy making? *Public Administration Review*, 76(3), 472-484.
- Kitchin, R. (2014). *The Data Revolution: Big data, open data, data infrastructures and their consequences*. London: Sage Publications.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., and Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915), 721-723.
- Lê Thị Hoàng Minh và Vũ Công Giao. (2025). Đổi mới quản trị quốc gia theo hướng hiện đại, hiệu lực, hiệu quả trong bối cảnh chuyển đổi số. *Tạp chí Nghề Luật*, số 5/2025.
- Lê Thị Hoàng Minh và Vũ Công Giao. (2026). Quản trị quốc gia gắn với phát triển xã hội số ở Việt Nam hiện nay - Thực trạng và giải pháp hoàn thiện pháp luật. *Tạp chí Dân chủ và Pháp luật*, số 1(444), 1/2026.

OECD. (2019). *Enhancing access to and sharing of data: Reconciling risks and benefits for data reuse across societies*. Geneva: OECD Publishing.

Sutcliffe, S., and Court, J. (2005). *Evidence-based Policymaking: What is it? How does it work? What relevance for developing countries?* London: Overseas Development Institute (ODI).

World Bank. (2021). *World Development Report 2021: Data for better lives*. Washington, DC: World Bank.