

TRENDS IN PUBLISHING SOCIAL SCIENCE DATA IN VIETNAM: A SCOPUS - BASED ANALYSIS

Le Thi Tuyet Trinh*¹, Le Minh Cuong²,
Nguyen Thuy Phuong Tram³, Do Thi Trinh⁴,
Doan Thi Phuong Thuc⁵

* Corresponding author
Email: letrinh1282@gmail.com

² Email: lmcuong@dthu.edu.vn

^{1,2} Dong Thap University
783 Pham Huu Lau, Cao Lanh City,
Dong Thap province, Vietnam

³ Email: nguyenthuyphuongtramdt@gmail.com
Duc Trong Highschool, Lam Dong province
320 Highway 20, Lien Nghia, Duc Trong,
Lam Dong province, Vietnam

⁴ Email: trinhdt@tnue.edu.vn
Thai Nguyen University of Education
No. 20 Luong Ngoc Quyen street, Thai Nguyen city,
Thai Nguyen province, Vietnam

⁵ Email: dtpthuc@thanhdouni.edu.vn
Institute for Research on Education
and Knowledge Transfer, Thanh Do University
Highway 32, Hoai Duc, Hanoi, Vietnam

Received: 11/9/2024

Revised: 25/9/2024

Accepted: 09/11/2024

Published: 20/12/2024

Abstract: Research data is one of the most valuable outcomes of most research. The public release of research datasets in a free and open-access format has consequently become a fundamental practice within the open science movement. In this context, publishing data papers is an optimal choice, not only for disseminating data but also for providing the necessary descriptions to enable colleagues and the research community to easily find, access, and reuse it. Despite its significance, data publishing remains uncommon and relatively unfamiliar to Vietnamese scientists. This study examines data papers in the field of social sciences authored by Vietnamese researchers and published in journals indexed in the SCOPUS database. The findings reveal that only 57 data papers were recorded, with the annual number surpassing 10 papers from 2020 onward, a trend that coincides with the outbreak of the COVID-19 pandemic. Notably, the majority of these publications were in the field of education, with most papers appearing in the journal Data in Brief. Furthermore, this study identifies the funding sources for these data papers and discusses the importance of this practice for the Vietnamese academic community.

Keywords: Data publication, data sharing, SCOPUS, open science.

XU HƯỚNG CÔNG BỐ BÀI BÁO MÔ TẢ DỮ LIỆU TRONG LĨNH VỰC KHOA HỌC XÃ HỘI Ở VIỆT NAM: MỘT PHÂN TÍCH DỰA TRÊN SCOPUS

Lê Thị Tuyết Trinh*¹, Lê Minh Cường²,
Nguyễn Thuy Phuong Tram³, Đỗ Thị Trinh⁴,
Đoàn Thị Phuong Thuc⁵

* Tác giả liên hệ
Email: letrinh1282@gmail.com

² Email: lmcuong@dthu.edu.vn

^{1,2} Trường Đại học Đồng Tháp
783 Phạm Hữu Lầu, thành phố Cao Lãnh,
tỉnh Đồng Tháp, Việt Nam

³ Email: nguyenthuyphuongtramdt@gmail.com
Trường Trung học phổ thông Đức Trọng, tỉnh Lâm Đồng
320 quốc lộ 20, Liên Nghĩa, Đức Trọng,
tỉnh Lâm Đồng, Việt Nam

⁴ Email: trinhdt@tnue.edu.vn
Trường Đại học Sư phạm - Đại học Thái Nguyên
Số 20 đường Lương Ngọc Quyến, thành phố Thái Nguyên,
tỉnh Thái Nguyên, Việt Nam

⁵ Email: dtpthuc@thanhdouni.edu.vn
Viện Nghiên cứu Giáo dục và Chuyển giao tri thức,
Trường Đại học Thành Đô
Quốc lộ 32, Hoài Đức, Hà Nội, Việt Nam

Nhận bài: 11/9/2024

Chỉnh sửa xong: 25/9/2024

Chấp nhận đăng: 09/11/2024

Xuất bản: 20/12/2024

Tóm tắt: Dữ liệu nghiên cứu là một trong những sản phẩm giá trị nhất của phần lớn các nghiên cứu. Việc công khai các bộ dữ liệu nghiên cứu dưới dạng miễn phí và truy cập mở do đó trở thành một trong những thực hành cốt lõi của phong trào khoa học mở. Trong bối cảnh này, xuất bản các bài báo dữ liệu là một lựa chọn tối ưu, không chỉ để công bố dữ liệu mà còn cung cấp các mô tả cần thiết giúp đồng nghiệp và cộng đồng nghiên cứu dễ dàng tìm kiếm, truy cập và tái sử dụng. Tuy nhiên, việc xuất bản dữ liệu vẫn là một thực hành chưa phổ biến và còn xa lạ đối với các nhà khoa học Việt Nam. Nghiên cứu này tập trung xem xét các bài báo dữ liệu thuộc lĩnh vực Khoa học xã hội của các tác giả Việt Nam được công bố trên các tạp chí được chỉ mục trong cơ sở dữ liệu SCOPUS. Kết quả cho thấy, có 57 bài báo dữ liệu được ghi nhận, với số lượng vượt mốc 10 bài/năm từ năm 2020, trùng với thời điểm đại dịch COVID-19 bùng phát, trong đó các công bố nổi bật chủ yếu thuộc lĩnh vực giáo dục. Phần lớn các bài báo được công bố trên Tạp chí Data in Brief. Ngoài ra, nghiên cứu cũng liệt kê các nguồn tài trợ cho các bài báo dữ liệu và thảo luận về tầm quan trọng của thực hành này đối với cộng đồng học thuật Việt Nam.

Từ khóa: Công bố dữ liệu, chia sẻ dữ liệu, SCOPUS, khoa học mở.

1. Đặt vấn đề

Khoa học giáo dục nói riêng và khoa học xã hội nói chung đều đang đối mặt với một vấn đề lớn là không thể tái lập (replicate) dẫn tới sự không khả thi trong nhân rộng kết quả và khái quát hóa giá trị từ các công trình nghiên cứu. Một trong những nguyên nhân của vấn đề này đó là tiếp cận nghiên cứu không minh bạch [1]. Trong bối cảnh đó, khoa học mở đã trở thành trào lưu mạnh mẽ và không thể đảo ngược trên toàn thế giới [2]. Phong trào này khuyến khích các nhà khoa học chia sẻ không chỉ kết quả cuối cùng mà còn cả dữ liệu cơ bản và phương pháp luận. Từ đó, công bố dữ liệu đang trở thành một tiêu chuẩn mới trong cộng đồng khoa học [3]. Xu hướng này cho thấy sự công nhận giá trị của việc chia sẻ dữ liệu nghiên cứu một cách cởi mở, không chỉ như một phần bổ sung cho các xuất bản phẩm truyền thống mà còn là một nguồn tài nguyên quan trọng theo đúng nghĩa của nó. Trước sự quan tâm về khoa học mở và công bố dữ liệu ở quốc tế, xu hướng này ít được chú ý ở Việt Nam. Do đó, nghiên cứu này nhằm mục đích cung cấp các thông tin tổng quan về xu hướng công bố dữ liệu trong thực hành khoa học và hiện trạng công bố dữ liệu ở Việt Nam, trong đó xem xét năng suất và sức ảnh hưởng của các công bố dữ liệu từ các nhà nghiên cứu khoa học xã hội, bao gồm cả khoa học giáo dục trong nước đã tham gia vào xu hướng này.

2. Nội dung nghiên cứu

2.1. Tổng quan về công bố dữ liệu

Trước đây, các phát hiện khoa học thường được chia sẻ thông qua trước tác nhưng dữ liệu thô thường được các nhà nghiên cứu tạo ra chúng lưu giữ. Trong một số ngành Khoa học xã hội như Tâm lý và Y học, việc chia sẻ dữ liệu lại càng ít phổ biến hơn so với các ngành Khoa học tự nhiên và Khoa học vũ trụ do lo ngại về quyền sở hữu dữ liệu, quyền riêng tư và quyền kiểm soát nghiên cứu [4], [5]. Trên thực tế, khái niệm chia sẻ dữ liệu không hoàn toàn mới nhưng việc chính thức hóa và áp dụng rộng rãi của nó mới tương đối gần đây được thúc đẩy bởi nhiều yếu tố khác nhau bao gồm tiến bộ công nghệ (sự ra đời của Internet và sự phát triển của cơ sở dữ liệu số), sự thừa nhận tầm quan trọng của khoa học mở, những thay đổi trong chính sách tài trợ nghiên cứu (Ví dụ: Kế hoạch S [6]) và đại dịch COVID-19.

Việc chia sẻ dữ liệu đến nay vẫn gặp nhiều thách thức việc chia sẻ dữ liệu không phải là không có thách thức. Các nhà nghiên cứu thường phải đối mặt với những rào cản như cơ sở hạ tầng không đầy đủ,

lo ngại về việc lạm dụng dữ liệu cũng như nhu cầu trích dẫn và ghi nhận phù hợp những người đóng góp dữ liệu [7]. Một trong những cách khắc phục và chính thức hóa việc chia sẻ dữ liệu, đảm bảo sự ghi nhận cho tác giả đó là công bố dữ liệu.

Công bố dữ liệu liên quan đến việc cung cấp các bộ dữ liệu thông qua các tạp chí hoặc kho dữ liệu chuyên dụng, nơi đảm bảo rằng, dữ liệu được quản lý, đánh giá ngang hàng và thuận tiện cho cộng đồng khoa học truy cập và sử dụng [8]. Thực tiễn này giúp tăng cường tính minh bạch, thúc đẩy khả năng tái lập (Replication) và tăng tốc đổi mới bằng cách cho phép các nhà nghiên cứu khác xác minh các phát hiện, tiến hành phân tích mới và đặc biệt là khám phá các ứng dụng liên ngành. Qua đó, công bố dữ liệu đang làm thay đổi cục diện nghiên cứu khoa học và thúc đẩy sự hợp tác.

Phù hợp với mục tiêu minh bạch và thúc đẩy hợp tác, nhiều tạp chí chuyên về công bố data trở thành thành phần thiết yếu của hệ sinh thái khoa học mở [9]. Các tạp chí riêng về công bố dữ liệu cũng tập trung vào việc xuất bản các bộ dữ liệu hoàn chỉnh, cung cấp các mô tả, phương pháp và siêu dữ liệu chi tiết để đảm bảo cộng đồng nghiên cứu rộng lớn hơn có thể truy cập và tái sử dụng. Các tạp chí dữ liệu cũng có quy trình đánh giá ngang hàng. Tuy nhiên, không giống như các bài báo nghiên cứu truyền thống (Research article) tập trung vào giả thuyết, phương pháp và kết quả, với các bài báo dữ liệu (Data paper), phản biện viên sẽ xem xét về quá trình và phương pháp thu thập, chất lượng và khả năng tiếp cận dữ liệu.

Cơ sở dữ liệu SCOPUS cũng ghi nhận bài công bố dữ liệu là một loại tài liệu được chỉ mục và khác biệt với bài báo nghiên cứu. Trong đó, hai loại tài liệu được SCOPUS định nghĩa và xác định tính chất như sau:

Article (tạm dịch: Bài báo nghiên cứu): Được định nghĩa là “Các công trình nghiên cứu hoặc ý kiến chính bản/chính gốc”. Các đặc điểm của Article bao gồm: “Các bài báo trên các tạp chí được đánh giá ngang hàng, thường có độ dài nhiều trang được chia thành các phần: Tóm tắt (Abstract), giới thiệu (Introduction), phương pháp (Materials & methods), kết quả (Results), kết luận (Conclusions), thảo luận (Discussion) và tài liệu tham khảo (References). Tuy nhiên, các báo cáo trường hợp (Case reports), ghi chú kỹ thuật (Technical & research notes) cũng như các thông tin truyền thông ngắn gọn cũng được coi là bài báo và có thể chỉ dài một trang. Các bài báo trên tạp chí thương mại thường ngắn hơn so với các

tạp chí được bình duyệt và cũng có thể ngắn gọn bằng một trang.”

Data paper (tạm dịch: Bài báo dữ liệu): Được định nghĩa là: “Các tài liệu về siêu dữ liệu (*Metadata*) có thể tìm kiếm, trong đó mô tả một bộ dữ liệu hoặc tập hợp các bộ dữ liệu có thể truy cập trực tuyến”. Các đặc điểm của *Data paper* bao gồm: “Mục đích của bài báo dữ liệu là cung cấp thông tin mô tả về (các) bộ dữ liệu liên quan, tập trung vào việc thu thập dữ liệu, phân biệt các khía cạnh, cách truy cập và khả năng tái sử dụng thay vì các thông tin về xử lý và phân tích dữ liệu” [10].

2.2. Phương pháp nghiên cứu

Nghiên cứu này sử dụng phương pháp thống kê mô tả về số lượng và số lượt trích dẫn của các tài liệu công bố, nguồn công bố và các thông tin về quỹ tài trợ nhằm xác định xu hướng, năng suất, sức ảnh hưởng của các công bố và các nguồn công bố dữ liệu tiêu biểu. Dữ liệu của nghiên cứu này được trích xuất từ cơ sở dữ liệu SCOPUS, là hệ thống chỉ mục có uy tín và cách thức tổ chức dữ liệu phù hợp với mục đích nghiên cứu khi phân loại tài liệu tách biệt bài công bố dữ liệu (*Data paper*) và bài báo khoa học (*Article*). Chúng tôi đã tiến hành tìm kiếm và truy xuất dữ liệu theo ba tiêu chí sau: 1/ Công bố có tác giả đến từ các đơn vị, cơ quan và tổ chức tại Việt Nam; 2/ Nội dung và phạm vi nghiên cứu thuộc các lĩnh vực khoa học xã hội; 3/ Các công bố được SCOPUS phân loại là công bố dữ liệu (*Data paper*).

Dựa trên các tiêu chí trên, lệnh truy vấn dữ liệu (20h00 ngày 23 tháng 8 năm 2024) được sử dụng như sau: AFFIL (“viet nam” OR vietnam) AND DOCTYPE (dp) AND (LIMIT-TO (SUBJAREA, “MULT”) OR LIMIT-TO (SUBJAREA, “SOCI”) OR LIMIT-TO (SUBJAREA, “DECI”).

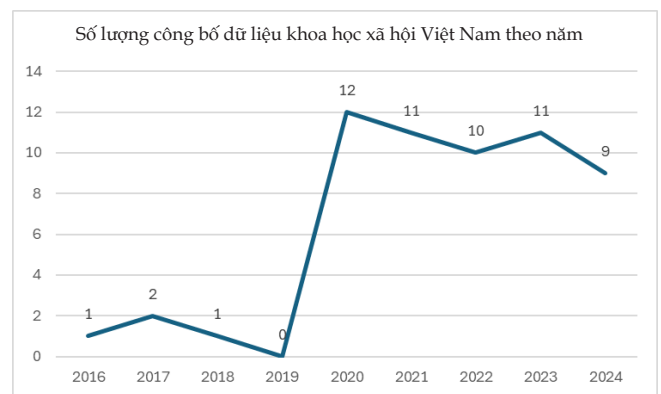
Kết quả từ SCOPUS bao gồm 131 bài công bố dữ liệu được truy xuất thành tệp Excel, sau đó chúng tôi tiến hành đọc, rà soát tiêu đề và tóm tắt để lọc các kết quả để đảm bảo ba yêu cầu trên. Sau quá trình này, 74 tài liệu bị loại do không đáp ứng tiêu chí thuộc lĩnh vực khoa học xã hội, 57 tài liệu hợp lệ được sử dụng để tiếp tục phân tích.

2.3. Kết quả nghiên cứu

2.3.1. Năng suất công bố theo năm

Kết quả thống kê các công bố dữ liệu theo thời gian được trình bày tại Hình 1. Trong số 57 công bố, công bố dữ liệu đầu tiên của tác giả Việt Nam được SCOPUS ghi nhận vào năm 2016 với tiêu đề “*Data on Vietnamese patients’ financial burdens and risk*

of destitution” đăng tải trên tạp chí *Data in Brief* của hai tác giả Vương Quân Hoàng và Nguyễn Trọng Khang. Công bố này chia sẻ bộ dữ liệu qua hai lần triển khai khảo sát các bệnh nhân từ bệnh viện miền Bắc với quy mô lần 1 với 330 bản ghi và lần 2 với 900 bản ghi. Bộ dữ liệu bao gồm các thông tin cụ thể về chi phí tự chi trả của bệnh nhân, tỉ lệ chi phí này so với thu nhập hộ gia đình của họ và hậu quả tài chính kéo theo. Ví dụ: Nguy cơ rơi vào nghèo đói. Bộ dữ liệu có thể là tài nguyên cho các nhà hoạch định chính sách và nhà nghiên cứu quan tâm đến việc tìm hiểu và giải quyết các thách thức tài chính liên quan đến chăm sóc sức khỏe tại Việt Nam [11].



Hình 1: Số lượng công bố dữ liệu trong khoa học xã hội Việt Nam theo năm

Giai đoạn sau đó, từ năm 2016 đến năm 2019, số lượng công bố không có sự thay đổi với trung bình 01 tài liệu/năm. Tuy vậy, từ năm 2020 ghi nhận sự tăng trưởng đột biến từ 0 tài liệu vào năm 2019 lên 13 tài liệu trong năm 2020. Đáng chú ý là, trong số 13 công bố trong năm 2020 có 07 tài liệu về dữ liệu trong lĩnh vực giáo dục với đối tượng khảo sát là giáo viên và học sinh. Đồng thời, 08 công bố ở năm 2020 có đề cập về COVID-19 (dựa trên xem xét phần tiêu đề và tóm tắt).

Từ năm 2020, năng suất công bố dữ liệu được duy trì từ 10 bài/năm. Tuy rằng, năng suất và tổng số lượng bài như trên không phải số lượng lớn nhưng điều này cũng cho thấy xu hướng chia sẻ dữ liệu đã nhận được sự chú ý từ các nhà nghiên cứu khoa học xã hội tại Việt Nam.

2.3.2. Các nguồn công bố chính

Phần lớn các bài báo dữ liệu của các tác giả đến từ Việt Nam được công bố trên Tạp chí *Data in Brief* với 54 trong 57 bài báo được phân tích được công bố trên tạp chí này. *Data in Brief* là một tạp chí truy cập mở đa ngành thuộc nhà xuất bản Elsevier, xuất bản các

bài báo dữ liệu ngắn gọn đa ngành nhằm mô tả và cung cấp con đường truy cập tới bộ dữ liệu nghiên cứu (xem Bảng 1).

Bảng 1: Danh sách các nguồn công bố dữ liệu

Nguồn/Tạp chí	Số công bố
Data in Brief	54
Scientific Data	2
Frontiers in Education	1

Chỉ có 3 bài báo được công bố trên tạp chí khác Data in Brief, trong đó có 02 bài được xuất bản trên Tạp chí Scientific Data (Nhà Xuất bản Nature) và 01 bài trên Tạp chí Frontiers in Education (Nhà Xuất bản Frontiers). Tương tự với Data in Brief, Scientific Data là tạp chí chuyên xuất bản các bài báo mô tả dữ liệu đa ngành. Trong khi đó, Frontiers in Education là tạp chí xuất bản các bài báo nghiên cứu giáo dục nói chung, trong đó có chấp nhận loại bài báo cáo dữ liệu (Data report).

2.3.3. Các công bố tiêu biểu

Dựa vào chỉ số trích dẫn của các công bố, chúng tôi đã xác định các công bố dữ liệu tiêu biểu nhất của các nhà khoa học xã hội Việt Nam. Bảng 2 liệt kê 10 tài liệu có số lượng trích dẫn nhiều nhất.

Đứng đầu danh sách này là tài liệu *Dataset of Vietnamese student's learning habits during COVID-19* của nhóm tác giả Trần Trung, Hoàng Anh Đức, Nguyễn Tiến Trung, Đinh Việt Hùng, Nguyễn Yến Chi và Phạm Hùng Hiệp công bố trên Tạp chí Data in Brief. Tài liệu được SCOPUS ghi nhận có 54 lượt trích dẫn. Nghiên cứu này trình bày bộ dữ liệu tìm hiểu thói quen học tập của học sinh Việt Nam trong thời gian trường học tạm dừng hoạt động do đại dịch COVID-19. Dữ liệu được thu thập nhằm tìm hiểu cách học sinh thích nghi với việc học từ xa và các yếu tố khác nhau ảnh hưởng đến thói quen học tập của họ trong một thời kì mang tính khủng hoảng. Bộ dữ liệu bao gồm thông tin về các loại thiết bị được sử dụng để học tập, tần suất và thời lượng của các buổi học cũng như các môn học mà học sinh tập trung vào. Nó cũng xem xét những thách thức mà sinh viên phải đối mặt, chẳng hạn như vấn đề kết

Bảng 2: Danh sách 10 công bố dữ liệu tiêu biểu của các nhà khoa học xã hội Việt Nam

STT	Tác giả	Tiêu đề	Năm	Tạp chí	Lượt trích dẫn
1	Trần Trung, Hoàng Anh Đức, Nguyễn Tiến Trung, Đinh Việt Hùng, Nguyễn Yến Chi và Phạm Hùng Hiệp	Dataset of Vietnamese student's learning habits during COVID-19	2020	Data in Brief	54
2	Nguyễn Văn Duy, Phạm Hoàng Giang, Nguyễn Ngọc Đạt	Impact of the COVID-19 pandemic on perceptions and behaviors of university students in Vietnam	2020	Data in Brief	43
3	Vũ Cẩm Tú, Hoàng Anh Đức, Thân Văn Quân, Nguyễn Mạnh Tuấn, Đinh Việt Hùng, Lê Thị Quỳnh Anh, Lê Thị Thu Trang, Phạm Hùng Hiệp, Nguyễn Yến Chi	Dataset of Vietnamese teachers' perspectives and perceived support during the COVID-19 pandemic	2020	Data in Brief	40
4	Bùi Thế Hợp, Lương Đình Hải, Nguyễn Xuân An, Nguyễn Hồng Liên, Ngô Thanh Thủy	Impact of female students' perceptions on behavioral intention to use video conferencing tools in COVID-19: Data of Vietnam	2020	Data in Brief	31
5	Huỳnh Lưu Đức Toàn	Data for understanding the risk perception of COVID-19 from Vietnamese sample	2020	Data in Brief	30

STT	Tác giả	Tiêu đề	Năm	Tạp chí	Lượt trích dẫn
6	Hsinguang Chi, Vũ Văn Thịnh, Tan Vo-Thanh, Nguyễn Phong Nguyên, Nguyễn Văn Duy	Workplace health and safety training, employees' risk perceptions, behavioral safety compliance, and perceived job insecurity during COVID-19: Data of Vietnam	2020	Data in Brief	20
7	Phạm Huyền, Trần Quế Nhi, La Gia Long, Đoàn Hà My, Vũ Tiến Đức	Readiness for digital transformation of higher education in the COVID-19 context: The dataset of Vietnam's students	2021	Data in Brief	14
8	Vương Quân Hoàng, Nguyễn Trọng Khang	Data on Vietnamese patients' financial burdens and risk of destitution	2016	Data in Brief	12
9	Nguyễn Thanh Hiếu, Dương Công Doanh	Dataset on the effect of perceived educational support on entrepreneurial intention among Vietnamese students	2021	Data in Brief	12
10	Khúc Văn Quý, Vũ Phú Trí, Lưu Phương	Dataset on the Hanoian suburbanites' perception and mitigation strategies towards air pollution	2020	Data in Brief	12

nổi Internet và thiếu khả năng tiếp cận tài liệu học tập. Bộ dữ liệu này với 450 bản ghi có thể phục vụ các mục tiêu nghiên cứu tìm hiểu tác động của đại dịch đối với việc học tập của học sinh và phát triển các chiến lược nhằm cải thiện giáo dục từ xa [12].

Đứng thứ hai là công bố của nhóm tác giả Nguyễn Văn Duy, Phạm Hoàng Giang, Nguyễn Ngọc Đạt trên Tạp chí Data in Brief. Công bố có tiêu đề "*Impact of the COVID-19 pandemic on perceptions and behaviors of university students in Vietnam*" cung cấp một bộ dữ liệu chi tiết tìm hiểu xem đại dịch đã ảnh hưởng như thế nào đến sinh viên đại học Việt Nam. Nghiên cứu đã thu thập 440 phản hồi từ sinh viên về nhận thức và thay đổi hành vi để ứng phó với đại dịch, đặc biệt trong các vấn đề như thói quen làm việc/học tập, kế hoạch du lịch và hoạt động hàng ngày. Tác giả đề cập bộ dữ liệu có thể cung cấp thông tin cần thiết và hướng dẫn ngành du lịch xây dựng giải pháp phục hồi sau khủng hoảng [13].

Trong 10 tài liệu có lượng trích dẫn cao nhất, có duy nhất một tài liệu do tác giả đứng tên. Công bố trên Tạp chí Data in Brief với tiêu đề "*Data for understanding the risk perception of COVID-19 from Vietnamese sample*" được ghi nhận 30 lượt trích dẫn và đứng trong top 5 của danh sách trên. Tác giả Huỳnh Lưu Đức Toàn trình bày một bộ dữ liệu làm sáng tỏ cách người dân Việt Nam nhìn nhận những rủi ro liên quan đến COVID-19 trong giai đoạn đầu

của đại dịch. Cuộc khảo sát được thực hiện vào ngày 01 tháng 02 năm 2020, ngay sau khi Thủ tướng Việt Nam ban bố các kịch bản khẩn cấp quốc gia thu về 391 phản hồi từ các cá nhân từ 15 đến 47 tuổi. Khảo sát tập trung vào nhận thức về nguy cơ, các yếu tố kinh tế xã hội và tiếp nhận các thông tin truyền thông. Tác giả đề xuất rằng, bộ dữ liệu nghiên cứu về vai trò của truyền thông trong định hình nhận thức của cộng đồng trong giai đoạn đầu của đại dịch COVID-19. Điều đó khiến bộ dữ liệu có giá trị hữu dụng cho các nhà hoạch định chính sách và các quan chức y tế công cộng quan tâm đến việc khám phá mối quan hệ giữa nhận thức rủi ro và phản ứng hành vi trong bối cảnh các nước đang phát triển đồng thời cung cấp cơ sở cho việc so sánh xuyên quốc gia và nghiên cứu theo chiều dọc về tác động của truyền thông và các yếu tố kinh tế xã hội đối với các phản ứng về sức khỏe cộng đồng [14]. Điều đặc biệt là, khác với hai tài liệu được giới thiệu bên trên, bài công bố này cung cấp một tập dữ liệu đã được tác giả sử dụng trong một công bố khác của mình [15].

2.3.4. Các nguồn tài trợ

Các nguồn tài trợ được ghi nhận trong số 57 bài báo được phân tích được liệt kê chi tiết ở Bảng 3. 26 bài báo trong số 57 bài được phân tích có công nhận nguồn tài trợ cho nghiên cứu. Trong đó, xuất hiện nhiều nhất là Đại học Kinh tế Quốc dân với 07 bài ghi nhận nguồn tài trợ này. 02 bài được công bố thuộc

diện đề tài cấp Nhà nước được tài trợ bởi Quỹ Phát triển Khoa học và Công nghệ Quốc gia (NAFOSTED), 05 bài được tài trợ bởi Bộ Giáo dục và Đào tạo. Tất cả các bài báo ghi nhận được tài trợ tài chính đều là các bài công bố trên Tạp chí Data in Brief.

Xem xét nguồn gốc của các cơ quan tài trợ, trong số 28 bài có ghi nhận hỗ trợ tài chính, có 05 bài nhận được tài trợ từ một cơ sở nước ngoài, 20 bài được tài trợ bởi một cơ quan Việt Nam và 01 bài ghi nhận sự hỗ trợ tài chính từ cả trong nước lẫn ngoài nước.

2.4. Thảo luận

Nhìn chung, tổng số công bố dữ liệu được SCOPUS chỉ mục là 57 tài liệu, là một con số vẫn còn khiêm tốn. Năm 2020, chúng kiến sự gia tăng đáng

kể trong việc xuất bản các tài liệu dữ liệu, phần lớn là do tác động toàn cầu của đại dịch COVID-19. Các nhà nghiên cứu trên nhiều lĩnh vực khác nhau đã nhanh chóng thu thập, phân tích và chia sẻ dữ liệu để hiểu sự lây lan của vi-rút, ảnh hưởng của nó đối với sức khỏe cộng đồng và các tác động kinh tế xã hội rộng hơn. Tình huống chưa từng có này đã dẫn đến sự bùng nổ trong nghiên cứu dựa trên dữ liệu với nhiều nghiên cứu tập trung vào việc lập mô hình kết quả và tìm hiểu hành vi của cộng đồng cũng như nhận thức về rủi ro trong đại dịch [13], [14]. Tình hình cấp bách đòi hỏi phải phổ biến dữ liệu nhanh chóng để cung cấp thông tin cho các quyết định và hoạch định chính sách về y tế công cộng, dẫn đến khối lượng tài liệu dữ liệu được xuất

Bảng 3: Nguồn tài trợ công bố dữ liệu trong khoa học xã hội Việt Nam

Nguồn tài trợ	Số bài tài trợ	Quốc gia
Đại học Kinh tế Quốc dân	7	Việt Nam
Bộ Giáo dục và Đào tạo	3	Việt Nam
Trường Đại học Hàng Hải	2	Việt Nam
Đại học Sư phạm Quốc gia Gwangju	2	Hàn Quốc
Quỹ Phát triển Khoa học và Công nghệ Quốc gia NAFOSTED	2	Việt Nam
Đại học Kinh tế Thành phố Hồ Chí Minh	2	Việt Nam
British Academy	1	Anh
Đại học Đà Nẵng	1	Việt Nam
Bộ Ngoại giao, Khối Thịnh vượng chung và Phát triển (Foreign, Commonwealth and Development Office - FCDO)	1	Anh
Trường Đại học Ngoại thương	1	Việt Nam
Trường Đại học FPT	1	Việt Nam
Bộ Khoa học Công nghệ Đài Loan	1	Đài Loan (Trung Quốc)
Chương trình Khoa học và Công nghệ Quốc gia giai đoạn kế hoạch 5 năm lần thứ 12	1	Việt Nam
Trường Đại học Thương mại	1	Việt Nam
Quỹ tài trợ Nghiên cứu đặc biệt Đại học Ghent Gent-BOF (Ghent University Special Research Fund)	1	Bỉ
Cơ quan Phát triển và Hợp tác Thụy Sĩ (Swiss Agency for Development and Cooperation - SDC)	1	Thụy Sĩ
Chương trình Nghiên cứu CGIAR về Hệ thống Nông sản Lúa gạo (CGIAR Research Program on Rice)	1	Tổ chức liên quốc gia

bản trong năm đó tăng mạnh. Những bài viết này rất quan trọng trong việc tạo điều kiện cho phản ứng khoa học tập thể đối với đại dịch, cho phép các nhà nghiên cứu trên toàn thế giới tiếp cận và xây dựng dựa trên công trình của nhau. Tuy rằng, số lượng công bố từ Việt Nam chưa phải lớn, sự phát triển này cũng phản ánh xu hướng chung của khoa học thế giới vào thời điểm đó. Đại dịch COVID-19 khiến giới khoa học buộc phải đẩy nhanh tốc độ chia sẻ thông tin với các thực hành chia sẻ dữ liệu được coi là rất quan trọng [16].

Về các công bố tiêu biểu, chúng tôi đã xác định 10 tài liệu tiêu biểu qua chỉ số lượng trích dẫn và giới thiệu cụ thể hơn về 03 tài liệu nhằm làm nổi bật các đặc tính của một bài công bố dữ liệu. Trong đó, tài liệu "Data for understanding the risk perception of COVID-19 from Vietnamese sample" minh họa một cách tiếp cận khác trong việc chia sẻ dữ liệu so với hai bài viết còn lại. Trong khi các bài viết trước đó cung cấp các bộ dữ liệu thô phục vụ các phân tích trong tương lai của các nhà nghiên cứu khác, tài liệu này lại chia sẻ một tập dữ liệu đã được áp dụng trong một nghiên cứu đã xuất bản trước đó. Điều này cho thấy một cách tiếp cận tích hợp hơn để chia sẻ dữ liệu, trong đó tác giả không chỉ cung cấp dữ liệu mà còn thể hiện ứng dụng của nó trong nghiên cứu trong thế giới thực. Cách thức này cung cấp một ví dụ rõ ràng về cách sử dụng bộ dữ liệu, đưa ra những lý thuyết sâu sắc để hiểu và các phát hiện mà các nhà nghiên cứu khác có thể dựa vào. Bằng cách chia sẻ dữ liệu đã đóng góp cho nghiên cứu được công bố, tác giả cung cấp nguồn dữ liệu đã được kiểm nghiệm và xác nhận với mong muốn tăng giá trị và khả năng sử dụng của nó trong tương lai. Thực hành này làm tăng cường tính minh bạch và khuyến khích các nghiên cứu nhân rộng và tái lập, cho phép cộng đồng học thuật xác minh, thách thức hoặc mở rộng dựa trên những phát hiện ban đầu.

Trong nghiên cứu này, chúng tôi cũng xem xét các nguồn tài trợ cho các công trình nghiên cứu có công bố dữ liệu. Trên thế giới, các nguồn tài trợ đã và đang đóng vai trò quan trọng trong xu hướng công bố dữ liệu. Sự ra đời của Kế hoạch S (Plan S) vào năm 2018 [6] đã tạo ra một bước ngoặt lớn trong phong trào khoa học mở khi liên minh gồm các quỹ nghiên cứu lớn nhất thế giới này yêu cầu các nghiên cứu được họ tài trợ bắt buộc phải công khai kết quả nghiên cứu công khai miễn phí. Bên cạnh đó, hướng dẫn của Quỹ này cũng khuyến khích các nhà nghiên cứu công khai bộ dữ liệu nghiên cứu theo nhiều cách thức. Trong những năm gần đây, việc chia sẻ

công khai dữ liệu đã dần trở thành một tiêu chuẩn bắt buộc ở một số quỹ nghiên cứu lớn. Chẳng hạn, Quỹ Khoa học Quốc gia Hoa Kỳ (National Science Foundation) yêu cầu trong hồ sơ xin tài trợ bắt buộc phải bao gồm một kế hoạch quản lý dữ liệu (Data management plan) trong đó nêu rõ cách thức các dữ liệu của nghiên cứu được lưu trữ và chia sẻ [17].

3. Kết luận

Nghiên cứu cho thấy một góc nhìn tổng quan về hiện trạng công bố dữ liệu trong lĩnh vực khoa học xã hội của Việt Nam. Qua quá trình nghiên cứu về xu hướng công bố và thống kê các tài liệu liên quan được chỉ mục trong cơ sở dữ liệu SCOPUS, kết quả cho thấy trào lưu khoa học mở và xu hướng chia sẻ dữ liệu đóng vai trò quan trọng đối với các lĩnh vực khoa học xã hội. Trong bối cảnh đó, công bố dữ liệu là cách thức giúp chia sẻ dữ liệu một cách minh bạch và có giá trị khi được trải qua đánh giá ngang hàng và đăng tải trên các tạp chí uy tín giúp gia tăng chỉ số trích dẫn cho nhà khoa học, đồng thời tăng sự hiện thị của dữ liệu, thúc đẩy sự hợp tác và các khám phá xa hơn. Hoạt động công bố dữ liệu ở Việt Nam đã có và đi theo xu hướng của quốc tế. Tuy nhiên, số lượng vẫn còn khiêm tốn và chưa được phổ biến. Trong đó, chính sách và các quỹ tài trợ có thể đóng vai trò lớn thúc đẩy sự phát triển của xu hướng này.

Mặc dù có các đóng góp như trên song nghiên cứu này vẫn còn hạn chế khi hệ thống phân loại của SCOPUS vẫn chưa hoàn toàn chính xác về bài báo nghiên cứu và bài báo dữ liệu, dẫn đến bỏ sót một số bài công bố dữ liệu được phân loại thành bài báo nghiên cứu trong cơ sở dữ liệu này. Điển hình như một số bài báo dữ liệu đăng tải trên Tạp chí Data (Nhà Xuất bản MPDI) được SCOPUS phân loại thành bài báo nghiên cứu (Article), do đó không nằm trong dữ liệu trích xuất của nghiên cứu này. Các nghiên cứu tương lai về vấn đề này có thể quan tâm và xây dựng câu lệnh tìm kiếm bao quát hơn. Cuối cùng, dựa trên kết quả nghiên cứu, chúng tôi hi vọng việc công bố dữ liệu sẽ được quan tâm và trở thành phong trào phổ biến trong hoạt động khoa học của Việt Nam. Nghiên cứu này cũng giới thiệu các nguồn công bố tiêu biểu và nội dung cách thức đa dạng của các công bố tiêu biểu, có thể đóng vai trò như nguồn tham khảo cơ bản cho các nhà nghiên cứu Việt Nam muốn tham gia vào phong trào này.

Lời cảm ơn: *Nghiên cứu này được tài trợ bởi Quỹ Phát triển Khoa học và Công nghệ Quốc gia (NAFOSTED) trong đề tài mã số 503.01- 2021.27.*

Tài liệu tham khảo

- [1] Van der Zee, T., & Reich, J., (2018), *Open Education Science*. *Science*, 359(6376), 794-795.
- [2] Vicente-Saez, R., & Martinez-Fuentes, C., (2018), *Open Science now: A systematic literature review for an integrated definition*, *Journal of Business Research*, 88, 428-436, <https://doi.org/10.1016/j.jbusres.2017.12.043>
- [3] The World Bank (n.d.), *Publishing Data*, Dimewiki, Retrieved from https://dimewiki.worldbank.org/Publishing_Data.
- [4] Michener, W., (2015), *Ecological data sharing*, *Ecol. Informatics*, 29, 33-44, <https://doi.org/10.1016/j.ecoinf.2015.06.010>.
- [5] Hauser, R., (2018), *Sharing data: Some examples*, *Chinese Journal of Sociology*, 4, 262 – 276, <https://doi.org/10.1177/2057150X18762025>.
- [6] Burgelman, J., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M., (2019), *Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century*, *Frontiers in Big Data*, 2, 469872, <https://doi.org/10.3389/fdata.2019.00043>.
- [7] Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., Read, E., Manoff, M., & Frame, M., (2011), *Data Sharing by Scientists: Practices and Perceptions*, *PLoS ONE*, 6. <https://doi.org/10.1371/journal.pone.0021101>
- [8] Borgman, C. L., (2012), *The conundrum of sharing research data*, *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.
- [9] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B., (2016), *The FAIR Guiding Principles for scientific data management and stewardship*, *Scientific data*, 3, 160018.
- [10] Elsevier. (n.d.). *What content is indexed in SCOPUS?* Retrieved from https://service.elsevier.com/app/answers/detail/a_id/11274/c/10547/supporthub/SCOPUS/kw/Data+paper/
- [11] Vuong, Q., & Nguyen, T., (2016), *Data on Vietnamese patients' financial burdens and risk of destitution*, *Data in Brief*, 9, 543-548. <https://doi.org/10.1016/j.dib.2016.09.040>
- [12] Trung, T., Hoang, A. D., Nguyen, T. T., Dinh, V. H., Nguyen, Y. C., & Pham, H. H., (2020), *Dataset of Vietnamese student's learning habits during COVID-19*, *Data in Brief*, 30, 105682.
- [13] Van Nguyen, D., Pham, G. H., & Nguyen, D. N., (2020), *Impact of the Covid-19 pandemic on perceptions and behaviors of university students in Vietnam*, *Data in Brief*, 31, 105880.
- [14] Huynh, T. L. D., (2020), *Data for understanding the risk perception of COVID-19 from Vietnamese sample*, *Data in brief*, 30, 105530.
- [15] Huynh, T. L., (2020), *The COVID-19 risk perception: A survey on socioeconomics and media attention*, *Econ. Bull.*, 40(1), 758-764
- [16] Strcic, J., Civljak, A., Glozanic, T. et al, (2022), *Open data and data sharing in articles about COVID-19 published in preprint servers medRxiv and bioRxiv*, *Scientometrics* 127, 2791–2802, <https://doi.org/10.1007/s11192-022-04346-1>.
- [17] U.S. National Science Foundation (n.d.), *Chapter VI - Other Post Award Requirements and Considerations*. Office of Budget Finance and Award Management (BFA) Retrieved from https://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4.