Public Disclosure Authorized
Public Disclosure Authorized
Public Disclosure Authorized
Public Disclosure Authorized



CHƯƠNG TRÌNH ĐỐI TÁC CHIẾN LƯỢC ÔTXTRÂYLIA – NHÓM NGÂN HÀNG THẾ GIỚI

Chuyên đề Phát triển Số

Kích hoạt giá trị dữ liệu mở:
Động lực mới cho đổi mới sáng tạo và kinh tế dữ liệu
Báo cáo đánh giá dữ liệu mở ở Việt Nam

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

© 2025 Ngân hàng Tái thiết và Phát triển Quốc tế / Ngân hàng Thế giới
1818 H Street NW
Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

Báo cáo này là sản phẩm do chuyên gia của Ngân hàng Thế giới và các chuyên gia tư vấn thực hiện. Các kết quả, giải thích và kết luận đưa ra trong báo cáo này không phản ánh quan điểm chính thức của Ngân hàng Thế giới, Ban Giám đốc điều hành Ngân hàng Thế giới hoặc các Chính phủ mà họ đại diện.

Ngân hàng Thế giới không đảm bảo tính chính xác, đầy đủ hoặc tính thời sự của dữ liệu được sử dụng trong báo cáo và không chịu trách nhiệm về bất kỳ sai sót, thiếu sót hoặc tính thiếu nhất quán trong thông tin cũng như trách nhiệm pháp lý đối với việc sử dụng hoặc không sử dụng thông tin, phương pháp, quy trình hoặc kết luận đưa ra. Đường biên giới, màu sắc, tên gọi và các thông tin khác biểu hiện trên các bản đồ trong báo cáo này không hàm ý bất kỳ đánh giá nào của Ngân hàng Thế giới về vị thế pháp lý của bất kỳ vùng lãnh thổ nào và cũng không thể hiện bất kỳ sự ủng hộ hay chấp nhận nào của Ngân hàng Thế giới về các đường biên giới đó.

Không có nội dung nào trong báo cáo này cấu thành hoặc được hiểu hoặc được coi là hạn chế hay từ bỏ các đặc quyền và miễn trừ của Ngân hàng Thế giới, tất cả những điều trên được bảo lưu.

Ảnh bìa: Wright Studio@shutterstock

Kích hoạt giá trị của dữ liệu mở: Động lực mới cho đổi mới sáng tạo và kinh tế dữ liệu

Báo cáo đánh giá dữ liệu mở ở Việt Nam

Official Use Only

MỤC LỤC

LỜI CẢM ƠN iv

TÓM TẮT v

PHẦN I. GIỚI THIỆU CHUNG 1

1. Tổng quan về bối cảnh và mục tiêu nghiên cứu 1

2. Phương pháp 2

2.1. Phương pháp tổng quan, phân tích chính sách, quy định về dữ liệu mở 2

2.2. Phương pháp đánh giá cổng dữ liệu mở tại Việt Nam 3

2.3. Phương pháp khảo sát thực trạng cung cấp dữ liệu mở tại Việt Nam 4

2.4. Phương pháp khảo sát nhu cầu sử dụng dữ liệu mở tại Việt Nam 4

PHẦN II. MỘT SỐ KẾT QUẢ NGHIÊN CỨU CHÍNH 6

1. Các quy định, chính sách hiện hành về dữ liệu mở 6

1.1. Khái niệm và đặc điểm của dữ liệu mở 7

1.2. Nguyên tắc tiếp cận, sử dụng và cung cấp dữ liệu mở 10

1.3. Quy trình cung cấp dữ liệu mở của cơ quan nhà nước 12

1.4. Trách nhiệm phối hợp và triển khai thực hiện của các cơ quan, đơn vị 16

2. Nhu cầu khai thác, sử dụng dữ liệu mở 19

2.1. Mục đích sử dụng dữ liệu mở 19

2.2. Nhu cầu về định dạng dữ liệu mở 20

2.3. Nhu cầu đối với các tính năng của cổng dữ liệu mở 21

3. Hiện trạng cung cấp dữ liệu mở 22

3.1. Một số ưu điểm của các cổng dữ liệu mở 23

3.2. Một số hạn chế của các cổng dữ liệu mở 23

3.3. Đánh giá của người sử dụng về cổng dữ liệu mở tại Việt Nam 25

4. Một số khó khăn trong quá trình cung cấp dữ liệu mở của cơ quan nhà nước 27

4.1. Cơ sở hạ tầng, kĩ thuật và công nghệ thông tin 29

4.2. Kinh phí và nguồn lực 29

4.3. Nhận thức của các bên liên quan 30

PHẦN III. KẾT LUẬN VÀ ĐỀ XUẤT 33

1. Kết luận 33

2. Đề xuất 35

Phụ lục: Yêu cầu cơ bản về chức năng, tính năng kĩ thuật của cổng dữ liệu mở 39

TÀI LIỆU THAM KHẢO 50

Official Use Only

Bảng

Bảng 1: Mục đích sử dụng dữ liệu của các TCDN và người dùng cá nhân 20
Bảng 2: Nhu cầu về các tính năng trên cổng DLM (n=372) 21
Bảng 3: Những khó khăn của các CQNN trong quá trình cung cấp DLM (n=280) 29
Bảng A.1: Giải thích từ ngữ và chữ viết tắt 39
Bảng A.2: Chức năng cơ bản đối với người sử dụng 43
Bảng A.3: Chức năng cơ bản đối với người xuất bản 443
Bảng A.4: Chức năng cơ bản đối với quản trị viên 454
Bảng A.5: Các yêu cầu phi chức năng của Cổng dữ liệu mở 465

Hình

Hình 1: Nhu cầu về định dạng DLM của các TCDN và người dùng cá nhân (n=420) 21
Hình 2: Kết quả đánh giá cổng DLM cấp Quốc gia và Bộ (n=4) 25
Hình 3: Kết quả đánh giá cổng DLM cấp tỉnh (n=30) 25
Hình 4: Mức độ hài lòng về các cổng DLM mà người dùng cá nhân đã từng sử dụng (n=54) 26
Hình 5: Tỷ lệ người dùng cá nhân thường xuyên và rất thường xuyên gặp những khó khăn khi truy cập các cổng DLM (n=54) 26

Hộp

Hộp 1: Định nghĩa về dữ liệu mở của các nước trên thế giới 9
Hộp 2: Kĩ thuật ẩn danh dữ liệu nhằm bảo mật thông tin cá nhân, theo Quy định bảo vệ dữ liệu chung của châu Âu 12
Hộp 3: Kinh nghiệm quốc tế liên quan đến danh mục dữ liệu mở 14

Official Use Only

DANH MỤC TỪ VIẾT TẮT

API	Giao diện lập trình ứng dụng (Application Programming Interface)
CCCD	Căn cước công dân
CMND	Chứng minh nhân dân
CNTT	Công nghệ thông tin
CQNN	Cơ quan nhà nước
CSDL	Cơ sở dữ liệu
DEPOCEN	Trung tâm Nghiên cứu Chính sách và Phát triển
DLM	Dữ liệu mở
ĐH	Đại học
GDĐT	Giao dịch điện tử
GDS	Cơ quan Dịch vụ số (Government Digital Service)
PVS	Phỏng vấn sâu
TCDN	Tổ chức doanh nghiệp
TT&TT	Thông tin và Truyền thông

Official Use Only

LỜI CẢM ƠN

Báo cáo này được nhóm công tác Ngân hàng Thế giới và Trung tâm Nghiên cứu Chính sách và Phát triển (DEPOCEN) thực hiện theo yêu cầu của Cục Chuyển đổi số Quốc gia (CĐSQG) thuộc Bộ Khoa học và Công nghệ Việt Nam (Bộ KH&CN). Nhóm công tác Ngân hàng Thế giới do bà Trần Thị Lan Hương làm trưởng nhóm, gồm có ông Ngô Quốc Thái, ông Lưu Danh Hiếu, bà Đỗ Thanh Phương, ông Trần Lương Thành và bà Lê Thị Khánh Linh. Nhóm DEPOCEN do ông Nguyễn Ngọc Anh làm trưởng nhóm, gồm có bà Nguyễn Thị Phương Mai, bà Trần Thị Thanh Thuỷ, bà Trần Thị Ngọc Ánh, ông Lê Vũ Huy, ông Đoàn Mạnh Đức, bà Vũ Phương Mai, ông Vũ Đình Quân, và bà Nguyễn Hải Hồng.

Chúng tôi xin chân thành cảm ơn các ý kiến chỉ đạo của ông Nguyễn Thành Phúc (Cục trưởng, Cục CĐSQG), ông Hồ Đức Thắng (Phó Cục trưởng, Cục CĐSQG) cũng như tất cả các góp ý của ông Nguyễn Trọng Khánh, ông Mai Thanh Hải và bà Lê Thị Thùy Trang từ Cục CĐSQG trong quá trình thực hiện đánh giá. Nhóm cũng trân trọng cảm ơn những chia sẻ nhiệt tình của cán bộ, công chức từ các cơ quan trung ương tới địa phương của Việt Nam, cũng như các cá nhân đã tham gia khảo sát trực tuyến và phỏng vấn sâu. Sự đóng góp và tham gia của các bên liên quan này đã giúp nhóm đánh giá hiệu quả và đảm bảo các khuyến nghị dựa trên các cuộc tham vấn và phân tích sâu rộng về tình hình triển khai Dữ liệu mở hiện tại của Việt Nam. Nhóm cũng vô cùng biết ơn những góp ý của ông Nguyễn Anh Tuấn từ Trung tâm Dữ liệu quốc gia thuộc Bộ Công an.

Chúng tôi biết ơn sự chỉ đạo và hỗ trợ của bà Mariam J. Sherman (Giám đốc quốc gia Ngân hàng Thế giới tại Việt Nam, Campuchia và Lào), ông Mahesh Uttamchandani (Giám đốc khu vực, Ban Phát triển số), bà Kathleen Whimp (Giám đốc Quản lý hoạt động và Danh mục đầu tư), ông James Neumann (Quyền Quản lý, Ban Phát triển số) và bà Zayra Romo (Trưởng Ban Cơ sở hạ tầng – Phát triển số) trong quá trình hoàn thiện báo cáo.

Nhóm xin chân thành cảm ơn sự hỗ trợ tài chính từ Chương trình Đối tác chiến lược Ốtxtrâylia - Nhóm Ngân hàng Thế giới (ABP).

Official Use Only

TÓM TẮT

Ngân hàng Thế giới phối hợp với Cục Chuyển đổi số quốc gia thuộc Bộ Thông tin và Truyền thông (Bộ TT&TT)[1] và Trung tâm Nghiên cứu Chính sách và Phát triển (DEPOCEN) thực hiện báo cáo **Đánh giá dữ liệu mở "Kích hoạt giá trị của dữ liệu mở - Động lực mới cho đổi mới sáng tạo và kinh tế dữ liệu ở Việt Nam"** nhằm ba mục tiêu chính: 1) Phân tích các khoảng trống, các điểm còn hạn chế về chính sách, quy định hiện hành và kế hoạch triển khai của các Bộ, ngành, địa phương về cung cấp và phát triển dữ liệu mở (DLM); phân tích kế hoạch của các Bộ, ngành, địa phương về cung cấp DLM, xây dựng cổng DLM, 2) Đánh giá những kết quả và hạn chế của các cổng thông tin DLM dựa trên các quy định hiện hành, 3) Khuyến nghị nâng cao hiệu quả triển khai dữ liệu mở hướng đến nền kinh tế dữ liệu ở Việt Nam.Nghiên cứu bao gồm bốn nội dung:

(i) Tổng quan chính sách của Việt Nam về DLM và một số thực tiễn tốt trên thế giới;

(ii) Khảo sát về nhu cầu sử dụng DLM tại Việt Nam;

(iii) Khảo sát về thực trạng cung cấp DLM tại Việt Nam;

(iv) Đánh giá cổng DLM tại Việt Nam.

Thời gian thực hiện các khảo sát là từ ngày 01 tháng 10 năm 2024 đến ngày 29 tháng 12 năm 2024; thời gian đánh giá cổng DLM được thực hiện từ ngày 15 tháng 10 năm 2024 đến ngày 15 tháng 11 năm 2024. Mỗi một cấu phần sẽ tương ứng với một báo cáo, trình bày chi tiết về những phân tích, kết quả và khuyến nghị. Sau đó, nội dung tổng thể được tóm lược một cách cô đọng trong báo cáo này, nhấn mạnh vào những phát hiện chính của toàn bộ nghiên cứu.

Bên cạnh các khảo sát trong nước, nghiên cứu cũng đã tiến hành tổng quan tài liệu và nghiên cứu chuyên sâu về các thông lệ quốc tế trong triển khai DLM. Nhóm nghiên cứu đã rà soát kinh nghiệm điển hình từ các quốc gia như Pháp, Hàn Quốc, Hoa Kỳ và Úc, tập trung vào các tiêu chuẩn kĩ thuật, quy định pháp lý, mô hình quản trị và cơ chế vận hành cổng DLM. Qua đó, báo cáo phân tích đối chiếu chi tiết những điểm tương đồng và khác biệt giữa Việt Nam và các nước, làm cơ sở cho việc đề xuất khuyến nghị phù hợp với bối cảnh trong nước. Các thông tin này đã được tích hợp trong từng phần nội dung nghiên cứu, đặc biệt trong phần đánh giá chính sách và chức năng kĩ thuật của cổng DLM.

Báo cáo này trình bày kết quả nghiên cứu với các nội dung nhằm đạt được ba mục tiêu nêu trên, bao gồm kết quả đánh giá toàn diện về các chính sách, quy định hiện hành về cung cấp DLM, kết quả đánh giá việc triển khai cung cấp DLM của các cơ quan nhà nước (CQNN) trên thực tế, kết quả khảo sát nhu cầu khai thác, sử dụng DLM của các tổ chức, cá nhân, từ đó chỉ ra các khoảng trống, các điểm còn hạn chế và đưa ra các khuyến nghị nhằm cải thiện các quy định, chính sách hiện hành và thúc đẩy việc cung cấp và sử dụng DLM.

Báo cáo thực hiện tổng quan các quy định pháp lý liên quan đến DLM ở Việt Nam, bao gồm các văn bản quy phạm pháp luật, các văn bản pháp lý của nhà nước, các kế hoạch triển khai thực hiện các chính sách, quy định ở cấp địa phương và các văn bản khác. Các vấn đề được tổng quan, rà soát theo khung phân tích đi từ mục đích, ý nghĩa của việc cung cấp DLM, khái niệm, định nghĩa về DLM đến các quy định về triển khai cung cấp và quản lí DLM.

Đối với các Cổng DLM hiện hành, nghiên cứu đã thực hiện đánh giá từng cổng để có được bức tranh tổng quan hiện trạng các cổng DLM. Bộ công cụ đã được thiết kế để đánh giá cụ thể theo 27 tiêu chí, từ đó so sánh và chỉ ra điểm mạnh, điểm yếu của từng cổng DLM của tất các Bộ/ngành/địa phương hiện có. Tổng cộng có 37 cổng DLM đã được đánh giá, trong đó có 4 cổng cấp bộ, 1 cổng quốc gia, 32 cổng địa phương. Tuy nhiên, cuối cùng chỉ có 34 cổng truy cập được và có cung cấp các bộ dữ liệu để đánh giá nên báo cáo này

[1] Bộ TT&TT đã được hợp nhất với Bộ Khoa học và Công nghệ (KH&CN) theo Kế hoạch số 141/KH-BCDDTKNQ18 ngày 06/12/2024 của Ban Chỉ đạo về tổng kết việc thực hiện Nghị quyết số 18-NQ/TW và Nghị quyết số 176/2025/QH15 ngày 18/02/2025 của Quốc hội về cơ cấu tổ chức của Chính phủ nhiệm kỳ Quốc hội khóa XV. Tuy nhiên, trong phạm vi của báo cáo này và các tài liệu liên quan, tên gọi Bộ TT&TT vẫn được giữ nguyên.

Official Use Only

trình bày kết quả đánh giá chung đối với 34 cổng gồm 3 cổng cấp bộ và 1 cổng quốc gia và 30 cổng địa phương. Các cổng DLM đều đang trong quá trình xây dựng và phát triển do vậy các kết quả trình bày trong nghiên cứu này là tại thời điểm thực hiện đánh giá cổng DLM từ ngày 15/10/2024 đến ngày 15/11/2024.

Báo cáo nghiên cứu cũng đã thực hiện khảo sát hai nhóm đối tượng, gồm: (i) nhóm cung cấp DLM là cán bộ các bộ, cơ quan, đơn vị cấp Trung ương (TW) và các cơ quan, đơn vị cấp địa phương.; (ii) nhóm sử dụng DLM (người dùng) là tổ chức, doanh nghiệp (TCDN) và cá nhân sử dụng, khai thác DLM. Nhóm nghiên cứu đã gửi phiếu hỏi cung cấp DLM tới 438 cán bộ và thu về 280 phản hồi đại diện từ các CQNN (tỉ lệ phản hồi là 64%), Đối với nhóm sử dụng DLM, có 1200 phiếu được phát ra và có 420 người dùng tham gia khảo sát (tỷ lệ phản hồi đạt 35%), bao gồm 95 người dùng cá nhân và 325 TCDN (bao gồm các viện nghiên cứu/trường đại học, doanh nghiệp, tổ chức quốc tế/tổ chức phi chính phủ, cơ quan truyền thông, báo chí).

Ngoài ra, báo cáo nghiên cứu đã thực hiện phỏng vấn đại diện các bộ, cơ quan, địa phương về việc triển khai cung cấp DLM, xây dựng cổng DLM, về các quy định, chính sách hiện hành có liên quan. Một số TCDN, cá nhân sử dụng, khai thác DLM cũng được phỏng vấn sâu để tìm hiểu ý kiến, quan điểm, nhu cầu khai thác và sử dụng DLM. Nghiên cứu cũng đã nhận được các ý kiến đóng góp của chuyên gia Cục Chuyển đổi số quốc gia và Trung tâm dữ liệu quốc gia.

Một số kết quả chính của nghiên cứu

Về nhu cầu sử dụng, khai thác dữ liệu mở

(i) DLM có vai trò quan trọng trong việc thúc đẩy giao dịch điện tử (GDĐT), chuyển đổi số, phát triển kinh tế số và xã hội số. Các tổ chức và cá nhân (người dùng) cho thấy có nhu cầu cao trong việc sử dụng DLM phục vụ cho mục đích kinh doanh, sản xuất, phân tích thị trường, nghiên cứu sản phẩm mới, nghiên cứu khoa học...

(ii) Người dùng mong muốn DLM được cung cấp dưới các định dạng dễ sử dụng như Excel, CSV và API, nhằm tối ưu hóa khả năng khai thác và phân tích dữ liệu. Đối với các cổng DLM, người dùng mong muốn có các tính năng quan trọng như mô tả dữ liệu chi tiết, giao diện trực quan, dễ dàng sử dụng, có tính năng phản hồi dữ liệu, có liên kết với cổng DLM quốc gia, có công cụ tìm kiếm nâng cao giúp tìm kiếm và kết nối dữ liệu hiệu quả hơn.

Về hiện trạng cung cấp dữ liệu mở

(iii) Một số ưu điểm của các cổng DLM hiện hành ở Việt Nam là đã cho phép người dùng có thể khai thác dữ liệu mọi lúc, mọi nơi và bất kỳ ai đều có thể truy cập và tải về miễn phí, có khả năng truy cập ổn định trên các thiết bị điện tử; về giao diện, các cổng DLM được thiết kế tối giản, dễ nhìn và thân thiện với người dùng, tương thích với màn hình máy tính; đã tích hợp một số tính năng cơ bản như quản lý danh mục (tìm kiếm, lọc dữ liệu), quản lý và lưu trữ dữ liệu (tải xuống dữ liệu, truy cập API), chức năng cộng đồng (bình luận, đánh giá, chia sẻ trên mạng xã hội), quản lý tài khoản (cá nhân hóa tài khoản người dùng).

(iv) Cổng DLM vẫn còn những hạn chế, chưa đáp ứng nhu cầu của người dùng như: Chức năng, tính năng kĩ thuật trên các cổng DLM chưa đồng đều; phần lớn các cổng chưa khai thác hiệu quả tiềm năng của API; nhiều cổng chưa cho phép người dùng gửi yêu cầu dữ liệu, không cung cấp bất kỳ cơ chế phản hồi về bộ dữ liệu; hiện không có cổng nào cung cấp ví dụ tái sử dụng dữ liệu.

(v) Số lượng và chất lượng dữ liệu được cung cấp trên các cổng DLM còn ít và chưa đáp ứng theo kế hoạch. Nhiều cổng chủ yếu cung cấp dữ liệu dưới định dạng PDF, điều này làm giảm khả năng sử dụng của dữ liệu, vì định dạng này không thể dễ dàng đọc hoặc phân tích tự động. Nhiều cổng DLM hiện nay vẫn sử dụng dữ liệu cũ và không được cập nhật thường xuyên, dẫn đến việc chất lượng thông tin không còn chính xác và thiếu tính thời sự, không đáp ứng được nhu cầu sử dụng dữ liệu mới và chính xác để phục vụ cho nghiên cứu, ra quyết định hay các hoạt động phân tích mặc dù theo quy định. Các cổng DLM hiện nay đã bước đầu cung cấp các mô tả về dữ liệu. Tuy

Official Use Only

nhiên, các trường dữ liệu trong lớp Dataset chưa đáp ứng được yêu cầu của Tiêu chuẩn DCAT-VN 14171:2024 về Đặc tả tập DLM cho cổng dữ liệu.

(vi) Nhiều cổng DLM chưa đảm bảo tính bảo mật, làm lộ lọt dữ liệu cá nhân: Điều 20 Nghị định 47/2020/NĐ-CP yêu cầu không tiết lộ thông tin cá nhân khi công bố DLM của CQNN. Việc lộ lọt thông tin cá nhân gây nguy cơ rủi ro bảo mật và vi phạm quyền riêng tư của công dân. Tuy nhiên, theo rà soát, có 18/34 cổng DLM hiện cung cấp dữ liệu chưa đảm bảo tính bảo mật, làm lộ dữ liệu cá nhân.

Về các chính sách, quy định hiện hành liên quan đến dữ liệu mở

(vii) Tổng cộng chín mươi bốn (94) văn bản pháp luật về dữ liệu mở đã được nghiên cứu, rà soát. Trong đó, bao gồm tám (8) luật, năm (5) nghị định, bốn (4) thông tư, nhiều văn bản quyết định và hướng dẫn do cả chính quyền trung ương và địa phương ban hành.

(viii) Trong các văn bản pháp luật hiện nay có một số khái niệm liên quan đến dữ liệu và DLM như "thông tin", "dữ liệu", "dữ liệu số", "dữ liệu mở", "dữ liệu mở của cơ quan nhà nước", "dữ liệu dùng chung", "dữ liệu dùng riêng", "dữ liệu gốc"… DLM được hiểu là một loại thông tin của CQNN (theo Luật tiếp cận thông tin). Theo quy định, thông tin của CQNN phải được người có thẩm quyền của CQNN ký, đóng dấu hoặc xác nhận bằng văn bản. Điều này lý giải cho hiện trạng trên cổng DLM của nhiều địa phương có nhiều tệp dữ liệu định dạng PDF được chụp từ các văn bản của CQNN với chữ ký, con dấu của người có thẩm quyền.

(ix) Hiện nay có ba văn bản quy phạm pháp luật đưa ra định nghĩa về DLM (Nghị định 47/2020/NĐ-CP của Chính phủ năm 2020, Luật GDĐT năm 2023 và Luật Dữ liệu năm 2024), các định nghĩa không hoàn toàn thống nhất (và chỉ áp dụng trong phạm vi văn bản đó) nhưng có điểm chung là: (i) DLM là "dữ liệu số" (dữ liệu "được biểu diễn bằng tín hiệu số" hay "dùng tín hiệu số" hay "được thể hiện dưới dạng kĩ thuật số") và (ii) mọi cơ quan, tổ chức, cá nhân được tiếp cận, sử dụng miễn phí. Một số đặc điểm quan trọng của DLM về định dạng máy đọc được hay không chứa thông tin cá nhân hiện chưa được nhắc đến khi giải thích thuật ngữ và chưa thể chế hóa các quy định cụ thể hoặc có hướng dẫn để đảm bảo các yêu cầu này được thực hiện.

(x) Quy định hiện nay về cổng dữ liệu và cổng DLM còn rải rác trong nhiều văn bản (Nghị định 47/2020/NĐ-CP, Nghị định 42/2022/NĐ-CP, Tiêu chuẩn TCVN 14171:2024) mà chưa có hướng dẫn cụ thể về tiêu chuẩn kĩ thuật để xây dựng cổng DLM cấp bộ, cấp tỉnh, dẫn đến tình trạng mỗi cổng thông tin của các địa phương được thiết kế và triển khai theo những cách thức khác nhau, gây khó khăn trong việc tạo dựng hệ thống đồng nhất trên phạm vi quốc gia.

(xi) Một số hướng dẫn về quy chuẩn, tiêu chuẩn, quy định kĩ thuật kết nối, chia sẻ DLM hiện được thực hiện thông qua văn bản hành chính (các công văn của Bộ TT&TT), không phải là quy phạm pháp luật do đó tính tuân thủ và hiệu quả thực thi không cao. Mặt khác, nhiều hướng dẫn đối với việc kết nối, chia sẻ dữ liệu nói chung, không dành riêng cho DLM dẫn đến tình trạng DLM được các bộ, địa phương cung cấp hiện nay dưới dạng máy không đọc được hay có chứa thông tin cá nhân tương đối phổ biến.

(xii) Quy trình các bước chuẩn bị dữ liệu cho đến kết nối, chia sẻ DLM chưa được quy định chi tiết trách nhiệm của các bên liên quan, cụ thể là giữa cơ quan chủ quản dữ liệu và cơ quan cung cấp dữ liệu.

(xiii) Chưa có quy định về việc kiểm tra, đánh giá, duy trì DLM. Quy định hiện hành áp dụng đối với trường hợp dữ liệu do CQNN chia sẻ nói chung, không phù hợp với DLM. Theo quy định hiện hành, chủ quản CSDL có trách nhiệm tự tổ chức kiểm tra đánh giá định kỳ dữ liệu và gửi báo cáo về đơn vị chuyên trách về công nghệ thông tin (CNTT), còn các đơn vị chuyên trách về CNTT có trách nhiệm hướng dẫn chủ quản CSDL tự kiểm tra, đánh giá dữ liệu. Như vậy, trong trường hợp DLM, khi cơ quan chủ quản CSDL cung cấp DLM cho đơn vị chuyên trách về CNTT để chia sẻ cho tổ chức, cá nhân sử dụng trên cổng DLM thì việc cơ quan chủ quản CSDL có trách nhiệm kiểm

tra, đánh giá định kỳ theo quy định là không phù hợp, đòi hỏi cần có quy định riêng về trách nhiệm kiểm tra, đánh giá việc cung cấp, chia sẻ DLM của các cơ quan, đơn vị liên quan.

(xiv) Mặc dù có gần một tram quy định về dữ liệu mở, nhưng chưa có quy định về trình tự xử lý vướng mắc trong cung cấp, khai thác và sử dụng DLM.

Một số khó khăn trong triển khai thực hiện cung cấp dữ liệu mở

Ngoài việc hệ thống quy định pháp luật về DLM chưa đầy đủ khiến việc triển khai thực hiện của các CQNN ở trung ương và địa phương gặp nhiều khó khăn, thách thức, các CQNN còn gặp một số khó khăn, thách thức, như khảo sát cho thấy, cụ thể như sau:

(xv) Cơ sở hạ tầng CNTT nhằm đáp ứng các yêu cầu phức tạp trong kĩ thuật xử lý dữ liệu đa miền và hạ tầng công nghệ chưa được đầu tư đúng mức. Việc tích hợp, xử lý, và chuẩn hóa các loại dữ liệu khác nhau để có thể khai thác hiệu quả đặt ra một thách thức lớn cho các đơn vị thiếu cơ sở hạ tầng CNTT. Trên thực tế, hệ thống lưu trữ và tính toán hiện nay chưa đáp ứng được nhu cầu cung cấp DLM của cả quốc gia.

(xvi) Hiện nay, các nguồn kinh phí cho hoạt động về DLM toàn bộ được lấy từ ngân sách địa phương vì vậy còn rất hạn chế và có sự khác biệt giữa các địa phương khác nhau, nguồn xã hội hóa cho DLM chưa được thúc đẩy. Quy trình chi tiêu tài chính vẫn chưa rõ ràng, chính sách mua sắm hiện tại còn nhiều bất cập hạn chế khả năng cung cấp các sản phẩm, dịch vụ có chất lượng tốt hơn nếu áp dụng các tiêu chuẩn về mua sắm hiện hành.

(xvii) Thiếu nhân lực chuyên môn, đặc biệt là các chuyên gia kĩ thuật và chuyên gia về dữ liệu như kĩ sư dữ liệu (data engineer). Đây là những vị trí then chốt trong việc xây dựng và vận hành cổng DLM, nhưng nguồn nhân lực hiện tại chưa đáp ứng được yêu cầu cả về số lượng lẫn chất lượng.

(xviii) Nhận thức về DLM còn hạn chế là một thách thức làm chậm tiến trình phát triển DLM tại Việt Nam. Phần lớn các CQNN, tổ chức doanh nghiệp (TCDN) và người dùng cá nhân đã biết về các lợi ích của DLM nhưng vẫn hiểu sai về DLM, gồm khái niệm, định dạng, khả năng truy cập và cách thức cung cấp.

Đề xuất, khuyến nghị

DLM được xác định theo tính chất chia sẻ, trong đó quan trọng nhất là tất cả mọi cơ quan, tổ chức, cá nhân có nhu cầu đều được tiếp cận, chia sẻ, khai thác, sử dụng miễn phí. DLM không chỉ do CQNN mà còn có thể do các tổ chức khác cung cấp, cần khuyến khích các cơ quan, tổ chức tạo lập dữ liệu, chủ sở hữu dữ liệu chia sẻ DLM nhằm phát triển kinh tế số và xã hội số. Trong đó, để thực hiện Luật tiếp cận thông tin, Luật Giao dịch điện tử (GDĐT), Luật Dữ liệu, các CQNN phải cung cấp DLM, theo đó, cần xây dựng và ban hành một văn bản dưới luật riêng cho việc cung cấp DLM của CQNN dưới hình thức Nghị định, khắc phục các vấn đề như đã được nhận diện và kết luận ở trên. Ngoài ra, nhóm nghiên cứu cũng đề xuất một số bước hành động cụ thể như sau:

Bước 1: Cải thiện khung pháp lý và chính sách liên quan đến DLM

- Các quy định cần phân biệt rõ các loại dữ liệu chia sẻ giữa các CQNN, dữ liệu chia sẻ trong nội bộ một CQNN và DLM, dựa trên quyền truy cập của người dùng và các đặc điểm chia sẻ, kèm theo các tiêu chuẩn kĩ thuật và nội dung tương ứng.
- Cần quy định định dạng chuẩn, máy đọc được và đặc tả dữ liệu cho DLM (ví dụ như XML, CSV và JSON).
- Yêu cầu thiết lập cơ cấu quản trị dữ liệu ở cấp địa phương, ví dụ như nhóm công tác DLM do lãnh đạo đứng đầu, bao gồm các cán bộ lãnh đạo, cán bộ phụ trách CNTT và cán bộ quản lý dữ liệu. Nhóm này có nhiệm vụ báo cáo cho lãnh đạo cấp cao về lợi ích của DLM và việc chia sẻ dữ liệu công nhằm thúc đẩy sự ủng hộ từ trên xuống.

- Thiết lập cơ chế tài chính rõ ràng: Thực hiện một cơ chế tài chính minh bạch và cụ thể ở cấp trung ương nhằm giảm khoảng cách giữa các tỉnh. Trong đó, cần quy định mức phân bổ ngân sách tối thiểu cho các sáng kiến về DLM ở cấp địa phương, đặc biệt là tại các tỉnh khó khăn, nhằm đảm bảo có đủ nguồn lực thực hiện hiệu quả các nhiệm vụ liên quan đến DLM và thúc đẩy sự phát triển cân bằng.
- Quy trình cung cấp DLM cần được quy định rõ ràng, với việc phân công vai trò và trách nhiệm cụ thể cho từng bên liên quan. Quy trình này cần nêu rõ nghĩa vụ của các đơn vị sở hữu dữ liệu trong việc đảm bảo tính chính xác, quyền riêng tư và tuân thủ quy định pháp luật. Đồng thời, cần xác định nhiệm vụ giám sát của các cơ quan quản lý trong quá trình xử lý dữ liệu, tiếp nhận phản hồi từ người dùng và kiểm tra, giám sát. Cũng cần ban hành hướng dẫn cụ thể cho người dùng và cơ chế xử lý khi có vấn đề phát sinh. Cần thiết lập hệ thống giám sát, kiểm tra và báo cáo nghiêm ngặt để duy trì chất lượng DLM. Cần thành lập đội ngũ chuyên trách để thực hiện đánh giá định kỳ và tổ chức thu thập ý kiến nhằm thúc đẩy cải tiến liên tục.

Bước 2: Kiểm kê và các bộ dữ liệu cần ưu tiên

- Cần ban hành danh mục DLM thống nhất của cả nước để đảm bảo tính nhất quán của DLM và khả năng kết nối giữa các hệ thống. Danh mục này phải quy định rõ ràng về cấu trúc dữ liệu, các trường thông tin và tiêu chuẩn định dạng tối thiểu nhằm đảm bảo khả năng kết nối, tính thống nhất và khả năng sử dụng của dữ liệu.
- Danh mục DLM nên được xây dựng, áp dụng và triển khai theo hồ sơ ứng dụng DCAT cho tất cả các cổng DLM tại Việt Nam, đồng thời tận dụng định dạng dữ liệu liên kết mở RDF.
- Như báo cáo đã chỉ ra, nhiều địa phương hiện đã xây dựng danh mục DLM, tuy nhiên điều này là chưa đủ mà chỉ nên được coi là điểm khởi đầu. Cần rà soát lại danh mục để xác định những tập dữ liệu có thể công khai ngay (dữ liệu không nhạy cảm, được công chúng quan tâm) và những tập dữ liệu chỉ chia sẻ nội bộ (có tính nhạy cảm nhưng cần thiết cho việc phối hợp giữa các cơ quan).
- Tập trung cung cấp các bộ dữ liệu đã được công khai dưới một hình thức nào đó (ví dụ báo cáo trên website) nhưng chưa có định dạng DLM - các dữ liệu này có thể nhanh chóng được chuyển đổi sang định dạng CSV/Excel và công bố.
- Ưu tiên các bộ dữ liệu phù hợp với danh mục DLM quốc gia (nếu đã được ban hành) hoặc những bộ dữ liệu được yêu cầu thường xuyên, chẳng hạn như dữ liệu về thủ tục hành chính, thông tin địa chính,... Danh mục DLM nên đề xuất các chủ đề để phân loại những bộ dữ liệu theo các lĩnh vực, xây dựng lịch trình công bố dữ liệu, ví dụ: đặt mục tiêu “Mỗi quý công bố X tập dữ liệu mới”. Bắt đầu công bố với số lượng vừa phải, có thể từ 10–20 bộ dữ liệu trọng điểm trong giai đoạn đầu để tạo đà triển khai hiệu quả.

Bước 3: Phát triển cổng DLM (hoặc tích hợp với hệ thống hiện có) kết nối với trung tâm dữ liệu và/hoặc hệ thống trao đổi dữ liệu

- Đối với các địa phương chưa có cổng DLM, cần lựa chọn nền tảng triển khai và bắt buộc sử dụng phần mềm mã nguồn mở như CKAN hoặc DKAN. Cần xây dựng lộ trình kĩ thuật theo giai đoạn để phát triển cổng theo khung tiêu chuẩn.
- Đảm bảo cổng DLM có các tính năng thiết yếu: tìm kiếm và duyệt các bộ dữ liệu, tải về ở nhiều định dạng, truy cập API cơ bản cho người dùng nâng cao, hiển thị dữ liệu đặc tả và biểu mẫu liên hệ/phản hồi.
- Quy định các tiêu chuẩn kĩ thuật cho cổng DLM ở các cấp, đảm bảo tính thống nhất và khả năng tương tác. Đồng thời, cần phù hợp với tiêu chuẩn kĩ thuật quốc gia để phục vụ việc tích hợp sau này.
- Đối với những địa phương đã có cổng nhưng hoạt động kém hiệu quả (ví dụ: tốc độ chậm, thiếu tính năng), cần có kế hoạch nâng cấp như: cải thiện giao diện người dùng, bổ sung điểm truy cập API và khắc phục các liên kết không hoạt động. Trước khi triển khai chính thức nên thử nghiệm với một nhóm người dùng nhỏ (ví dụ: sinh viên địa phương) để đảm bảo khả năng sử dụng của cổng.

- Với các cơ sở dữ liệu dùng nội bộ trong một CQNN, nếu địa phương đã có hoặc đang xây dựng kho dữ liệu tập trung (như trung tâm tích hợp dữ liệu tỉnh), cần kết nối với cổng DLM để tự động công bố những dữ liệu không nhạy cảm.
- Đề xuất thành lập trung tâm dữ liệu hoặc hệ thống trao đổi dữ liệu ở cấp tỉnh để các sở ban ngành có thể tải dữ liệu lên trước khi đưa ra cổng DLM. Một số địa phương đã có quy trình tương tự, ví dụ Đà Nẵng và Thành phố Hồ Chí Minh, trong đó dữ liệu được đưa vào kho dữ liệu dùng chung trước khi công bố.

Bước 4: Triển khai quy trình quản lý dữ liệu, cụ thể là khi cổng được đưa vào hoạt động, thiết lập quy trình để duy trì vận hành

- *Định kỳ cập nhật dữ liệu*: Phối hợp với các đơn vị để cập nhật định kỳ theo từng loại dữ liệu (hằng tháng, quý, năm). Tự động hóa cập nhật khi có thể (qua API hoặc kịch bản xuất dữ liệu). Với trường hợp cập nhật thủ công, cần gửi thông báo đôn đốc các đơn vị và theo dõi việc thực hiện.
- *Kiểm soát chất lượng*: Trước khi công bố, đơn vị phụ trách công khai DLM cần kiểm tra định dạng, tính đầy đủ và tính bảo mật của dữ liệu. Áp dụng các kĩ thuật ẩn danh cơ bản: loại bỏ tên cá nhân, tổng hợp dữ liệu cá nhân nếu cần thiết (công bố các kết quả thống kê thay vì dữ liệu cá nhân gốc). Đảm bảo hiển thị đầy đủ dữ liệu đặc tả (ngày, nguồn, liên hệ…).
- *Phản hồi người dùng*: Theo dõi các kênh phản hồi. Khi người dùng yêu cầu một bộ dữ liệu hoặc báo lỗi, phải phản hồi kịp thời. Ngay cả khi chưa thể cung cấp dữ liệu ngay lập tức, vẫn cần ghi nhận và lưu lại các yêu cầu của người dùng, từ đó xây dựng lòng tin từ người dùng. Có thể thiết lập một địa chỉ email riêng hoặc một địa chỉ trang Facebook để công chúng thảo luận về DLM - việc tiếp cận người dùng sẽ thu được nhiều ý kiến hữu ích để cải thiện hệ thống.
- *Kết nối với cổng dữ liệu quốc gia*: Thường xuyên đồng bộ với cổng quốc gia (data.gov.vn). Tận dụng tính năng tự động tải dữ liệu địa phương lên cổng quốc gia (nếu có). Các cơ quan/địa phương cần phân công cán bộ phụ trách đảm bảo dữ liệu của mình được hiển thị/trích dẫn trên cổng quốc gia để tăng khả năng tiếp cận.

Bước 5: Tăng cường sử dụng cơ sở dữ liệu dùng chung trong nội bộ

- Cải thiện việc chia sẻ dữ liệu nội bộ song song với việc công bố DLM.
- Tiếp cận từng bước: Bắt đầu kết nối một số CSDL chính của các đơn vị (sở, ban ngành) với trung tâm dữ liệu/hệ thống trao đổi dữ liệu. Sau đó kết nối các hệ thống CSDL trung tâm đến trung tâm dữ liệu/hệ thống chia sẻ dữ liệu của tỉnh/thành phố.
- Tập trung vào kết quả cụ thể: nếu việc chia sẻ dữ liệu liên sở, ban, ngành giúp giải quyết một vấn đề cụ thể (chẳng hạn rút ngắn thời gian xử lý một thủ tục xuống vài ngày) cần ghi nhận và báo cáo kết quả này, làm căn cứ để các đơn vị nhận thức được về giá trị của việc giải quyết các vấn đề an toàn kĩ thuật bằng cách phối hợp với bên cung cấp giải pháp hoặc chuyên gia an ninh mạng để đảm bảo rằng việc kết nối dữ liệu có thể thực hiện một cách an toàn. Từng bước xây dựng một môi trường CSDL liên thông theo mô hình phân quyền, trong đó - tùy theo phân quyền - các cơ quan có thể truy xuất thông tin cần thiết từ nhau. Cách làm này cũng sẽ giúp DLM trở nên phong phú và cập nhật hơn (vì dữ liệu trong các hệ thống CSDL dùng chung cũng có thể cung cấp trực tiếp cho các bộ DLM).

Bước 6: Nâng cao năng lực cho cán bộ

- Tổ chức tập huấn cho đội ngũ cán bộ và các đầu mối phụ trách tại các sở, ngành với hỗ trợ chuyên môn từ các bộ ngành trung ương hoặc các dự án tài trợ.
- Nội dung đào tạo có thể bao gồm: cách sử dụng phần mềm cổng DLM, cách chuẩn bị bộ dữ liệu (làm sạch và định dạng dữ liệu), những kiến thức cơ bản về bảo mật và quyền riêng tư (để biết phần nào cần được ẩn danh), cũng như cách sử dụng các công cụ phân tích dữ liệu (giúp cán bộ chủ động khai thác DLM trong công việc).

Official Use Only

- Khuyến khích đội ngũ cán bộ CNTT tham gia các diễn đàn cấp quốc gia hoặc cộng đồng trực tuyến về DLM. Bố trí kinh phí cử cán bộ tham dự các hội thảo chuyên đề hoặc chương trình học hỏi quốc tế.
- Về lâu dài, nên xây dựng một nhóm "đại sứ dữ liệu" (data champions) trong nội bộ - có thể là các cán bộ trẻ, nhiệt huyết - để lan tỏa lợi ích của DLM đến đồng nghiệp, khuyến khích học hỏi giữa các đơn vị. Ví dụ, Cục Thống kê, Chi cục thống kê có kĩ năng dữ liệu tốt nên có thể huy động họ hỗ trợ các cơ quan khác trong quản lý và sử dụng dữ liệu.

Bước 7: Tăng cường gắn kết cộng đồng và đối tác

Để khai thác hiệu quả DLM, cần kết nối với hệ sinh thái ngoài nhà nước. Thông báo rộng rãi việc ra mắt cổng DLM qua truyền thông địa phương, mời các trường đại học, doanh nghiệp, tổ chức phi chính phủ cùng tham gia sử dụng và phản hồi.

Bước 8: Giám sát, đánh giá và điều chỉnh

- Thiết lập quy trình theo dõi tiến độ và điều chỉnh phù hợp. Theo dõi các chỉ số như: số lượng bộ dữ liệu đang được công bố (và cập nhật), lượng truy cập cổng dữ liệu, số lượt tải xuống hoặc sử dụng API cũng như phản hồi nhận được từ người dùng. Đối với những bộ dữ liệu ít được truy cập, cần tìm hiểu nguyên nhân (có thể do không thực sự hữu ích hoặc chưa được quảng bá tốt). Nếu có bộ dữ liệu được yêu cầu nhưng chưa cung cấp, cần lên kế hoạch để công bố.
- Báo cáo định kỳ cho lãnh đạo - ví dụ, có thể gửi báo cáo 6 tháng hoặc 1 năm cho Lãnh đạo bộ/ngành, UBND tỉnh/thành phố về tình hình cung cấp DLM (thể hiện sự tăng trưởng về số lượng hoặc mức độ sử dụng) nhằm duy trì sự quan tâm và ủng hộ từ cấp trên. So sánh với các địa phương khác (nếu có số liệu hoặc bảng xếp hạng), việc so sánh, đối chiếu giữa các địa phương ở mức độ có chừng mực có thể là động lực để các địa phương nỗ lực hơn.
- Cần phải xem đây là một quá trình cải tiến liên tục: thu thập ý kiến người dùng và rút kinh nghiệm. Nếu người dùng phản ánh chức năng tìm kiếm khó sử dụng, cần phải cải tiến. Bộ, ngành trung ương cần ban hành hướng dẫn mới (về chuẩn kĩ thuật, danh mục, các quy định chi tiết triển khai dữ liệu mở của Luật Dữ liệu v.v.), cụ thể, chi tiết và có cơ chế hướng dẫn, phân bổ kinh phí để hỗ trợ các địa phương cập nhật và áp dụng đồng bộ.

Bước 9: Bảo đảm an toàn và bảo mật dữ liệu xuyên suốt: Đây là bước xuyên suốt toàn bộ quá trình - luôn đảm bảo không có dữ liệu nhạy cảm nào vô tình bị công bố và các hệ thống dữ liệu dùng chung luôn được bảo vệ an toàn.

- Cập nhật quy trình ngay khi có hướng dẫn mới về quyền riêng tư dữ liệu.
- Trước khi mở dữ liệu, cần xác minh và ẩn danh dữ liệu cá nhân, chẳng hạn như số CMND/CCCD hoặc tên trong tập dữ liệu.
- Kiểm tra định kỳ các tập dữ liệu để đảm bảo không có thông tin cá nhân bị lọt ra (nhất là sau những lần cập nhật bổ sung thông tin mới).
- Xây dựng quy trình phản ứng sự cố: nếu có ai đó phát hiện vấn đề về quyền riêng tư hoặc lỗ hổng bảo mật, cần xử lý nhanh chóng – tạm thời gỡ bộ dữ liệu, khắc phục sự cố rồi mới công bố lại.
- Việc chủ động bảo vệ quyền riêng tư không chỉ giúp tăng niềm tin của người dân, mà còn tạo sự yên tâm cho các CQNN khi tham gia chia sẻ dữ liệu qua một quy trình DLM có biện pháp bảo vệ chắc chắn.

PHẦN I. GIỚI THIỆU CHUNG

1. Tổng quan về bối cảnh và mục tiêu nghiên cứu

Trên thế giới, ngày càng nhiều chính quyền xem dữ liệu và dữ liệu mở (DLM) là công cụ hữu hiệu để thúc đẩy minh bạch hoá, khơi dậy đổi mới sáng tạo và tạo động lực cho phát triển kinh tế - xã hội. Thực tiễn tại các quốc gia như Pháp, Mexico, Hàn Quốc hay Úc cho thấy khi các chính phủ ban hành chính sách đồng bộ, quy định kĩ thuật rõ ràng, việc phát huy giá trị của dữ liệu và triển khai dữ liệu mở diễn ra thuận lợi và mang lại giá trị rõ rệt. Chính phủ Anh, thông qua Cơ quan Dịch vụ số (GDS), đã lồng ghép nguyên tắc DLM vào chiến lược chuyển đổi số quốc gia một cách bài bản, nhờ vào các chính sách thống nhất và phân định rõ trách nhiệm giữa các cơ quan. Ở Hàn Quốc, thành công trong triển khai DLM phần lớn đến từ việc ban hành chỉ đạo tập trung từ trung ương, giúp liên thông hiệu quả dữ liệu giữa các cấp chính quyền. Bài học từ các quốc gia như Anh hay Hàn Quốc càng củng cố thêm nhận thức về tầm quan trọng của một khung chính sách rõ ràng và tập trung.

Việt Nam cũng không đứng ngoài xu hướng này. Nghị định số 47/2020/NĐ-CP ngày 09/4/2020 về quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước (sau đây gọi tắt là Nghị định số 47/2020/NĐ-CP) ra đời như một bước đi đầu tiên xác lập vai trò của dữ liệu và DLM. Nghị định đã quy định các nguyên tắc và quy trình quản lý DLM của các cơ quan nhà nước. Nghị quyết số 57-NQ/TW ngày 22/12/2024 của Bộ Chính trị nhằm tạo ra đột phá trong phát triển khoa học công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia đã tái khẳng định dữ liệu là tài nguyên và tài nguyên này phải đáp ứng ba yêu cầu: chất lượng cao, có khả năng chia sẻ và bảo đảm an toàn, bảo mật.

Việt Nam triển khai nhiều sáng kiến về dữ liệu trong những năm gần đây, trong đó bao gồm việc triển khai Nghị định 47/2020/NĐ-CP và đưa DLM vào chương trình hành động của các bộ, ngành, địa phương. Cụ thể, khoản 4 Điều 3 của Nghị định số 47/2020/NĐ-CP đã đưa ra định nghĩa về "dữ liệu mở của cơ quan nhà nước". Theo đó, DLM của cơ quan nhà nước (CQNN) là "*dữ liệu được cơ quan nhà nước có thẩm quyền công bố rộng rãi cho cơ quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ*". Khoản 1, Điều 13 Nghị định số 47/2020/NĐ-CP nêu cổng dữ liệu quốc gia là đầu mối cung cấp DLM của các CQNN trên môi trường mạng nhằm tăng cường tính minh bạch trong hoạt động của Chính phủ và thúc đẩy sáng tạo, phát triển kinh tế, xã hội. Ngoài các quy định tại Nghị định số 47/2020/NĐ-CP, việc cung cấp DLM cũng tuân thủ các nguyên tắc quy định tại Chương I, II của Luật Tiếp cận thông tin năm 2016, Điều 40 và 43 của Luật Giao dịch điện tử (GDĐT) năm 2023 (thay thế Luật GDĐT số 51/2005/QH11 năm 2005) và các quy định pháp luật khác có liên quan *(xem thêm Báo cáo Tổng quan chính sách của Việt Nam về DLM*). Theo đó, các CQNN cần công bố DLM thông qua các cổng DLM chuyên dụng.

Theo quy định, các bộ, cơ quan, địa phương có trách nhiệm ban hành danh mục DLM của CQNN, tổ chức xây dựng kế hoạch và triển khai công bố DLM của CQNN trong phạm vi quản lý của mình (Điều 17, Nghị định 47/2020/NĐ-CP). Tại thời điểm thực hiện nghiên cứu, có **15/18** Bộ đã ban hành kế hoạch hoặc danh mục DLM, có **61/63** địa phương ban hành kế hoạch hoặc danh mục DLM của tỉnh/thành phố[2], trong đó có 33 tỉnh/thành phố đã triển khai xây dựng, thử nghiệm cổng dữ liệu chuyên dụng theo quy định tại Nghị định số 42/2022/NĐ-CP ngày 24/6/2022 của Chính phủ về việc cung cấp thông tin và dịch vụ công trực tuyến của CQNN trên môi trường mạng, 30 tỉnh/thành đã bắt đầu cung cấp các bộ dữ liệu trên cổng và mỗi cổng cung cấp trung bình 181 bộ dữ liệu.

Việc triển khai xây dựng cổng DLM trên thực tế của các Bộ, ngành, địa phương còn gặp nhiều khó khăn, thách thức. Cụ thể, giao diện và chức năng của các cổng DLM của Bộ, ngành, địa phương vẫn chưa được chuẩn hóa, một số cổng vẫn đang trong giai đoạn thử nghiệm và cần nhiều cải tiến, các cổng hình thành còn rời rạc, không thể chia sẻ dữ liệu với nhau[3]. Số lượng các bộ DLM ở Việt Nam còn ít, theo báo cáo của Bộ Nội

[2] Tính đến ngày 28/11/2024, công cụ tìm kiếm Internet.

[3] Bộ TT&TT và Cục chuyển đổi số Quốc gia, Thuyết minh Dự thảo tiêu chuẩn Quốc gia: Đặc tả DLM cho Cổng dữ liệu DCAT-VN. 2023.

Official Use Only

vụ và Viện Khoa học tổ chức nhà nước, tính đến tháng 9/2022, cổng DLM quốc gia có 10.616 bộ dữ liệu, bao gồm nhiều chủ đề khác nhau, tuy nhiên, về mặt định dạng, khoảng 90% bộ dữ liệu được cung cấp dưới dạng khó khai thác (định dạng PDF). Các tập dữ liệu ở định dạng thuận tiện cho việc sử dụng và khai thác chỉ chiếm chưa tới 9% số tập dữ liệu[4]. Tới nay, trong các đơn vị thuộc Bộ, ngành, địa phương thực hiện mở dữ liệu, vẫn tồn tại những bất cập về cách hiểu, triển khai và trình bày DLM (dạng bộ dữ liệu thô, dạng API hoặc dạng thông tin tra cứu trực tiếp).

Bối cảnh trên đặt ra yêu cầu nghiên cứu các quy định, chính sách và kế hoạch của các Bộ, ngành, địa phương trong việc triển khai cung cấp DLM; nghiên cứu kinh nghiệm, thực tiễn tốt, tiêu chuẩn thế giới về cổng DLM; đánh giá nhu cầu sử dụng DLM; đánh giá hiện trạng các cổng DLM của Bộ, ngành, địa phương từ đó đưa ra các đề xuất, khuyến nghị xây dựng bộ tiêu chí về cổng DLM cho các Bộ, ngành, địa phương nhằm nâng cao chất lượng, hiệu quả, đẩy nhanh tiến độ của hoạt động chuyển đổi số quốc gia, góp phần nâng cao thứ hạng chỉ số DLM (Open Government Data Index-OGDI) của Việt Nam trong đánh giá của Liên hợp quốc, và quan trọng hơn là đưa đất nước vào con đường vững chắc hướng tới một nền kinh tế mới sáng tạo và dựa trên dữ liệu.Nghiên cứu **Đánh giá dữ liệu mở - Kích hoạt giá trị của dữ liệu mở - Động lực mới cho đổi mới sáng tạo và kinh tế dữ liệu ở Việt Nam** được thực hiện với ba mục tiêu chính như sau:

1. Phân tích các khoảng trống, các điểm còn hạn chế về chính sách, quy định hiện hành về cung cấp DLM, dựa trên thực tiễn tốt trên thế giới; phân tích kế hoạch của các Bộ, ngành, địa phương về cung cấp DLM, xây dựng cổng DLM.
2. Đánh giá những kết quả và hạn chế của các cổng thông tin DLM dựa trên các quy định hiện hành (đánh giá cổng DLM).
3. Khuyến nghị nâng cao hiệu quả triển khai Dữ liệu mở hướng đến nền kinh tế dữ liệu ở Việt Nam.

Báo cáo này trình bày kết quả nghiên cứu với các nội dung nhằm đạt được ba mục tiêu nêu trên, bao gồm kết quả đánh giá toàn diện về các chính sách, quy định hiện hành về cung cấp DLM, cung cấp kinh nghiệm thực tiễn tốt trên thế giới về DLM, kết quả đánh giá việc triển khai cung cấp DLM của các CQNN trên thực tế, kết quả khảo sát nhu cầu khai thác, sử dụng DLM của các tổ chức, cá nhân, từ đó chỉ ra các khoảng trống, các điểm còn hạn chế và đưa ra các khuyến nghị nhằm cải thiện các quy định, chính sách hiện hành và thúc đẩy việc cung cấp và sử dụng DLM.

2. Phương pháp

Để đáp ứng mục tiêu đề ra, nghiên cứu đã được thực hiện với 4 nội dung chính:

(i) Tổng quan chính sách của Việt Nam về DLM và rà soát thực tiễn tốt trên thế giới;

(ii) Khảo sát về nhu cầu sử dụng DLM tại Việt Nam;

(iii) Khảo sát về thực trạng cung cấp DLM tại Việt Nam;

(iv) Đánh giá cổng DLM tại Việt Nam.

Với mỗi nội dung, nhóm nghiên cứu đã thực hiện các phương pháp như trình bày dưới đây để đạt được mục tiêu đề ra.

2.1. Phương pháp tổng quan, phân tích chính sách, quy định về dữ liệu mở

Để thực hiện mục tiêu thứ nhất, nhóm nghiên cứu đã thực hiện phương pháp nghiên cứu tại bàn, tập trung tổng quan, phân tích các quy định pháp lý liên quan đến DLM ở Việt Nam, bao gồm các văn bản quy phạm pháp luật (Luật, Nghị định), các văn bản pháp lý của nhà nước (Quyết định của Thủ tướng Chính phủ), các kế hoạch triển khai thực hiện các chính sách, quy định ở cấp địa phương và các văn bản khác. Các vấn đề được

[4] Bộ Nội vụ và Viện Khoa học tổ chức nhà nước. Thực trạng và giải pháp phát triển DLM tại Việt Nam. 2022 [trích dẫn ngày 8 tháng 5 năm 2024]; Có tại: https://isos.gov.vn/cds/thuc-trang-va-giai-phap-phat-trien-du-lieu-mo-tai-viet-nam-43241.html

Official Use Only

tổng quan, rà soát theo khung phân tích đi từ mục đích, ý nghĩa của việc cung cấp DLM, khái niệm, định nghĩa về DLM đến các quy định về triển khai cung cấp và quản lí DLM, cụ thể: (i) mục đích, ý nghĩa của việc cung cấp DLM; (ii) khái niệm và đặc điểm về DLM; (iii) nguyên tắc tiếp cận, sử dụng và cung cấp DLM; (iv) cách thức công bố DLM; (v) trách nhiệm phối hợp và thực hiện.

Các văn bản quy phạm pháp luật trong Luật[5], Nghị định[6], các Chiến lược, đề án quốc gia do Thủ tướng chính phủ ban hành[7], các Quyết định của các bộ, cơ quan, địa phương ban hành kế hoạch, danh mục DLM nhằm triển khai thực hiện cung cấp DLM đã được nhóm nghiên cứu tổng hợp để rà soát, tổng quan các nội dung liên quan. Tổng cộng có chín mươi bốn văn bản quy phạm pháp luật được rà soát. Danh sách các văn bản liên quan đến DLM xem tại *Báo cáo Tổng quan chính sách của Việt Nam về DLM và một số thực tiễn tốt trên thế giới*. Song song với phân tích khung pháp lý trong nước, nhóm nghiên cứu cũng tiến hành rà soát các tài liệu quốc tế và thông lệ thực hành tốt từ nhiều quốc gia đã triển khai DLM thành công như Pháp, Anh, Hàn Quốc, Mỹ và Úc. Việc nghiên cứu tại bàn này tập trung vào các yếu tố như tiêu chuẩn kĩ thuật, định dạng dữ liệu, quy định bảo mật, mô hình vận hành, cơ chế quản lý và chính sách pháp lý điều chỉnh DLM. Kết quả từ quá trình đối sánh này giúp làm rõ những điểm tương đồng và khác biệt giữa Việt Nam và các quốc gia đi đầu, từ đó củng cố thêm căn cứ cho các khuyến nghị chính sách và kĩ thuật phù hợp với bối cảnh Việt Nam.

Ngoài ra, trong quá trình thực hiện phỏng vấn các bộ, cơ quan, địa phương về việc triển khai cung cấp DLM, xây dựng cổng DLM, nhóm nghiên cứu cũng thu thập ý kiến của các CQNN về các quy định, chính sách hiện hành có liên quan.

2.2. Phương pháp đánh giá cổng dữ liệu mở tại Việt Nam

Nhóm nghiên cứu đã thực hiện đánh giá từng cổng để có được bức tranh tổng quan hiện trạng các cổng DLM. Nhóm nghiên cứu đã áp dụng bộ công cụ gồm các tiêu chí đánh giá cụ thể, từ đó so sánh và chỉ ra điểm mạnh, điểm yếu của từng cổng DLM của tất các Bộ/ngành/địa phương hiện có. Các dữ liệu liên quan từ các cổng đã được trích xuất, bao gồm thông tin giới thiệu cổng, đơn vị quản lý, cấu trúc của cổng, danh mục DLM, thông tin về người dùng, thời lượng và thời điểm truy cập, các bộ dữ liệu hiện có và các thông tin đặc tả dữ liệu. Các thông tin, dữ liệu này được sử dụng để cung cấp bằng chứng cho đánh giá toàn diện về cổng, ví dụ như tốc độ truy cập, số lượng bộ dữ liệu hiện có, định dạng và thời gian tải dữ liệu từ cổng, cùng các tiêu chí kĩ thuật khác.

Hệ thống tiêu chí đánh giá được xây dựng dựa trên việc đánh giá các thực tiễn tốt về DLM trên thế giới, bao gồm các quốc gia như Anh, Úc, Hàn Quốc, Singapore, các quy định hiện hành tại Việt Nam, Dự thảo Quyết định Ban hành yêu cầu cơ bản về chức năng, tính năng kĩ thuật của Cổng DLM (Phiên bản 1.0) của Bộ TT&TT và Tiêu chuẩn quốc gia về Đặc tả tập DLM cho Cổng dữ liệu (DCAT-VN) ban hành năm 2024.

Bộ tiêu chí đánh giá các cổng DLM tại Việt Nam bao gồm 27 tiêu chí, chia làm hai phần chính *(chi tiết xem tại Báo cáo đánh giá cổng DLM tại Việt Nam)*: **(I) Đặc điểm chung của các cổng DLM**: bao gồm 08 tiêu chí về yêu cầu chung của cổng và 08 tiêu chí về công cụ hỗ trợ người dùng; **(II) Đặc điểm của DLM được cung cấp trên cổng**: bao gồm 07 tiêu chí về *Yêu cầu liên quan đến dữ liệu và 04 tiêu chí về Yêu cầu đặc tả dữ liệu (metadata).* Mỗi tiêu chí có từ 1 đến 8 yếu tố cụ thể được đánh giá, nếu cổng đáp ứng mỗi yếu tố sẽ được 1

[5] Như Luật Công nghệ thông tin năm 2006; Luật Bảo vệ bí mật nhà nước năm 2018; Luật Tiếp cận thông tin năm 2016; Luật Giao dịch điện tử (GDĐT) năm 2023; Luật Dữ liệu năm 2024.

[6] Như Nghị định 47/2020/NĐ-CP ngày 9 tháng 4 năm 2020 của Chính phủ về quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước; Nghị định số 42/2022/NĐ-CP ngày 24 tháng 6 năm 2022 của Chính phủ quy định về việc cung cấp thông tin và dịch vụ công trực tuyến của cơ quan Nhà nước trên môi trường mạng; Nghị định 13/2023/NĐ-CP ngày 17 tháng 4 năm 2023 của Chính phủ về bảo vệ dữ liệu cá nhân.

[7] Như Quyết định 749/QĐ-TTg ngày 03 tháng 6 năm 2020 của Thủ tướng Chính phủ phê duyệt "Chương trình chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030"; Quyết định 942/QĐ-TTg ngày 15 tháng 6 năm 2021 của Thủ tướng Chính phủ phê duyệt chiến lược phát triển Chính phủ điện tử hướng tới Chính phủ số giai đoạn 2021-2025, định hướng đến năm 2030; Quyết định 411/QĐ-TTg ngày 31 tháng 3 năm 2022 của Thủ tướng Chính phủ phê duyệt chiến lược Quốc gia phát triển kinh tế số và xã hội số đến năm 2025, định hướng đến năm 2030; Quyết định 146/QĐ-TTg ngày 28 tháng 01 năm 2022 của Thủ tướng Chính phủ phê duyệt Đề án "Nâng cao nhận thức, phổ cập kĩ năng và phát triển nguồn nhân lực chuyển đổi số Quốc gia đến năm 2025, định hướng đến năm 2030.

Official Use Only

điểm. Điểm của từng tiêu chí sau đó sẽ được quy đổi về thang điểm từ 0 đến 1. Tổng điểm cuối cùng được tính dựa trên tổng điểm của từng phần (1 và 2) nhân với trọng số của phần đó (50% mỗi phần). Tổng điểm tối đa của một cổng là 27 điểm.

Tổng cộng có 37 cổng DLM đã được đánh giá, trong đó có 4 cổng cấp bộ, 1 cổng quốc gia, 32 cổng địa phương. Tuy nhiên, cuối cùng chỉ có 34 cổng truy cập được và có cung cấp các bộ dữ liệu để đánh giá nên báo cáo này trình bày kết quả đánh giá chung đối với 34 cổng gồm 3 cổng cấp bộ và 1 cổng quốc gia và 30 cổng địa phương. Các cổng DLM đều đang trong quá trình xây dựng và phát triển do vậy các kết quả trình bày trong nghiên cứu này là tại thời điểm thực hiện đánh giá cổng DLM từ ngày 15/10/2024 đến ngày 15/11/2024.

2.3. Phương pháp khảo sát thực trạng cung cấp dữ liệu mở tại Việt Nam

Nghiên cứu kết hợp phương pháp khảo sát định lượng và định tính để thu thập dữ liệu từ các CQNN, bao gồm các bộ, cơ quan, đơn vị cấp Trung ương (TW) và các cơ quan, đơn vị cấp địa phương. Dữ liệu được thu thập thông qua khảo sát trực tuyến, kết hợp phỏng vấn sâu (PVS) với đại diện các CQNN trực tiếp tham gia triển khai cung cấp DLM và quản lý cổng DLM. Khảo sát định lượng được thực hiện trực tuyến đối với các cán bộ của các cơ quan, đơn vị cung cấp và quản lý DLM ở TW và địa phương gồm các nội dung về:

1. Nhận thức về DLM và các quy định pháp luật liên quan;
2. Thực trạng triển khai cung cấp DLM;
3. Khó khăn trong quá trình cung cấp DLM.

Trong tổng số 438 phiếu phát ra, tổng cộng có 280 đại diện từ các CQNN đã tham gia khảo sát định lượng (tỷ lệ phản hồi đạt 64%), gồm 78 đại diện CQNN cấp TW (28%), 202 đại diện CQNN cấp địa phương (72%). Các CQNN được khảo sát gồm 105 đơn vị chuyên trách công nghệ thông tin (CNTT) của các bộ, ban, ngành và của địa phương (là các Cục, Vụ, Trung tâm CNTT hoặc tương đương và Sở TT&TT các địa phương) và 175 đơn vị cung cấp dữ liệu ở địa phương (như Sở Giáo dục và Đào tạo, Sở Khoa học và Công nghệ, Sở Kế hoạch và Đầu tư, Sở Tài nguyên và Môi trường, Sở Tài chính, Sở Nông nghiệp và Phát triển Nông thôn, Sở Lao động - Thương binh và Xã hội, Sở Xây dựng, Sở Giao thông vận tải, Sở Y tế...).

PVS đã được thực hiện nhằm tìm hiểu ý kiến, quan điểm của các cán bộ, lãnh đạo CQNN về việc triển khai cung cấp DLM trên thực tế và tầm quan trọng, thách thức, những kỳ vọng đối với việc triển khai DLM. Có 07 cuộc PVS đã được thực hiện với 12 cán bộ đại diện 7 CQNN (2 CQNN cấp TW và 5 CQNN cấp địa phương). 02 Bộ, 01 tỉnh được liên hệ mời tham gia phỏng vấn nhưng không sắp xếp/bố trí được thời gian.

2.4. Phương pháp khảo sát nhu cầu sử dụng dữ liệu mở tại Việt Nam

Tương tự như khảo sát về cung cấp DLM, nghiên cứu kết hợp phương pháp khảo sát định lượng và PVS với các đối tượng sử dụng DLM là các tổ chức, doanh nghiệp (bao gồm viện nghiên cứu/trường đại học công lập và ngoài công lập, tổ chức quốc tế, tổ chức phi chính phủ, tổ chức xã hội và cơ quan truyền thông/báo chí) và người dùng cá nhân. Nội dung khảo sát gồm 3 phần chính*:*

(i) Nhận thức về DLM và các quy định pháp luật liên quan;

(ii) Nhu cầu sử dụng DLM;

(iii) Đánh giá chung về trải nghiệm khai thác, sử dụng các cổng DLM và DLM tại Việt Nam.

Khảo sát định lượng được thực hiện trực tuyến với tổng số 1200 phiếu phát ra, thu được 420 phiếu trả lời của 325 tổ chức, doanh nghiệp (TCDN) và 95 người dùng cá nhân (tỷ lệ phản hồi đạt 35%). Trong đó, nhóm TCDN có hơn 60% là các viện nghiên cứu/trường đại học (ĐH), 22% là doanh nghiệp, 8% là các tổ chức quốc tế/tổ chức phi chính phủ và 6% là các cơ quan truyền thông, báo chí. Trong nhóm người dùng cá nhân, có 95 người thuộc lĩnh vực chuyên môn khác nhau như kinh tế, tài chính; quản lý và chính sách công; xã hội; pháp luật; môi trường, khí hậu và phát triển bền vững; công nghệ và đổi mới sáng tạo. Ngoài ra, trong khảo sát

Official Use Only

các CQNN là các đơn vị cung cấp DLM và quản lý cổng DLM của các bộ, cơ quan và địa phương, nhóm nghiên cứu cũng lồng ghép tìm hiểu thực trạng và nhu cầu sử dụng DLM của cán bộ các cơ quan này.

Bên cạnh khảo sát định lượng, nghiên cứu còn thực hiện PVS tìm hiểu ý kiến, quan điểm của cá nhân, TCDN sử dụng DLM (04 cuộc PVS đã được thực hiện với các cá nhân, TCDN là người đã từng sử dụng DLM).

PHẦN II. MỘT SỐ KẾT QUẢ NGHIÊN CỨU CHÍNH

1. Các quy định, chính sách hiện hành về dữ liệu mở

DLM có vai trò quan trọng trong việc thúc đẩy phát triển kinh tế - xã hội, là mũi nhọn chiến lược để phát triển chính phủ số và xã hội số. Luật GDĐT năm 2023 nêu rõ việc công bố DLM cho các cơ quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ *"nhằm thúc đẩy giao dịch điện tử, chuyển đổi số, phát triển kinh tế số và xã hội số"*. Luật Dữ liệu năm 2024 quy định DLM là một trong ba loại dữ liệu do CQNN chia sẻ[8], theo đó, mọi cơ quan, tổ chức, cá nhân có nhu cầu đều được tiếp cận, chia sẻ, khai thác, sử dụng DLM. Ngoài ra, dữ liệu là một loại thông tin[9], do vậy, một số quy định pháp luật được hiểu cũng áp dụng đối với DLM như Luật Tiếp cận thông tin (năm 2016), Luật Bảo vệ bí mật nhà nước (năm 2018), Luật Công nghệ thông tin (năm 2006, sửa đổi, bổ sung năm 2017).

Các Chiến lược phát triển kinh tế số, xã hội số, Chiến lược phát triển Chính phủ điện tử, Chương trình Chuyển đổi số quốc gia khuyến khích các CQNN mở dữ liệu để phục vụ cho sự phát triển chung nhằm: i) minh bạch hóa và tăng cường hoạt động trao đổi, chia sẻ dữ liệu; ii) phục vụ chuyển đổi số quốc gia; iii) và đáp ứng nhu cầu sử dụng dữ liệu của công dân tạo ra các sản phẩm kinh doanh, dịch vụ. Cụ thể, Chương trình Chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030 (Quyết định số 749/QĐ-TTg*)* đã đặt mục tiêu chuyển đổi số và phát triển Chính phủ điện tử hướng tới Chính phủ số, bao gồm việc phát triển hạ tầng số và tạo lập DLM. Chương trình Chuyển đổi số quốc gia đã đề cập đến mục tiêu cung cấp cơ sở dữ liệu (CSDL) quốc gia về Dân cư, Đất đai, Đăng ký doanh nghiệp, Tài chính, Bảo hiểm được hoàn thành và kết nối, chia sẻ trên toàn quốc; từng bước mở dữ liệu của các CQNN để cung cấp dịch vụ công kịp thời, một lần khai báo, trọn vòng đời phục vụ người dân và phát triển kinh tế - xã hội.

Chiến lược phát triển Chính phủ điện tử hướng tới Chính phủ số giai đoạn 2021-2025, định hướng đến năm 2030 (Quyết định số 942/QĐ-TTg) và Chiến lược quốc gia phát triển kinh tế số và xã hội số đến năm 2025, định hướng đến năm 2030 (Quyết định số 411/QĐ-TTg ngày 31/3/2022 của Thủ tướng Chính phủ) xác định dữ liệu số là yếu tố quan trọng của kinh tế số và xã hội số, với mục tiêu chuyển từ sở hữu riêng dữ liệu sang chia sẻ và khai thác chung. Cụ thể, các quyết định đặt ra hai mục tiêu liên quan trực tiếp đến DLM, bao gồm: (1) 100% CQNN cấp bộ, tỉnh tham gia mở dữ liệu và cung cấp DLM phục vụ phát triển Chính phủ số, kinh tế số, xã hội số; (2) Xếp hạng quốc gia của Việt Nam về phát triển Chính phủ điện tử, Chính phủ số theo đánh giá của Liên Hợp Quốc có sự thay đổi đột phá, trong đó Việt Nam thuộc nhóm 50 nước dẫn đầu về chỉ số DLM.

Chiến lược phát triển Chính phủ điện tử hướng tới Chính phủ số giai đoạn 2021-2025, định hướng đến năm 2030 yêu cầu ưu tiên chuyển đổi số với một số ngành và lĩnh vực trước, trong đó, chú trọng tới việc triển khai các sáng kiến nhằm liên kết giữa các ngành, lĩnh vực để cung cấp một trải nghiệm mới, hoàn toàn khác, mang lại giá trị cho người dân, doanh nghiệp và xã hội bao gồm y tế, giáo dục, tài chính- ngân hàng, nông nghiệp, giao thông vận tải, tài nguyên môi trường. Chiến lược quốc gia phát triển kinh tế số và xã hội số đến năm 2025, định hướng đến năm 2030 nêu DLM trong ba lĩnh vực trọng điểm, bao gồm: nông nghiệp, lao động - việc làm - an sinh xã hội và du lịch cần được ưu tiên phát triển nhằm hướng tới kinh tế số vã xã hội số nhằm cung cấp thông tin, tư vấn, dự báo tình hình cho người dân.

Mặc dù đã có nhiều văn bản quy phạm pháp luật và các quy định, chính sách có liên quan đến cung cấp DLM song kết quả nghiên cứu, tổng quan chỉ ra vẫn còn nhiều khoảng trống và hạn chế trong các chính sách, quy

[8] Điểm a, Khoản 1, Điều 13. Phân loại dữ liệu, Luật Dữ liệu năm 2024 quy định "*1. Cơ quan nhà nước phải phân loại dữ liệu dựa trên yêu cầu quản trị, xử lý, bảo vệ dữ liệu, bao gồm: a) Phân loại theo tính chất chia sẻ dữ liệu gồm: dữ liệu dùng chung, dữ liệu dùng riêng, dữ liệu mở*".

[9] Khoản 1, Điều 2. Giải thích từ ngữ, Luật Tiếp cận thông tin quy định năm 2016: *"Thông tin là tin, dữ liệu được chứa đựng trong văn bản, hồ sơ, tài liệu có sẵn, tồn tại dưới dạng bản viết, bản in, bản điện tử, tranh, ảnh, bản vẽ, băng, đĩa, bản ghi hình, ghi âm hoặc các dạng khác do cơ quan nhà nước tạo ra".*

định hiện hành cần được giải quyết nhằm nâng cao hiệu quả, hiệu lực thực thi trong việc cung cấp DLM. Một số vấn đề đáng chú ý được tổng hợp trong các mục dưới đây.

1.1. Khái niệm và đặc điểm của dữ liệu mở

Trong các văn bản pháp luật hiện nay có một số khái niệm liên quan đến dữ liệu và DLM như "thông tin"[10], "dữ liệu"[11], "dữ liệu số"[12], "dữ liệu mở"[13], "dữ liệu mở của cơ quan nhà nước"[14], "dữ liệu dùng chung"[15], "dữ liệu dùng riêng"[16], "dữ liệu gốc"[17].

Hiện nay có ba văn bản quy phạm pháp luật đưa ra định nghĩa về DLM (Nghị định 47/2020/NĐ-CP, Luật GDĐT năm 2023 và Luật Dữ liệu năm 2024), các định nghĩa không hoàn toàn thống nhất (và chỉ áp dụng trong phạm vi văn bản đó) nhưng có những điểm chung cho thấy DLM là "dữ liệu số" (dữ liệu "được biểu diễn bằng tín hiệu số" hay "dùng tín hiệu số" hay "được thể hiện dưới dạng kĩ thuật số"), do CQNN chia sẻ/cung cấp và mọi cơ quan, tổ chức, cá nhân được tiếp cận, sử dụng miễn phí.

Một số đặc điểm khác của DLM không được nêu trong định nghĩa của các văn bản quy phạm pháp luật hiện hành mà nêu trong một số quy định có liên quan như:

- DLM ở định dạng mở (điểm e, Khoản 1, Điều 17. Nguyên tắc cung cấp dữ liệu mở của cơ quan nhà nước, Nghị định 47/2020/NĐ-CP).
- Không tiết lộ thông tin cá nhân (Khoản 1, Điều 20. Yêu cầu trong việc công bố dữ liệu mở của cơ quan nhà nước, Nghị định 47/2020/NĐ-CP).

Kết quả khảo sát của nghiên cứu đối với các cán bộ CQNN cả ở cấp trung ương và địa phương cho thấy phần lớn (80,7%) cán bộ đã từng nghe đến khái niệm "dữ liệu mở" và 75,4% đã tìm hiểu, nghiên cứu các quy định liên quan đến DLM như Luật GDĐT, Nghị định 47/2020/NĐ-CP. Song phần lớn cán bộ CQNN cho rằng DLM có thể là PDF, DOCX (với tỷ lệ lần lượt là 74,3% và 63,2%) trong khi theo các thực hành tốt trên thế giới và

---

[10] "***Thông tin***": Theo khoản 1, Điều 2. Giải thích từ ngữ, Luật tiếp cận thông tin năm 2016, thông tin là *"tin, dữ liệu được chứa đựng trong văn bản, hồ sơ, tài liệu có sẵn, tồn tại dưới dạng bản viết, bản in, bản điện tử, tranh, ảnh, bản vẽ, băng, đĩa, bản ghi hình, ghi âm hoặc các dạng khác do cơ quan nhà nước tạo ra*"; Theo khoản 2, Điều 2. Giải thích từ ngữ, Luật tiếp cận thông tin năm 2016, "***Thông tin do cơ quan nhà nước tạo ra***" là *"tin, dữ liệu được tạo ra trong quá trình cơ quan nhà nước thực hiện chức năng, nhiệm vụ, quyền hạn theo quy định của pháp luật, được người có thẩm quyền của cơ quan nhà nước đó ký, đóng dấu hoặc xác nhận bằng văn bản*".

[11] "***Dữ liệu***": Theo khoản 6 Điều 3. Giải thích từ ngữ, Luật Giao dịch điện tử năm 2023, dữ liệu là *"ký hiệu, chữ viết, chữ số, hình ảnh, âm thanh hoặc dạng tương tự khác"*.

[12] "***Dữ liệu số"****:* Theo khoản 1, Điều 3, Nghị định 47/2020/NĐ-CP dữ liệu số là *"dữ liệu dưới dạng ký hiệu, chữ viết, chữ số, hình ảnh, âm thanh hoặc dạng tương tự được biểu diễn bằng tín hiệu số. Dữ liệu số mang thông tin số và được chia sẻ dưới dạng thông điệp dữ liệu. Trong Nghị định này, dữ liệu được hiểu là dữ liệu số*"; Theo khoản 8, Điều 3, Luật GDĐT, dữ liệu số là "*dữ liệu điện tử được tạo lập bằng phương pháp dùng tín hiệu số*"; theo khoản 1, Điều 3. Giải thích từ ngữ, Luật Dữ liệu năm 2024, dữ liệu số là *"dữ liệu về sự vật, hiện tượng, sự kiện, bao gồm một hoặc kết hợp các dạng âm thanh, hình ảnh, chữ số, chữ viết, ký hiệu được thể hiện dưới dạng kĩ thuật số (sau đây gọi là dữ liệu)*".

[13] "***Dữ liệu mở***": là dữ liệu mà "*mọi cơ quan, tổ chức, cá nhân có nhu cầu đều được tiếp cận, chia sẻ, khai thác, sử dụng"* (Khoản 4, Điều 3. Giải thích từ ngữ, Luật Dữ liệu năm 2024)

[14] "***Dữ liệu mở của cơ quan nhà nước***": Theo khoản 4, Điều 3. Giải thích từ ngữ, Nghị định 47/2020/NĐ-CP ngày 09/4/2020, dữ liệu mở của cơ quan nhà nước là "*dữ liệu được CQNN có thẩm quyền công bố rộng rãi cho cơ quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ. Trong Nghị định này, dữ liệu mở được hiểu là dữ liệu mở của cơ quan nhà nước"*. Định nghĩa "*Dữ liệu mở của cơ quan nhà nước*" tại Luật GDĐT 2023 (Khoản 1, Điều 43) cũng tương tự như Nghị định 47/2020/NĐ-CP.

[15] "***Dữ liệu dùng chung****":* Theo khoản 2, Điều 3. Giải thích từ ngữ, Luật Dữ liệu năm 2024, dữ liệu dùng chung là dữ liệu *được tiếp cận, chia sẻ, khai thác, sử dụng chung trong các cơ quan Đảng, Nhà nước, Ủy ban Mặt trận Tổ quốc Việt Nam và các tổ chức chính trị - xã hội*. Nghị định 47/2020/NĐ-CP nêu khái niệm *"****dữ liệu danh mục dùng chung***", theo đó dữ liệu danh mục dùng chung là *"dữ liệu về các danh mục, bảng mã phân loại do cơ quan nhà nước có thẩm quyền ban hành, được sử dụng chung trong các hệ thống thông tin, cơ sở dữ liệu bảo đảm việc tích hợp, trao đổi, chia sẻ dữ liệu đồng bộ, thống nhất"* (Điều 3. Giải thích từ ngữ)

[16] "***Dữ liệu dùng riêng***": Theo khoản 3, Điều 3. Giải thích từ ngữ, Luật Dữ liệu năm 2024 là dữ liệu "*được tiếp cận, chia sẻ, khai thác, sử dụng trong phạm vi nội bộ của cơ quan Đảng, Nhà nước, Ủy ban Mặt trận Tổ quốc Việt Nam và các tổ chức chính trị - xã hội".*

[17] "***Dữ liệu gốc***": là dữ liệu *"được tạo lập trong quá trình hoạt động của cơ quan, tổ chức, cá nhân hoặc thu thập, tạo lập từ số hóa bản chính giấy tờ, tài liệu, các dạng vật chất khác"* (Khoản 5, Điều 3. Giải thích từ ngữ, Luật Dữ liệu năm 2024)

Official Use Only

các định nghĩa chuẩn về DLM được cung cấp bởi World Bank[18], DLM phải được cung cấp dưới dạng máy đọc được (machine readable) như XLSX, CSV, JSON hay dạng API. Tại châu Âu, dữ liệu định dạng PDF không được coi là DLM[19]. Việc các cán bộ chưa có nhận thức đầy đủ về các định dạng của DLM và cung cấp DLM là bởi các quy định hiện hành chưa đủ rõ ràng và phù hợp với thông lệ quốc tế.

Theo các quy định pháp luật hiện hành, có thể tổng hợp lại một số vấn đề đáng chú ý liên quan đến khái niệm, đặc điểm của DLM như sau:

(i) DLM được hiểu là một loại thông tin của CQNN (theo Luật tiếp cận thông tin). Theo quy định, thông tin của CQNN phải được người có thẩm quyền của CQNN ký, đóng dấu hoặc xác nhận bằng văn bản. Điều này lí giải cho việc trên cổng DLM của nhiều địa phương có nhiều tệp dữ liệu định dạng PDF được chụp từ các văn bản của CQNN với chữ ký, con dấu của người có thẩm quyền *(xem thêm Báo cáo đánh giá cổng DLM tại Việt Nam)*.

(ii) Định dạng của DLM (định dạng mở): Theo Khoản 8, Điều 3, Nghị định 47/2020/NĐ-CP, "*định dạng mở*" là "*định dạng của tập tin hoặc thông điệp dữ liệu được quy định theo công bố kĩ thuật của tổ chức hoạt động về tiêu chuẩn và không bị hạn chế khi sử dụng công bố kĩ thuật này*". Ngày 22/03/2024, Bộ TT&TT đã có Công văn số 1016/BTTTT-CĐSQG hướng dẫn xây dựng, áp dụng quy chuẩn, tiêu chuẩn, quy định kĩ thuật kết nối, chia sẻ dữ liệu ban hành đã hướng dẫn các định dạng dữ liệu về kết nối và chia sẻ dữ liệu. Theo đó, nếu dữ liệu là văn bản, các định dạng như TXT, RTF, DOCX, DOC, PDF hoặc ODT được lựa chọn tùy theo nhu cầu thể hiện. Đối với dữ liệu dạng bảng tính, Bộ TT&TT khuyến nghị nên sử dụng các định dạng CSV, XLSX, XLS hoặc ODS. Trong trường hợp dữ liệu phục vụ trình diễn, các tiêu chuẩn như HTM, PPTX, PPT, PDF hoặc ODP sẽ được áp dụng. Với dữ liệu ảnh đồ họa, các định dạng JPEG, GIF, TIFF hoặc PNG là phù hợp, còn với ảnh cần lưu trữ thông tin vị trí địa lý kèm theo, định dạng GEO TIFF sẽ được sử dụng. Tuy nhiên, hướng dẫn này của Bộ TT&TT hiện nay chỉ là văn bản hành chính, chưa phải "công bố kĩ thuật của tổ chức hoạt động về tiêu chuẩn" nên đến nay các địa phương vẫn có các hướng dẫn, cách hiểu và thực hiện khác nhau về định dạng của DLM. Thực trạng các tệp dữ liệu trên các cổng hiện nay chủ yếu phổ biến định dạng PDF và một số định dạng khác không phải là định dạng máy đọc được theo thông lệ quốc tế *(xem thêm Báo cáo đánh giá cổng DLM tại Việt Nam).*

(iii) Đặc điểm dữ liệu đọc được bằng máy (machine-readable) chưa được nêu rõ như các nước trên thế giới, dẫn đến sự thiếu tương thích với các tiêu chuẩn quốc tế. Theo Ngân hàng thế giới và nhiều quốc gia, DLM được quy định hoặc khuyến khích cung cấp ở các định dạng máy đọc được như JSON, XML, CSV, TSV, TDP, RDF, geoJSON và KML. Ngân hàng thế giới không công nhận DLM dưới định dạng PDF hay PNG[20]. Như vậy, các quy định hiện hành của Việt Nam về định dạng của DLM chưa đủ rõ và chưa tương thích với thông lệ quốc tế. Điều này đặc biệt quan trọng trong bối cảnh hội nhập và hợp tác quốc tế ngày càng gia tăng, nơi mà việc sử dụng các định dạng dữ liệu thống nhất có thể giúp nâng cao hiệu quả khai thác và chuyển đổi dữ liệu.

Với nhiều thuật ngữ liên quan đến dữ liệu như hiện nay, các cơ quan, đơn vị khi triển khai thực hiện còn nhầm lẫn, chưa phân biệt "dữ liệu mở" và "dữ liệu dùng chung" cũng như chưa định hình DLM thuộc CSDL nào, lúng túng trong việc cung cấp DLM, e ngại cung cấp DLM không đúng quy định, có các trường hợp sử dụng dữ liệu dùng chung để cung cấp như là DLM dẫn đến không đảm bảo yêu cầu không tiết lộ thông tin cá nhân và một số vấn đề khác. Hiện nay, Luật Dữ liệu năm 2024 đã phân loại dữ liệu của CQNN theo tính chất chia sẻ một cách rõ ràng hơn, gồm "dữ liệu dùng chung, dữ liệu dùng riêng, dữ liệu mở" (điểm a, Khoản 1 Điều 13. Phân loại dữ liệu) nhưng chưa có các quy định cụ thể hơn về DLM trong Luật này.

[18] Ngân hàng Thế giới, Khóa học về dữ liệu dành cho các nhà cung cấp dữ liệu. https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/essentials

[19] Data.europa.edu. Chọn đúng định dạng cho DLM. *https://data.europa.eu/elearning/en/module9/#/id/co-01*

[20] Ngân hàng Thế giới, Khóa học về dữ liệu dành cho các nhà cung cấp dữ liệu. https://blogs.worldbank.org/en/opendata/machine-readable-open-data-how-it-s-applicable-developing-countries#:~:text=For%20example%2C%20consider%20a%20PDF,spreadsheet%20would%20be%20machine%20readable

Official Use Only

Hộp 1: Định nghĩa về dữ liệu mở của các nước trên thế giới

Pháp: Cổng DLM của Pháp cung cấp danh mục thuật ngữ liên quan đến DLM trong đó định nghĩa DLM là dữ liệu có quyền truy cập mở, miễn phí và mọi người đều có thể dễ dàng sử dụng lại. Dữ liệu này không chỉ được tạo ra bởi chính quyền (các bộ, chính quyền địa phương, v.v...) mà còn bởi các tổ chức tư nhân hoặc thậm chí là công dân. Tại "tài nguyên OpenDataFrance"[21] - một bộ công cụ có mục đích hướng dẫn, tạo điều kiện thuận lợi cho các đơn vị hành chính công của Pháp trong quá trình mở dữ liệu, DLM được định nghĩa là dữ liệu số có thể truy cập miễn phí, định dạng máy có thể sử dụng được và mọi người có thể sử dụng lại. OpenDataFrance liệt kê các định dạng mở để giải thích rõ hơn về định nghĩa. Danh sách định dạng này tương thích với cấu trúc tập dữ liệu để công bố theo hướng dẫn tại cổng DLM quốc gia của Pháp[22] hay Tiêu chí lựa chọn cổng thông tin DLM (Critère de choix d'un portail Open data). Điểm đặc biệt của các danh sách các định dạng mở này là đều tập trung vào các định dạng phổ biến như CSV và chỉ ra rằng định dạng PDF không được tính là DLM.

Mỹ: Các định nghĩa về DLM hoặc có liên quan đến DLM được quy định rõ ràng trong Đạo luật DLM của Chính phủ (OPEN Government Data Act) (Tiêu đề II của Đạo luật Nền tảng để Xây dựng Chính sách dựa trên Bằng chứng năm 2018, Luật Công 115-435)[23]. Theo đó, thuật ngữ *"tài sản DLM của Chính phủ"* (open Government data asset) là tài sản dữ liệu công khai có thể đọc được bằng máy; có sẵn (hoặc có thể được cung cấp) dưới định dạng mở; không bị giới hạn bởi các ràng buộc, ngoại trừ quyền sở hữu trí tuệ, bao gồm quyền theo điều khoản 17 và 35, có thể cản trở việc sử dụng hoặc tái sử dụng tài sản này; và dựa trên một tiêu chuẩn mở được duy trì bởi một tổ chức tiêu chuẩn. Trong đó, thuật ngữ "đọc được bằng máy" (machine-readable), khi được áp dụng cho dữ liệu, có nghĩa là dữ liệu ở định dạng có thể dễ dàng được xử lý bởi máy tính mà không cần sự can thiệp của con người và vẫn đảm bảo không mất ý nghĩa ngữ nghĩa. Ngoài ra, Mỹ còn định nghĩa thuật ngữ "*tài sản dữ liệu công*" (public data asset) đó là một tài sản dữ liệu hoặc một phần của nó, được duy trì bởi chính phủ liên bang và đã hoặc có thể được công khai cho công chúng.

Úc: DLM được định nghĩa là dữ liệu do Chính phủ sở hữu và là dữ liệu "có sẵn, miễn phí, dễ dàng tìm kiếm, truy cập và được công bố dưới nhiều hình thức và được cấp phép tái sử dụng" [24]. Ngoài ra, DLM còn được biết đến với tính có cấu trúc (dữ liệu được lưu lại một cách đồng nhất dựa trên một khuôn mẫu chung); có thể đọc được bằng máy, cho phép dữ liệu được số hóa và tự động xử lý để truy xuất, tải xuống, lập chỉ mục và tìm kiếm; được cấp phép để giảm thiểu giới hạn về cách mà mọi người có thể sử dụng và phân phối lại dữ liệu. Tuyên bố Chính sách dữ liệu công của Chính phủ Úc ngày 7/12/2015 nêu rõ các dữ liệu không mang tính nhạy cảm sẽ được mở mặc định và có thể truy cập thông qua các Giao diện Lập trình Ứng dụng (API) miễn phí, có chất lượng cao và đáng tin cậy.

Anh: DLM được định nghĩa là dữ liệu đáp ứng các tiêu chí sau: có thể truy cập được (lý tưởng nhất là qua internet) với chi phí không vượt quá chi phí tái bản, danh tính người dùng hoặc mục đích sử dụng không bị hạn chế; dưới dạng có thể đọc được bằng máy để tương tác được với các dữ liệu khác; và không bị hạn chế sử dụng hoặc hạn chế phân phối lại theo các điều kiện được cấp phép[25]. Ngoài ra, DLM của chính phủ được hiểu là các thông tin của khu vực công được cung cấp tới công chúng dưới dạng DLM. Tất cả DLM của chính phủ sẽ có thể truy cập được thông qua trang data.gov.uk, mặc dù dữ liệu có thể được lưu trữ trên các trang web của từng cơ quan và các điểm truy cập khác, ví dụ như của Cơ quan Thống kê Quốc gia.

Hàn Quốc: DLM tại Hàn Quốc (hay còn lại là dữ liệu công - "public data") được định nghĩa tại khoản 2 Điều 2 là bất kỳ dữ liệu hoặc thông tin nào, bao gồm CSDL và tệp điện tử được tổng hợp dưới dạng quang học hoặc điện tử, hoặc thu thập và quản lý bởi các tổ chức công, để phục vụ các mục đích được quy định trong luật; bao gồm thông tin hành chính, dữ liệu của các tổ chức công, hồ sơ điện tử và các dữ liệu khác được quy định theo Nghị định của

[21] Ngân hàng Thế giới, Khóa học về dữ liệu dành cho các nhà cung cấp dữ liệu: https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/essentials

[22] Data.gouv.fr, Cấu trúc một tập dữ liệu: https://guides.data.gouv.fr/guides-open-data/guide-qualite/preparer-un-jeu-de-donnees-de-qualite/structurer-un-jeu-de-donnees

[23] Đạo luật dữ liệu chính phủ MỞ https://www.congress.gov/bill/115th-congress/house-bill/4174

[24] Vì sao dữ liệu Chính phủ mở có vai trò quan trọng?: https://www.infrastructure.gov.au/sites/default/files/documents/BCR%20Open%20government%20data%20report%20-%20final%20-%205%20Feb2016.pdf

[25] Bộ trưởng Văn phòng Nội các và Tổng cục trưởng Tài chính. Sách trắng về DLM. 2012

Tổng thống[26]. Một trong những nội dung chủ chốt trong chính sách DLM của Hàn Quốc là nhấn mạnh vào định dạng có thể đọc được bằng máy. Luật đề cập "định dạng có thể đọc bằng máy" là dữ liệu có thể được xử lý, chỉnh sửa hoặc phân tích bởi phần mềm. Yêu cầu về định dạng đọc được bằng máy rất quan trọng để đảm bảo DLM không chỉ được cung cấp mà còn có thể sử dụng một cách hiệu quả cho việc phân tích, phát triển ứng dụng và các mục đích khác.

1.2. Nguyên tắc tiếp cận, sử dụng và cung cấp dữ liệu mở

Nguyên tắc tiếp cận và sử dụng dữ liệu mở

Theo Điều 25[27] của Hiến pháp năm 2013 và Điều 3 của Luật Tiếp cận thông tin 2016, quyền tiếp cận thông tin là một trong những quyền cơ bản của công dân. Theo Luật tiếp cận thông tin 2016, mọi công dân đều bình đẳng, không bị phân biệt đối xử trong việc thực hiện quyền tiếp cận thông tin, được tiếp cận với dữ liệu thông qua việc đọc, xem, nghe, ghi chép, sao chép, chụp thông tin một cách đầy đủ, chính xác và không phải trả phí, lệ phí, trừ trường hợp luật khác có quy định. Một số quyền tiếp cận và sử dụng DLM đã được trực tiếp đề cập trong khoản 3 Điều 43 Luật GDĐT 2023 và Điều 18 Nghị định 47/2020/NĐ-CP như sau:

- Quyền truy cập và sử dụng DLM mà không cần phải khai báo danh tính hay thực hiện bất kỳ thủ tục xác minh nào;
- Quyền tự do sao chép, chia sẻ, trao đổi, sử dụng DLM hoặc kết hợp DLM với dữ liệu khác để phát triển sản phẩm hoặc dịch vụ, dù là mục đích thương mại hay phi thương mại, trừ trường hợp có quy định khác của pháp luật.
- Cơ quan, tổ chức, cá nhân không được bán DLM đã được khai thác nguyên trạng từ CQNN cho tổ chức, cá nhân khác. Khi sử dụng DLM trong sản phẩm, dịch vụ thương mại của mình phải cung cấp miễn phí DLM kèm theo sản phẩm, dịch vụ thương mại đó.
- Cá nhân và tổ chức sử dụng cần trích dẫn, ghi nhận thông tin sử dụng DLM trong các sản phẩm, tài liệu liên quan có sử dụng DLM. CQNN không chịu trách nhiệm về bất kỳ thiệt hại nào phát sinh từ việc sử dụng DLM bởi cá nhân, tổ chức hay doanh nghiệp.

Trên thế giới có khái niệm "giấy phép DLM" (open data license) nhằm đặt ra các giới hạn bảo vệ quyền sở hữu DLM. Ngân hàng thế giới đưa ra định nghĩa về "giấy phép DLM" để giải thích rõ ràng các điều kiện để sử dụng DLM do Ngân hàng thế giới cung cấp. Trong Đạo luật DLM của Chính phủ Mỹ tại §3561 (11) nêu thuật ngữ "giấy phép mở" là bảo đảm pháp lý được áp dụng cho một tài sản dữ liệu công khai rằng tài sản dữ liệu đó được cung cấp: (A) không mất phí cho công chúng và (B) không có hạn chế nào về việc sao chép, xuất bản, phân phối, truyền tải, trích dẫn hoặc điều chỉnh. Một số ví dụ về giấy phép DLM tiêu chuẩn cho phép người dùng có thể tái sử dụng và cung cấp các điều khoản sử dụng nhất quán bao gồm Creative Commons, đặc biệt là CC-By và CC0, Giấy phép Chính phủ Mở của Vương quốc Anh (UK Open Government License), Điều khoản Sử dụng của Ngân hàng Thế giới (The World Bank Terms of Use), hay Giấy phép Ouverte của Chính phủ Pháp.

Những quy định hiện có của Việt Nam (được nêu ở trên) đã đặt ra một số yêu cầu và điều kiện sử dụng DLM, tương đương như các điều khoản trong giấy phép cung cấp DLM của các CQNN, tuy nhiên chưa trực tiếp giải thích khái niệm "giấy phép DLM". Mặc dù các quy định này đã góp phần cải thiện những hạn chế trong quy định về giấy phép tại Việt Nam, như được nêu trong Báo cáo chính sách về Chính phủ số và DLM năm 2019 liên quan đến bản quyền và quyền cấp phép dữ liệu[28], vẫn cần có những quy định cụ thể hơn để làm rõ và chuẩn hóa việc áp dụng các giấy phép DLM, đảm bảo tính minh bạch và khả năng thực thi trong thực tế.

[26] Đạo luật về thúc đẩy việc cung cấp và sử dụng dữ liệu công cộng: https://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=47133&type=part&key=4

[27] Điều 25 Hiến pháp năm 2013: "Công dân có quyền tự do ngôn luận, tự do báo chí, tiếp cận thông tin, hội họp, lập hội, biểu tình. Việc thực hiện các quyền này do pháp luật quy định."

[28] Ngân hàng Thế giới. Báo cáo Chính phủ Nước CHXNCN Việt Nam: "Đánh giá Mức độ sẵn sàng về Chính phủ Số và DLM". 2019.

Official Use Only

Tại Điều 17 của Luật Tiếp cận thông tin 2016 đã quy định 14 thông tin cụ thể phải được các CQNN công khai, trong đó, có 8 loại thông tin cụ thể CQNN bắt buộc phải công khai trên cổng thông tin điện tử, trang thông tin điện tử29. Tuy nhiên, hiện nay chưa có văn bản quy phạm pháp luật nào quy định phạm vi nội dung DLM hay loại DLM công khai cho công dân tiếp cận, khiến một số CQNN gặp khó khăn trong lựa chọn, xây dựng danh mục DLM. Trên thực tế, các trang thông tin điện tử của các bộ, cơ quan ở trung ương và các địa phương hiện đã thực hiện tốt việc cung cấp cho doanh nghiệp, tổ chức, cá nhân tiếp cận các thông tin về văn bản pháp luật, thủ tục hành chính; hướng dẫn thực hiện pháp luật; các chiến lược, kế hoạch phát triển kinh tế - xã hội, quy hoạch ngành; chức năng, nhiệm vụ, cơ cấu tổ chức và thông tin liên hệ của cơ quan. Tuy nhiên, doanh nghiệp, tổ chức, cá nhân chưa được tiếp cận nhiều thông tin thể hiện số liệu như báo cáo tài chính, thông tin thống kê trên các trang thông tin điện tử và các cổng DLM của các bộ, cơ quan ở trung ương và các địa phương.

Nguyên tắc cung cấp dữ liệu mở

Điều 17 của Nghị định 47/2020/NĐ-CP nêu tám nguyên tắc cung cấp DLM của CQNN30. Ngoài ra, tại Điều 20. Yêu cầu trong việc công bố DLM của CQNN, Nghị định 47/2020/NĐ-CP còn quy định:

(i) Không tiết lộ thông tin cá nhân; đánh giá và hạn chế nguy cơ, rủi ro tiết lộ thông tin cá nhân khi kết hợp với các thông tin khác;

(ii) Bảo đảm an toàn, an ninh khi sử dụng DLM bao gồm cả rủi ro do DLM gây ra hoặc rủi ro khi DLM kết hợp với dữ liệu khác gây ra.

Điều 2. Giải thích từ ngữ của Nghị định 13/2023/NĐ-CP của Chính phủ về Bảo vệ dữ liệu cá nhân (Nghị định 13/2023/NĐ-CP) đã giải thích dữ liệu cá nhân là *"thông tin dưới dạng ký hiệu, chữ viết, chữ số, hình ảnh, âm thanh hoặc dạng tương tự trên môi trường điện tử gắn liền với một con người cụ thể hoặc giúp xác định một con người cụ thể. Dữ liệu cá nhân bao gồm dữ liệu cá nhân cơ bản và dữ liệu cá nhân nhạy cảm"*.

Theo Điều 3. Nguyên tắc bảo vệ dữ liệu cá nhân của Nghị định 13/2023/NĐ-CP, dữ liệu cá nhân được xử lý theo quy định của pháp luật và được áp dụng các biện pháp bảo vệ, bảo mật trong quá trình xử lý. Bên Kiểm soát dữ liệu cá nhân, Bên Xử lý dữ liệu cá nhân chịu trách nhiệm tuân thủ các nguyên tắc xử lí dữ liệu. Như vậy, các CQNN khi xây dựng và cung cấp DLM cũng cần có bộ phận kiểm soát và xử lí dữ liệu cá nhân, chịu trách nhiệm xử lý dữ liệu để đảm bảo việc bảo vệ dữ liệu cá nhân.

Ngoài ra, Điều 5. Nguyên tắc chung về quản lý, kết nối và chia sẻ dữ liệu tại Nghị định 47/2020/NĐ-CP, quy định việc chia sẻ dữ liệu giữa các CQNN không được làm ảnh hưởng tới quyền lợi và trách nhiệm của tổ chức, cá nhân có liên quan, không được xâm phạm quyền về đời sống riêng tư, bí mật cá nhân, bí mật gia đình trừ trường hợp pháp luật có quy định khác. Tuy nhiên, những quy định này mới chỉ mang tính nguyên tắc chưa có hướng dẫn hoặc quy định cụ thể về việc tạo dữ liệu phi cá nhân hóa trước khi cung cấp DLM nhằm vừa đảm bảo cung cấp toàn vẹn và đầy đủ thông tin, vừa đảm bảo quyền riêng tư cá nhân. Do đó, cần có quy định, hướng dẫn những biện pháp để phi cá nhân hóa DLM để cân bằng giữa hai yếu tố "mở" và "bảo mật", đảm bảo DLM để tạo ra giá trị cho xã hội, trong khi vẫn tôn trọng các quyền của công dân.

[29] Điều 19. Công khai thông tin trên cổng thông tin điện tử, trang thông tin điện tử, Luật Tiếp cận thông tin 2016: (i) Văn bản pháp luật, điều ước quốc tế, thỏa thuận quốc tế, thủ tục hành chính, và quy trình công việc của CQNN; (ii) Hướng dẫn thực hiện pháp luật, chế độ, chính sách trong lĩnh vực quản lý của cơ quan; (iii) Chiến lược, dự án, kế hoạch phát triển kinh tế - xã hội, quy hoạch ngành và kết quả thực hiện; (iv) Danh mục dự án, chương trình, và kết quả thực hiện đầu tư công, mua sắm công, vốn vay; (v) Thông tin về chức năng, nhiệm vụ, cơ cấu tổ chức, và thông tin liên hệ của cơ quan; (vi) Báo cáo tài chính, thông tin thống kê, và kết quả chương trình khoa học; (vii) Danh mục thông tin công khai kèm địa chỉ, hình thức, thời gian công khai; (viii) Thông tin cần thiết vì lợi ích công cộng hoặc sức khỏe cộng đồng.

[30] (i) DLM phải toàn vẹn và phản ánh đầy đủ nội dung thông tin CQNN cung cấp; (ii) DLM được cung cấp là dữ liệu được cập nhật mới nhất; (iii) DLM phải có khả năng truy cập sử dụng được trên môi trường mạng;(iv) DLM phải đảm bảo khả năng thiết bị số có thể gửi, nhận, lưu trữ và xử lý được; (v) Tổ chức, cá nhân được tự do truy cập sử dụng DLM, không yêu cầu khai báo định danh khi khai thác, sử dụng DLM; (vi) DLM ở định dạng mở; (vii) Sử dụng DLM là miễn phí; (viii) Ưu tiên cung cấp DLM đối với dữ liệu mà người dân, doanh nghiệp có nhu cầu sử dụng cao.

Official Use Only

Hộp 2: Kĩ thuật ẩn danh dữ liệu nhằm bảo mật thông tin cá nhân, theo Quy định bảo vệ dữ liệu chung của châu Âu

Quy định bảo vệ dữ liệu chung 2016/679 của Châu Âu (GDPR) yêu cầu các bộ phận kiểm soát dữ liệu phải thực hiện biện pháp bảo vệ dữ liệu ngay từ khâu thiết kế và đảm bảo cài đặt quyền riêng tư ở mức cao nhất theo mặc định (Điều 25). Điều này bao gồm ẩn danh dữ liệu cá nhân, mã hóa và các biện pháp kĩ thuật khác để bảo đảm dữ liệu được bảo vệ xuyên suốt vòng đời xử lý. Trong đó, kĩ thuật ẩn danh dữ liệu có thể bao gồm việc:

- Mã hóa chuyển đổi dữ liệu gốc thành dạng không thể hiểu được và chỉ có thể giải mã bằng khóa phù hợp.
- Tokenisation thay thế dữ liệu nhạy cảm bằng mã thông báo không có giá trị bên ngoài nhưng vẫn cho phép xử lý và phân tích dữ liệu.

Ngoài ra, bút danh được khuyến nghị để giảm thiểu rủi ro và hỗ trợ việc tuân thủ nghĩa vụ bảo vệ dữ liệu của các bộ phận kiểm soát và xử lý (Điều 28).

1.3. Quy trình cung cấp dữ liệu mở của cơ quan nhà nước

Xây dựng kế hoạch cung cấp dữ liệu mở và danh mục dữ liệu mở

Việc xây dựng danh mục DLM được quy định tại khoản 2 Điều 17 Nghị định 47/2020/NĐ-CP, theo đó, Bộ trưởng, Thủ trưởng cơ quan ngang bộ, cơ quan thuộc Chính phủ, Chủ tịch Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương có trách nhiệm ban hành danh mục DLM và tổ chức xây dựng kế hoạch và triển khai công bố DLM. Danh mục DLM được rà soát, sửa đổi (nếu có) theo định kỳ, tối thiểu 6 tháng một lần; DLM đã công bố theo danh mục phải được cập nhật hoặc bổ sung (nếu có) trong thời hạn không quá 03 tháng kể từ thời điểm công bố (khoản 3 Điều 20 Nghị định 47/2020/NĐ-CP). Luật Dữ liệu năm 2024 yêu cầu CQNN có trách nhiệm công bố danh mục DLM và tổ chức công khai DLM để tổ chức, cá nhân khai thác, sử dụng, chia sẻ (Điều 21. Công khai dữ liệu).

Tại Điều 11. Cơ sở dữ liệu trong CQNN, Nghị định 47/2020/NĐ-CP, Danh mục CSDL của các bộ, ngành, địa phương cần bao gồm các thông tin như: *tên của các CSDL; mô tả về mục đích, phạm vi và nội dung của từng CSDL; cơ chế thu thập, cập nhật và nguồn gốc của dữ liệu được thu thập cho mỗi CSDL; và danh sách các hạng mục dữ liệu* (bao gồm DLM, dữ liệu chia sẻ theo hình thức mặc định và dữ liệu chia sẻ theo yêu cầu đặc thù), tùy thuộc vào từng đối tượng khai thác, sử dụng. Quy định này về cơ bản tương đồng với các quy định đã có trong khoản 3 Điều 19 Luật Tiếp cận thông tin 2016 yêu cầu các loại dữ liệu công khai phải được sắp theo theo danh mục; khoản 1 Điều 19 Luật Tiếp cận thông tin 2016 và điểm c khoản 4 Điều 40 Luật GDĐT 2023 về nội dung danh mục thông tin phải được công khai trong danh mục CSDL của Bộ, ngành, địa phương.

Ngoài các quy định này, hiện chưa có quy định mang tính pháp lý về xây dựng danh mục DLM khiến các CQNN còn nhiều lúng túng khi xây dựng danh mục DLM. Để hướng dẫn các bộ, cơ quan, địa phương, Bộ TT&TT đã ban hành Công văn số 5752/BTTT-CĐSQG ngày 26/11/2022[31], đề nghị các bộ, cơ quan, địa phương xây dựng danh mục DLM, triển khai kế hoạch cung cấp DLM theo mẫu gửi kèm theo công văn, đồng thời gửi kèm theo danh mục dữ liệu để tham khảo cung cấp dưới dạng DLM. Tuy nhiên, đây không phải văn bản quy phạm pháp luật nên không có hiệu lực bắt buộc thực thi thống nhất, đồng thời các hướng dẫn chưa đủ chi tiết và mang tính chất tham khảo.

Nhóm nghiên cứu ghi nhận một số ý kiến từ các cán bộ địa phương kiến nghị cần có sự thống nhất trong hướng dẫn xây dựng danh mục DLM. Có ý kiến cho rằng, dù Bộ TT&TT đã ban hành một số văn bản hành

[31] Công văn số 5752/BTTT-CĐSQG ngày 26/11/2022 của Bộ TT&TT gửi các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, Ủy ban nhân dân các tỉnh, thành phố trực thuộc Trung ương về việc đôn đốc triển khai Nghị định 47/2020/NĐ-CP, thúc đẩy kết nối, chia sẻ dữ liệu, cung cấp DLM.

Official Use Only

chính đôn đốc và yêu cầu xây dựng danh mục DLM yêu cầu các địa phương thực hiện, nhưng những hướng dẫn chia sẻ danh mục DLM của các cơ quan chưa được chuẩn hóa, dẫn đến sự không đồng bộ về tên danh mục, trường thông tin và hình thức chia sẻ dữ liệu.

> *"Bộ chưa ban hành một bộ danh mục dữ liệu chuẩn thống nhất trên toàn quốc nên mỗi địa phương có một danh mục khác nhau, không ai có thể xác nhận được nó có phải 1 loại không. Gần như không có cách định danh các danh mục dữ liệu đấy. Các địa phương lúng túng trong việc xây dựng danh mục dữ liệu dùng chung, từ đấy xem danh mục nào có thể mở ra." (Cán bộ Sở TT&TT cấp tỉnh #3)*
>
> *"Các đơn vị rất ngại khi công bố ra ngoài vì có thể bị quy trách nhiệm và ngại cả về tính chính xác do các đơn vị không biết dữ liệu có được công bố ra ngoài không, có vi phạm về bảo mật dữ liệu cá nhân không." (Cán bộ Sở TT&TT cấp tỉnh #4)*
>
> *"Để thúc đẩy quá trình cung cấp DLM, phải có quy định đưa vào luật. Bởi vì chữ "mở" vẫn chưa rõ ràng, không hình dung được nhu cầu của xã hội cần mở đến mức nào. Thứ hai, quy định nào cho phép mở hoặc mô hình nào đã thành công thì mình cũng không có, không rõ. Giờ vẫn trong quá trình nghiên cứu làm rõ những nội hàm về "mở"". (Cán bộ cấp Bộ #1)*
>
> *"Nghị định 47 đã có hẳn 1 chương về DLM nhưng quy định còn chung chung, các đơn vị khó khăn trong việc xác định các dữ liệu nào là có thể mở, được phép cung cấp ra ngoài. Ngay với cả TP.HCM cũng vừa mới ban hành danh mục DLM trong năm giữa năm nay với khoảng 91 đầu mục dữ liệu trong 13 lĩnh vực. Nhưng khi làm việc trực tiếp với đơn vị để cung cấp DLM ra ngoài thì các đơn vị gặp khó khăn trong việc xác định thông tin nào được công khai. Hiện tại các bộ ngành chưa có quy định về dữ liệu nào là được công bố rộng rãi. Nên các đơn vị khi cung cấp DLM, họ không cởi mở lắm về vấn đề đó." (Cán bộ cấp tỉnh #3)*

Tại thời điểm nghiên cứu, có **15/18** Bộ, cơ quan đã ban hành kế hoạch triển khai hoàn thiện các quy định liên quan và ban hành danh mục DLM[32]. Các bộ, cơ quan này đã ban hành Kế hoạch phát triển CSDL mở và danh mục dữ liệu cần triển khai, kèm với các thông tin cơ bản về đơn vị cung cấp, ngày cung cấp lần đầu, tần suất cung cấp; nguyên tắc thực hiện và phân công tổ chức thực hiện cho các đơn vị, các cục, vụ trực thuộc. Tuy nhiên, một số bộ mới chỉ có kế hoạch chia sẻ dữ liệu và xác định nhiệm vụ của các cơ quan trong hoạt động xây dựng danh mục tài liệu dùng chung như Bộ Nội vụ, Bộ Y tế, Bộ Giao thông Vận tải và Bộ Nông nghiệp và Phát triển nông thôn.

Đối với các địa phương, tại thời điểm nghiên cứu, 61/63 tỉnh thành đã xây dựng hoàn thiện Kế hoạch xây dựng CSDL nhằm đáp ứng Kế hoạch hoạt động của Ủy ban quốc gia về Chuyển đổi số năm 2023. Các văn bản đều xác định rõ việc xây dựng hệ thống DLM phải tuân theo các quy định của Nghị định số 47/2020/NĐ-CP. Đồng thời, dựa theo hướng dẫn của Công văn số 5752/BTTTT-CĐSQG ngày 26/11/2022 của Bộ TT&TT, đa số các tỉnh thành **(51/63)** đã thực hiện xây dựng đầy đủ Bảng Danh mục cơ sở DLM của thành phố theo các tiêu chí được hướng dẫn[33].

Danh mục dữ liệu mở

Dù đa số các địa phương đều có điểm chung ở một số lĩnh vực chính như Giáo dục, Giao thông Vận tải, Tài nguyên và Môi trường, Văn hóa Du lịch, hay Y tế, sức khỏe nhưng do tình hình thực tiễn của mỗi địa phương, các tỉnh thành vẫn quy định danh mục CSDL khác nhau. Một số địa phương đã có văn bản triển khai tương đối đầy đủ yêu cầu, song nhiều địa phương mới chỉ dừng lại ở việc xây dựng quy chế cho các hoạt động quản lý, kết nối và chia sẻ dữ liệu số hoặc có kế hoạch triển khai hệ thống CSDL như Thái Bình, Phú Thọ, Phú Yên,

[32] Bộ TT&TT, Bộ Tài nguyên và Môi trường, Bộ Tài chính, Bộ Công thương, Bộ Xây dựng, Bộ Tư pháp, Bộ Khoa học và công nghệ, Bộ Ngoại giao, Bộ Kế hoạch và Đầu tư, Bộ Văn hóa Thể thao và Du lịch, Bộ Lao động-Thương binh và Xã hội

[33] Các tiêu chí bao gồm: *(i) Tên tập dữ liệu: Tên dữ liệu phải được viết dễ hiểu cho nhiều đối tượng khác nhau, độ dài vừa đủ để mô tả được nội dung dữ liệu hàm chứa; (ii) Mô tả: Thông tin mô tả về mục đích, phạm vi, nội dung, cơ chế thu thập, cập nhật, nguồn dữ liệu được thu thập cũng như các thông tin khác về chất lượng, độ tin cậy của dữ liệu; (iii) Đơn vị chủ trì cung cấp: Cơ quan có thẩm quyền quản lý và cung cấp dữ liệu; (iv) Ngày cung cấp lần đầu: Ngày mà dữ liệu được cơ quan có thẩm quyền công bố rộng rãi cho cơ quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ; (v) Tần suất cung cấp: Tần suất mà dữ liệu được cung cấp. Ví dụ như hàng tuần, hàng tháng, hàng năm…; (vi) Ghi chú: Thông tin bổ sung nếu có.*

Official Use Only

Bến Tre. Một số địa phương khác như Ninh Thuận, Hải Dương, Điện Biên, Hòa Bình, Thanh Hóa, Hà Tĩnh, Đắk Nông, Bình Thuận, Bình Phước, Trà Vinh, Kiên Giang mới chỉ ban hành danh mục CSDL dùng chung chứa một số loại CSDL được ghi chú là cung cấp dưới dạng "mở" chứ chưa có danh mục cơ sở DLM chi tiết và đầy đủ như các địa phương khác.

Nhìn chung, danh mục DLM của nhiều địa phương còn sơ sài, số lượng CSDL ít, lĩnh vực không đa dạng (với lý do triển khai từng bước một do chưa có kinh nghiệm), danh mục DLM chưa được rà soát, cập nhật, sửa đổi định kì theo quy định.

Hộp 3: Kinh nghiệm quốc tế liên quan đến danh mục dữ liệu mở

Tại một số nước trên thế giới, gồm Pháp, Ba Lan, Estonia đã ban hành các danh mục DLM bắt buộc trong luật pháp quốc gia phù hợp với Chỉ thị về DLM của Liên minh Châu Âu và khái niệm về các bộ dữ liệu có giá trị cao.

- Tại Pháp, Luật về Nền Cộng hòa Kĩ thuật số (Loi pour une République numérique) được ban hành với mục tiêu thúc đẩy một xã hội kĩ thuật số cởi mở và toàn diện.[34] Luật yêu cầu công bố dữ liệu về kinh tế, xã hội, môi trường. Pháp đang triển khai sáu danh mục bộ dữ liệu giá trị cao được xác định trong Chỉ thị về DLM của Liên minh Châu Âu bao gồm: Dữ liệu không gian địa lý, Quan sát trái đất và môi trường, Khí tượng, Thống kê, Công ty và quyền sở hữu công ty và Giao thông[35]
- Tại Ba Lan, Luật ngày 11 tháng 8 năm 2021 về Dữ liệu Mở và Việc Tái sử dụng Thông tin Khu vực Công đã được ban hành đã chỉ ra các danh mục dữ liệu mới như dữ liệu động, dữ liệu nghiên cứu và bộ dữ liệu giá trị cao.[36] Các loại dữ liệu có sẵn trên cổng dữ liệu mở của Ba Lan bao gồm nông nghiệp, kinh tế, giáo dục, môi trường, chính phủ, y tế, v.v..
- Tại Estonia, Đạo luật Thông tin Công (Avaliku teabe seadus)[37] yêu cầu các bên nắm giữ thông tin phải tiết lộ nhiều loại thông tin hiện có, khoảng 40 loại - bao gồm thống kê kinh tế, thống kê tội phạm, quy chế chính phủ, mô tả công việc, cơ cấu cơ quan, thông tin môi trường, ngân sách, báo cáo nghiên cứu, v.v.

Xây dựng cổng dữ liệu mở

Điều 13 Nghị định 47/2020/NĐ-CP xác định "Cổng dữ liệu quốc gia" là *"đầu mối cung cấp DLM của các cơ quan nhà nước trên môi trường mạng nhằm tăng cường tính minh bạch trong hoạt động của Chính phủ và thúc đẩy sáng tạo, phát triển kinh tế, xã hội"*, đồng thời *"là đầu mối cung cấp hỗ trợ kĩ thuật, giám sát tình hình kết nối, chia sẻ dữ liệu giữa các hệ thống thông tin trong cơ quan nhà nước"*. CQNN có trách nhiệm theo dõi, đăng ký và cung cấp thông tin về dữ liệu, DLM; quản lý DLM; khả năng chia sẻ dữ liệu và nhu cầu khai thác dữ liệu trên Cổng DLM quốc gia. Ngoài việc cung cấp DLM trên Cổng DLM quốc gia, Nghị định số 42/2022/NĐ-CP yêu cầu các CQNN xây dựng các kênh khác nhằm cung cấp thông tin cho tổ chức, cá nhân trên môi trường mạng, trong đó có Cổng dữ liệu cấp bộ, cấp tỉnh[38].

Tuy nhiên, hiện tại chỉ có quy định về tính năng tích hợp và liên kết giữa các Cổng dữ liệu tại Nghị định 42/2022/NĐ-CP và yêu cầu đặc tả tập DLM cho cổng dữ liệu theo Tiêu chuẩn TCVN 14171:2024 mà chưa có các nguyên tắc, tiêu chuẩn về các chức năng, tính năng kĩ thuật khi xây dựng cổng DLM để các CQNN triển khai thực hiện. Cụ thể, điểm a khoản 5 Điều 2 Nghị định 42/2022/NĐ-CP quy định các cổng dữ liệu cấp bộ phải được tích hợp với Cổng dữ liệu quốc gia và là cổng thành phần của Cổng thông tin điện tử của bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ; còn Cổng dữ liệu cấp tỉnh phải tích hợp với Cổng dữ liệu quốc gia

[34] Đạo luật Cộng hòa Kỹ thuật số của Pháp: Các điều khoản chính và yêu cầu tuân thủ - Kiteworks, truy cập ngày 3 tháng 4 năm 2025, https://www.kiteworks.com/risk-compliance-glossary/french-digital-republic-act/.
[35] DLM và các bộ dữ liệu có giá trị cao: hướng dẫn truy cập, truy cập ngày 3 tháng 4 năm 2025, https://digital-strategy.ec.europa.eu/en/factpages/open-data-and-high-value-datasets-step-step-access-guide
[36] https://ruj.uj.edu.pl/bitstreams/493cfe7f-452c-46b6-9888-112bb63a2b6f/download
[37] https://www.kratid.ee/en/avaandmed
[38] Cổng dữ liệu cấp bộ, cấp tỉnh là "điểm đầu mối truy cập thông tin, dữ liệu trên môi trường mạng phục vụ việc công bố DLM" (khoản 3 Điều 3 Nghị định 42/2022/NĐ-CP)

và là cổng thành phần của Cổng thông tin điện tử của Ủy ban nhân dân tỉnh, thành phố trực thuộc trung ương và sử dụng tên miền quốc gia ".gov.vn" theo cấu trúc định sẵn, đảm bảo việc chia sẻ và liên kết thông tin (Điều 6 Nghị định 42/2022/NĐ-CP). Trên thực tế, tính năng liên kết và tích hợp này đã được quy định từ trước tại khoản 4 Điều 19 của Luật Tiếp cận thông tin 2016 nhằm đảm bảo tính liên kết, tích hợp với cổng thông tin điện tử, trang thông tin điện tử của cơ quan trực thuộc để cập nhật thông tin, tạo thuận lợi cho công dân trong việc tìm kiếm, khai thác thông tin. Ngoài ra, nhằm đảm bảo các trường thông tin được cung cấp đầy đủ lên cổng DLM, các cơ quan cần phải tham khảo định nghĩa "dữ liệu đặc tả"[39] tại khoản 11 Điều 2 của Nghị định 42/2022/NĐ-CP và các hướng dẫn khác như Công văn 1016/BTTTT-CĐSQG năm 2024 của Bộ TT&TT hướng dẫn xây dựng, áp dụng quy chuẩn, tiêu chuẩn, quy định kĩ thuật kết nối, chia sẻ dữ liệu và Tiêu chuẩn quốc gia số TCVN 14171:2024 đặc tả tập DLM cho cổng dữ liệu (DCAT-VN) do Bộ Khoa học và Công nghệ công bố.

> *"Bộ TT&TT chưa ban hành về chức năng kĩ thuật của cổng dữ liệu, trong khi có rất nhiều văn bản chỉ đạo triển khai. Nghị định 47/2020/NĐ-CP về quản lý, kết nối và chia sẻ dữ liệu số của CQNN yêu cầu phải công bố dữ liệu lên cả cổng dữ liệu quốc gia nhưng việc công bố lên cổng dữ liệu quốc gia lại chưa có hướng dẫn rõ ràng về hình thức, cách thức đăng tải, kết nối các danh mục dữ liệu đã được công bố." (Cán bộ Sở TT&TT cấp tỉnh #3)*

Tóm lại, quy định hiện nay về cổng DLM còn rải rác trong nhiều văn bản mà chưa có hướng dẫn cụ thể về tiêu chuẩn kĩ thuật để xây dựng cổng DLM cấp bộ, cấp tỉnh, dẫn đến tình trạng mỗi cổng thông tin của các địa phương được thiết kế và triển khai theo những cách thức khác nhau, gây khó khăn trong việc tạo dựng hệ thống đồng nhất trên phạm vi quốc gia *(Xem thêm Báo cáo đánh giá cổng DLM tại Việt Nam)*.

Quản trị dữ liệu mở, quản trị chia sẻ, khai thác dữ liệu mở

Điều 7, Nghị định 47/2020/NĐ-CP đã quy định yêu cầu trong việc quản lý, kết nối, chia sẻ dữ liệu. Theo đó, CQNN thực hiện kết nối, chia sẻ dữ liệu phải chỉ định một cán bộ phụ trách việc kết nối và chia sẻ dữ liệu; công khai thông tin liên hệ của cán bộ này để tiếp nhận, phối hợp, xử lý các vấn đề về kết nối, chia sẻ dữ liệu của cơ quan, đơn vị mình đối với cơ quan, đơn vị bên ngoài; công khai thông tin về đầu mối phụ trách kết nối, chia sẻ dữ liệu; các thông tin về sự sẵn sàng chia sẻ dữ liệu và các thông tin cần công bố. ***Hiện chưa có quy định riêng về việc quản lý, kết nối, chia sẻ DLM trong Nghị định 47/2020/NĐ-CP cũng như trong các văn bản quy phạm pháp luật khác, hoặc có thể hiểu các quy định trong Nghị định 47/2020/NĐ-CP cũng áp dụng với cả DLM***.

Về quy trình cung cấp DLM, chưa có văn bản bản quy định cụ thể thời điểm, quy trình đăng tải DLM mà cơ quan, địa phương tự quyết định. Một số văn bản đã quy định về việc thu thập, lưu trữ, kết nối, chia sẻ đến các tiêu chuẩn kĩ thuật của dữ liệu nói chung nhưng chưa có quy định riêng đối với DLM[40]. Ngoài ra, Bộ TT&TT, Văn phòng Chính phủ cũng đã có một số hướng dẫn liên quan đến quản trị dữ liệu, chia sẻ, khai thác dữ liệu nhưng đây đều là các văn bản hành chính, chưa được thể chế hóa thành văn bản quy phạm pháp luật và chưa có văn bản riêng cho DLM[41].

---

[39] Dữ liệu đặc tả là những thông tin mô tả các đặc tính của dữ liệu như nội dung, định dạng, chất lượng, điều kiện và các đặc tính khác nhằm tạo thuận lợi cho quá trình tìm kiếm, truy cập, quản lý và lưu trữ dữ liệu.

[40] Một số quy định pháp luật như: Nghị định 64/2007/NĐ-CP ngày 10/4/2007 của Chính phủ về Ứng dụng CNTT trong hoạt động của CQNN, bao gồm việc xây dựng, khai thác và duy trì CSDL. Nội dung nghị định nhấn mạnh vào việc bảo đảm truy cập thông tin và khai thác dịch vụ hành chính công, đóng vai trò hỗ trợ hoạt động hành chính công hiệu quả hơn; Thông tư số 13/2017/TT-BTTTT ngày 23/6/2017 của Bộ trưởng Bộ TT&TT quy định về việc cung cấp thông tin và bảo đảm khả năng truy cập thông tin trên trang thông tin điện tử hoặc cổng thông tin điện tử của cơ quan nhà nước; Thông tư 39/2017/TT-BTTTT ngày 15/12/2017 của Bộ trưởng Bộ TT&TT quy định về Danh mục tiêu chuẩn kĩ thuật về ứng dụng CNTT trong cơ quan nhà nước.

[41] Một số văn bản hướng dẫn của Bộ TT&TT, Văn phòng chính phủ như: Công văn số 3788/BTTTT-THH ngày 26/12/2014 của Bộ TT&TT hướng dẫn triển khai ứng dụng CNTT trong hoạt động của cơ quan nhà nước quy định về việc liên thông, trao đổi dữ liệu có cấu trúc bằng ngôn ngữ XML giữa các hệ thống thông tin trong cơ quan nhà nước; Công văn số 2051/BTTTT-THH ngày 04/06/2020 của Bộ TT&TT hướng dẫn triển khai một số nhiệm vụ thực thi Nghị định số 47/2020/NĐ-CP; Công văn số 677/BTTTT-THH ngày 3/3/2022 của

Như vậy, trong khi các quy định pháp luật về quản trị DLM, quản trị chia sẻ DLM và các nội dung về kĩ thuật liên quan đến DLM còn thiếu, đồng thời lại có nhiều hướng dẫn hành chính về dữ liệu nói chung có thể gây khó khăn cho các CQNN trong quá trình triển khai thực hiện. Các văn bản hiện bao gồm cả nghị định, thông tư và nhiều công văn hướng dẫn hành chính, trong đó mỗi văn bản tập trung vào những khía cạnh khác nhau như tiêu chuẩn kĩ thuật, kết nối CSDL, mô hình trao đổi dữ liệu và tích hợp nền tảng chia sẻ. Sự chồng chéo này không chỉ làm tăng áp lực hành chính mà còn tiềm ẩn nguy cơ xảy ra sai sót, ảnh hưởng đến hiệu quả quản lý và sử dụng DLM. Việc xây dựng một văn bản pháp luật riêng, đầy đủ và toàn diện về DLM do vậy là hết sức cần thiết.

1.4. Trách nhiệm phối hợp và triển khai thực hiện của các cơ quan, đơn vị

Theo Luật Dữ liệu năm 2024, có hai chủ thể chính liên quan đến dữ liệu được nêu ngay tại Điều 3. Giải thích thuật ngữ là "*chủ quản dữ liệu*", "*chủ sở hữu dữ liệu*"[42], các điều luật khác quy định các nội dung theo hai chủ thể này. Theo Nghị định 47/2020/NĐ-CP, có rất nhiều chủ thể được quy định xuyên suốt trong toàn bộ văn bản gồm: "*cơ quan nhà nước quản lý dữ liệu*" (Điều 6), "*cơ quan nhà nước thực hiện kết nối, chia sẻ dữ liệu* (Điều 7)", "*đơn vị chuyên trách công nghệ thông tin của các các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, các tỉnh, thành phố trực thuộc trung ương*" (Điều 11, Điều 49), "*chủ quản cơ sở dữ liệu*" (Điều 12, 16, 24, 40, 55), "*cơ quan cung cấp các dịch vụ chia sẻ dữ liệu*" (Điều 25), "*cơ quan cung cấp và sử dụng dữ liệu*" (Điều 27), "*cơ quan cung cấp dữ liệu*" (quy định tại 17 điều luật), "*chủ quản hệ thống thông tin*"[43], "*cơ quan khai thác dữ liệu*", v.v…

Để thực hiện các hoạt động liên quan đến DLM, cần có sự tham gia của nhiều bên liên quan, đặc biệt là các CQNN từ các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cho đến Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương tuy nhiên Nghị định 47/2020/NQ-CP nêu quá nhiều chủ thể mà không có định nghĩa khiến cho các quy định trở nên không rõ ràng đối với từng chủ thể. Các vai trò, trách nhiệm chung trong các hoạt động như xây dựng và triển khai kế hoạch cung cấp DLM, thực hiện cung cấp DLM, nâng cao chuyên môn và nhận thức của cán bộ, đảm bảo nguồn kinh phí hoạt động và theo dõi giám sát đã được quy định trong Nghị định 47/2020/NĐ-CP. Song như đã nêu, do các chủ thể không thống nhất và phần lớn các quy định là đối với việc chia sẻ dữ liệu số của CQNN nói chung, có những quy định không phù hợp với DLM nên các quy định trong Nghị định 47/2020/NĐ-CP khó áp dụng đối với DLM.

Trách nhiệm xử lý và cung cấp, chia sẻ dữ liệu mở

Các văn bản pháp luật hiện hành chưa có quy định cụ thể về sự tham gia, vai trò, trách nhiệm của các bên liên quan trong quy trình cung cấp DLM từ việc thu thập, xử lí dữ liệu cho đến chia sẻ DLM lên môi trường mạng. DLM hiện nay theo các quy định hiện hành là DLM của CQNN (các cơ quan, đơn vị thuộc Bộ, cơ quan ngang bộ ở cấp trung ương và các sở, ban, ngành ở cấp địa phương). Các cơ quan, đơn vị này có thể được hiểu là các "*cơ quan chủ quản dữ liệu/cơ sở dữ liệu*" (thuật ngữ được sử dụng trong Nghị định 47/2020/NĐ-CP), là cơ quan cung cấp DLM cho "*cơ quan cung cấp dữ liệu*" (có thể là cơ quan/đơn vị chuyên trách CNTT) để thực hiện chia sẻ lên môi trường mạng (qua cổng DLM). Nghị định 47/2020/NĐ-CP hiện quy định về việc chuẩn bị và chia sẻ dữ liệu nói chung, có thể hiểu bao gồm DLM, song cũng chưa có quy định riêng về trách

---

Bộ TT&TT hướng dẫn kết nối và chia sẻ dữ liệu thông qua Nền tảng tích hợp, chia sẻ dữ liệu quốc gia và hướng dẫn kết nối nền tảng chia sẻ dữ liệu phân tán; Công văn số 1552/BTTTT-THH ngày 26/4/2022 hướng dẫn kĩ thuật triển khai Đề án 06 (phiên bản 1.0) chỉ dẫn các thành phần CNTT thiết yếu phục vụ triển khai Đề án 06, hướng dẫn tiêu chuẩn, quy chuẩn, hướng dẫn kĩ thuật định mức phục vụ triển khai Đề án 06 và quy định đầu mối phối hợp, hỗ trợ, hướng dẫn triển khai; Văn bản số 761/VPCP-KSTT ngày 29/01/2022 của Văn phòng Chính phủ về việc hướng dẫn kết nối, tích hợp, chia sẻ với CSDLQG về dân cư hướng dẫn kết nối, tích hợp chia sẻ dữ liệu dân cư giữa CSDLQG về dân cư với Hệ thống thông tin giải quyết thủ tục hành chính cấp bộ, cấp tỉnh; Công văn 1016/BTTTT-CĐSQG ngày 22/03/2024 của Bộ TT&TT hướng dẫn xây dựng, áp dụng quy chuẩn, tiêu chuẩn, quy định kĩ thuật kết nối, chia sẻ dữ liệu.

[42] Theo Điều 3. Giải thích từ ngữ, Luật Dữ liệu năm 2024, "*Chủ quản dữ liệu*" là cơ quan, tổ chức, cá nhân thực hiện hoạt động xây dựng, quản lý, vận hành, khai thác dữ liệu theo yêu cầu của chủ sở hữu dữ liệu (Khoản 13); "*Chủ sở hữu dữ liệu*" là cơ quan, tổ chức, cá nhân có quyền quyết định việc xây dựng, phát triển, bảo vệ, quản trị, xử lý, sử dụng và trao đổi giá trị của dữ liệu do mình sở hữu (Khoản 14).

[43] Nêu tại Khoản 1. Điều 28. Bảo đảm an toàn trong quá trình kết nối, chia sẻ dữ liệu.

nhiệm của "*CQNN chủ quản CSDL*" (các cơ quan/đơn vị thuộc bộ ở trung ương hoặc các sở ban ngành ở địa phương) trong việc này, mà chỉ có quy định về trách nhiệm của "*cơ quan cung cấp dữ liệu*" (được hiểu là cơ quan/đơn vị chuyên trách CNTT hoặc cơ quan/đơn vị quản lý cổng cung cấp DLM) trong việc "*xác định, chuẩn hóa dữ liệu*"[44].

Quy trình các bước chuẩn bị dữ liệu cho đến kết nối, chia sẻ DLM chưa được quy định chi tiết trách nhiệm của các bên liên quan, cụ thể là giữa cơ quan chủ quản dữ liệu và cơ quan cung cấp dữ liệu. Ví dụ, cơ quan/đơn vị nào có trách nhiệm rà soát, xử lý dữ liệu cá nhân, đảm bảo DLM khi cung cấp công khai không còn chứa thông tin cá nhân chưa được quy định cụ thể dẫn đến việc thực hiện ở các bộ, cơ quan, các địa phương thực hiện khác nhau. Nếu theo quy định hiện nay thì cơ quan chủ quản dữ liệu không có trách nhiệm xử lý dữ liệu trước khi cung cấp cho cơ quan cung cấp DLM. Khi phỏng vấn một số CQNN ở địa phương, nhóm nghiên cứu nhận thấy việc chuẩn hóa dữ liệu có địa phương phân công cơ quan chủ quản CSDL thực hiện, có địa phương phân công cho cơ quan quản lý cổng DLM thực hiện xử lý, chuẩn hóa DLM trước khi chia sẻ. Việc đăng tải DLM lên cổng DLM của địa phương cũng được thực hiện khác nhau, có địa phương cho phép cơ quan chủ quản CSDL được quyền chia sẻ DLM lên cổng DLM, có địa phương chỉ cơ quan quản lý cổng DLM mới được chia sẻ dữ liệu. Do đó, cần có quy định rõ ràng về quy trình quản lý, chia sẻ DLM và trách nhiệm của các cơ quan, đơn vị liên quan để đảm bảo thực hiện thống nhất, đồng bộ.

> *"Hiện tại, khi các đơn vị khi đưa dữ liệu về, họ đưa dữ liệu vẫn còn sơ khai, tổng hợp bằng file PDF, DOCX. Các đơn vị cung cấp chưa có CSDL để thống nhất với kho dữ liệu chung của thành phố, để từ đó đưa lên cổng dữ liệu mở. Quy trình tiếp nhận, xử lý và công khai dữ liệu bao gồm: 1) Người dùng/ tổ chức đăng nhập vào tài khoản trên cổng dữ liệu, đăng tải dữ liệu muốn chia sẻ, 2) Quản trị cổng tiếp nhận dữ liệu từ hệ thống, kiểm tra thông tin; rà soát dữ liệu trên tiêu chí đầy đủ thông tin theo yêu cầu, dữ liệu đảm bảo tính hợp lệ, hợp pháp và không vi phạm quy định, 3) Dữ liệu đã được phê duyệt sẽ hiển thị công khai trên cổng. Dữ liệu công khai được truy cập, khai thác, tìm kiếm bởi mọi người dùng, 4) Dữ liệu được lưu trữ trên hệ thống để phục vụ việc truy xuất, theo dõi và báo cáo định kỳ. Các thông tin liên quan đến quá trình xử lý dữ liệu đều được hệ thống ghi nhận lại." (Cán bộ Sở TT&TT #2)*
>
> *"Công tác hướng dẫn khai thác cổng DLM chưa được triển khai đồng bộ, dẫn đến tổ chức không hiểu rõ cách cung cấp dữ liệu đúng yêu cầu." (Cán bộ Sở TT&TT #4)*
>
> *"Hiện tại, một số đơn vị khó khăn trong việc xác định DLM vì đơn vị cung cấp không chuyên về CNTT, họ chỉ lấy những dữ liệu họ đang khai thác, thao tác hàng ngày trong công việc, thì họ xác định đó là dữ liệu có thể cung cấp được. Họ tổng hợp bằng file XLSX, PDF hoặc DOCX, cung cấp cho Trung tâm Chuyển đổi số để đưa lên cổng DLM." (Cán bộ Sở TT&TT #5)*
>
> *"Cơ quan/đơn vị cung cấp dữ liệu với định dạng không đồng nhất hoặc không nằm trong các định dạng được hỗ trợ (ví dụ: các file không chuẩn, dữ liệu không có cấu trúc). Nhiều dữ liệu thiếu thông tin mô tả, chẳng hạn như nguồn gốc dữ liệu, kỳ hạn cập nhật, thời gian dữ liệu hoặc không tuân thủ theo quy định của Danh mục DLM theo Quyết định Số 962 của UBND thành phố[45] đã ban hành. Một số tổ chức cung cấp dữ liệu chứa thông tin cá nhân mà chưa được xử lý, rủi ro cao về quyền riêng tư". (Cán bộ Sở TT&TT cấp tỉnh #4)*
>
> *"Nếu muốn kết nối vào kho dữ liệu dùng chung của thành phố thì họ phải xây dựng API, thêm chức năng thì họ rất ngại. Tại vì phần mềm của họ đang ổn định, giờ mình kết nối với họ thì để đảm bảo an toàn thông tin hay gì đó thì họ chưa sẵn sàng kết nối. Thực tế, đến thời điểm hiện tại chỉ có Sở Kế hoạch và Đầu tư là kết nối với kho dữ liệu dùng chung của thành phố, cung cấp DLM thôi. Còn những đơn vị khác trong quá trình khảo sát mới đây thì bước đầu họ cung cấp dữ liệu dạng XLSX trước thôi,*

[44] Theo đó, cơ quan cung cấp dữ liệu thực hiện các nội dung sau để chuẩn bị và thực hiện cung cấp dữ liệu (Điều 33, Nghị định 47/2020/NĐ-CP): Xác định, chuẩn hóa dữ liệu; Xác định danh mục dữ liệu chia sẻ; Phát triển triển dịch vụ chia sẻ dữ liệu, các tài liệu hướng dẫn kết nối khai thác dữ liệu; Công khai các dịch vụ chia sẻ dữ liệu; Tiếp nhận các yêu cầu kết nối sử dụng các dịch vụ chia sẻ dữ liệu, khai thác dữ liệu; Thực hiện việc kết nối, chia sẻ dữ liệu.

[45] Quyết định 962 QĐ-UBND Quyết định ban hành danh mục dữ liệu mở thành phố Đà Nẵng . Link: https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Quyet-dinh-962-QD-UBND-2023-Danh-muc-du-lieu-mo-thanh-pho-Da-Nang-566456.aspx

chưa sẵn sàng kết nối server với server về trung tâm dữ liệu thành phố." (Cán bộ Sở TT&TT cấp tỉnh #5)

Trách nhiệm theo dõi, giám sát việc triển khai cung cấp dữ liệu mở

Trong các nội dung quản lý nhà nước về dữ liệu quy định tại Điều 8 Luật Dữ liệu năm 2024 có nội dung "*quản lý, giám sát các hoạt động xây dựng, phát triển, bảo vệ, quản trị, xử lý, sử dụng dữ liệu, bảo đảm an ninh, an toàn dữ liệu*" (điểm c); "*Thanh tra, kiểm tra, giải quyết khiếu nại, tố cáo và xử lý vi phạm pháp luật về dữ liệu*" (điểm đ). Tại Điều 12 Luật Dữ liệu năm 2024 quy định "*cơ quan nhà nước quản lý cơ sở dữ liệu*" có trách nhiệm "*thường xuyên kiểm tra, giám sát, khắc phục sai sót; thực hiện đồng bộ dữ liệu trong phạm vi cơ quan và phối hợp với các cơ quan, tổ chức có liên quan để thường xuyên cập nhật, điều chỉnh, bảo đảm chất lượng dữ liệu trong khai thác, sử dụng*" (điểm b, Khoản 2, Điều 12. Bảo đảm chất lượng dữ liệu, Luật Dữ liệu).

Điều 16, Nghị định 47/2020/NĐ-CP quy định về việc kiểm tra, đánh giá, duy trì dữ liệu do CQNN chia sẻ, có thể hiểu là áp dụng cả đối với DLM. Theo đó, dữ liệu phải được kiểm tra, đánh giá định kì hàng năm[46]. **Chủ quản CSDL** có trách nhiệm tự tổ chức kiểm tra đánh giá định kỳ dữ liệu và gửi báo cáo về **đơn vị chuyên trách về CNTT**, còn các đơn vị chuyên trách về CNTT có trách nhiệm hướng dẫn chủ quản CSDL tự kiểm tra, đánh giá dữ liệu. Như vậy, trong trường hợp DLM, khi cơ quan chủ quản CSDL cung cấp DLM cho đơn vị chuyên trách về CNTT để chia sẻ cho tổ chức, cá nhân sử dụng trên cổng DLM thì việc cơ quan chủ quản CSDL có trách nhiệm kiểm tra, đánh giá định kỳ theo quy định tại Điều 16 sẽ không phù hợp (đơn vị chuyên trách về CNTT/đơn vị quản lý cổng DLM thực hiện trách nhiệm hoặc/và phối hợp thực hiện sẽ phù hợp hơn). Điều này đòi hỏi cần có quy định riêng về trách nhiệm kiểm tra, đánh giá việc cung cấp, chia sẻ DLM của các cơ quan, đơn vị liên quan.

Điểm tích cực là trong quá trình triển khai cung cấp DLM, Nghị định 47/2020/NĐ-CP có quy định CQNN trong phạm vi quản lý của mình có trách nhiệm thu thập thông tin phản hồi của người sử dụng DLM (Điều 19. Kế hoạch và triển khai dữ liệu mở của các cơ quan nhà nước)[47]. Mặc dù vậy, theo kết quả rà soát, đánh giá cổng (xem thêm tại *Báo cáo đánh giá cổng DLM tại Việt Nam*), nhiều cổng DLM hiện chưa có cơ chế thu thập ý kiến phản hồi của người sử dụng.

Ngoài ra, trong Nghị định 47/2020/NĐ-CP có quy định về giải quyết vướng mắc kết nối, chia sẻ dữ liệu (Mục 5, Chương III. Thực hiện kết nối, chia sẻ, sử dụng dữ liệu)48. Điều 52, Nghị định 47/2020/NĐ-CP quy định tổ chức, cá nhân khai thác sử dụng dữ liệu (dữ liệu nói chung, có thể hiểu bao gồm DLM) có quyền "*Đề nghị cơ quan có thẩm quyền giải quyết các vướng mắc làm ảnh hưởng đến quyền khai thác, sử dụng dữ liệu chia sẻ của mình*" (điểm b, khoản 1) và có trách nhiệm "*Thông báo kịp thời cho cơ quan cung cấp dữ liệu về những sai sót của dữ liệu đã chia sẻ*" (điểm c, khoản 2), "*Phối hợp với cơ quan có thẩm quyền, cơ quan cung cấp dữ liệu để giải quyết các vướng mắc trong quá trình chia sẻ, sử dụng dữ liệu*" (điểm d, khoản 2). Cơ quan cung cấp dữ liệu có trách nhiệm tiếp nhận, giải quyết các vướng mắc phát sinh khi kết nối, sử dụng dữ liệu do mình cung cấp (Điều 48, Nghị định 47/2020/NĐ-CP).

Tuy nhiên, Điều 49 của Nghị định 47/2020/NĐ-CP quy định về trình tự xử lý vướng mắc lại chỉ quy định trình tự xử lý đối với các trường hợp vướng mắc của các cơ quan khai thác dữ liệu, hoặc vướng mắc giữa các bộ, ngành, địa phương mà không quy định đối với các vướng mắc của người sử dụng cá nhân, như vậy có thể

[46] Bao gồm: kiểm kê các hạng mục; đánh giá việc tuân thủ các tiêu chuẩn, quy chuẩn, quy định về dữ liệu; đánh giá chất lượng dữ liệu, đánh giá về duy trì, vận hành, cập nhật dữ liệu, nhật ký cập nhật, khai thác của dữ liệu; Đánh giá về chia sẻ dữ liệu, đối tượng sử dụng, mục đích chia sẻ phù hợp với các quy định, quy chế khai thác, sử dụng dữ liệu chia sẻ, v.v…

[47] Theo đó, trong phạm vi quản lý của mình, bộ, cơ quan, địa phương có trách nhiệm: Xác định cơ chế thu thập, phân tích thông tin phản hồi của cá nhân, tổ chức về việc sử dụng DLM; Chỉ định đầu mối liên hệ tiếp nhận thông tin phản hồi của tổ chức, cá nhân về các vấn đề liên quan đến chất lượng, tính khả dụng, định dạng, sự tuân thủ các quy định pháp luật liên quan đến DLM; Tiếp nhận ý kiến phản hồi của tổ chức, cá nhân để xác định mức độ ưu tiên và thực hiện công bố DLM phù hợp với nhu cầu; Hỗ trợ các tổ chức, cá nhân trong xã hội, cộng đồng thực hiện các hoạt động xây dựng, đóng góp để mở rộng DLM.

[48] Điều 47 nêu một số vướng mắc khi kết nối, chia sẻ và khai thác dữ liệu, có thể hiểu áp dụng cho dữ liệu nói chung, bao gồm DLM như: Vướng mắc về quyền và nghĩa vụ chia sẻ, khai thác dữ liệu; Vướng mắc về hạ tầng kĩ thuật triển khai kết nối, chia sẻ; Vướng mắc về việc áp dụng, tuân thủ tiêu chuẩn, quy chuẩn, quy định về dữ liệu; Vướng mắc về chất lượng dữ liệu chia sẻ, sử dụng dữ liệu sau khi chia sẻ; Vướng mắc về phương án kết nối, chia sẻ dữ liệu…

Official Use Only

hiểu quy định của Điều 49 chỉ áp dụng đối với dữ liệu dùng chung của các CQNN, theo đó, **chưa có quy định về xử lí vướng mắc đối với DLM.**

Về tổng thể, trách nhiệm theo dõi, giám sát chính công tác kết nối, chia sẻ dữ liệu tại Việt Nam (bao gồm DLM) được giao cho Bộ TT&TT tại khoản 1 Điều 15 và khoản 1 Điều 53 Nghị định 47/2020/NĐ-CP. Cụ thể, Bộ TT&TT có trách nhiệm giám sát việc quản trị dữ liệu, chia sẻ và khai thác dữ liệu và thu thập, đánh giá kết nối chia sẻ dữ liệu và yêu cầu chuyển đổi hình thức chia sẻ; triển khai giải pháp, tiện ích hỗ trợ chia sẻ dữ liệu và thủ tục hành chính trực tuyến. Bộ trưởng, Thủ trưởng cơ quan ngang bộ, cơ quan thuộc Chính phủ có trách nhiệm chỉ đạo, thanh tra, kiểm tra quá trình kết nối, chia sẻ dữ liệu tại bộ, ngành mình quản lý; Chủ tịch Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương chịu trách nhiệm chỉ đạo, thanh tra, kiểm tra hoạt động quản lý, kết nối, chia sẻ dữ liệu ở địa phương.

Đối với việc đảm bảo bảo mật quyền riêng tư và an ninh mạng, Bộ Công an có trách nhiệm thực hiện biện pháp bảo đảm an ninh mạng trong kết nối, chia sẻ dữ liệu số. Trường hợp phát hiện xảy ra vi phạm quy định bảo vệ dữ liệu cá nhân, Điều 23 Nghị định 13/2023/NĐ-CP quy định bên Kiểm soát dữ liệu cá nhân, Bên Kiểm soát và xử lý dữ liệu cá nhân thông báo cho Bộ Công an (Cục An ninh mạng và phòng, chống tội phạm sử dụng công nghệ cao) chậm nhất 72 giờ sau khi xảy ra hành vi vi phạm. Nếu xác định vi phạm trong việc cung cấp thông tin cá nhân, cá nhân bị xâm phạm quyền riêng tư có quyền yêu cầu bồi thường thiệt hại theo quy định của pháp luật, đồng thời các cá nhân, tổ chức có hành vi vi phạm bảo mật dữ liệu cá nhân sẽ phải chịu trách nhiệm xử lý kỷ luật, xử phạt vi phạm hành chính, xử lý hình sự theo quy định đối với các cán bộ CQNN (Điều 4 Nghị định 13/2023/NĐ-CP).

2. Nhu cầu khai thác, sử dụng dữ liệu mở

2.1. Mục đích sử dụng dữ liệu mở

DLM có vai trò quan trọng trong việc thúc đẩy phát triển kinh tế - xã hội, là mũi nhọn chiến lược để phát triển chính phủ số và xã hội số. Luật GDĐT 2023 nêu rõ việc công bố DLM cho các cơ quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ *"nhằm thúc đẩy giao dịch điện tử, chuyển đổi số, phát triển kinh tế số và xã hội số"*. Các Chiến lược phát triển kinh tế số, xã hội số, Chiến lược phát triển Chính phủ điện tử, Chương trình Chuyển đổi số quốc gia khuyến khích các CQNN mở dữ liệu để phục vụ cho sự phát triển chung nhằm: i) minh bạch hóa và tăng cường hoạt động trao đổi, chia sẻ dữ liệu; ii) phục vụ chuyển đổi số quốc gia; iii) và đáp ứng nhu cầu sử dụng dữ liệu của công dân tạo ra các sản phẩm kinh doanh, dịch vụ.

Trên thực tế, nhu cầu tiếp cận với DLM của các TCDN và người dùng cá nhân nhằm phục vụ cho mục đích kinh doanh, sản xuất, phân tích thị trường, nghiên cứu sản phẩm mới, được phát sinh thường xuyên. Trong số 280 CQNN tham gia khảo sát, có 16,8% đơn vị thường xuyên nhận được yêu cầu cung cấp dữ liệu, 58,2% đơn vị thỉnh thoảng nhận được yêu cầu cung cấp dữ liệu. Trong đó, các CQNN cấp trung ương thường xuyên nhận được yêu cầu cung cấp dữ liệu hơn so với cấp địa phương, cho thấy rằng nhu cầu về các dữ liệu vĩ mô ở cấp độ tổng thể, toàn quốc khá lớn.

Kết quả khảo sát nhu cầu sử dụng, khai thác DLM của nghiên cứu này cho thấy 76,9% TCDN có nhu cầu khai thác và sử dụng dữ liệu nhằm phục vụ hoạt động của mình, 80% người dùng cá nhân (đa số trong khu vực nghiên cứu, đại học) có nhu cầu sử dụng dữ liệu thường xuyên. Mục đích sử dụng dữ liệu của TCDN, cá nhân rất đa dạng, gồm các mục đích chính như 1) Nghiên cứu, giảng dạy; 2) Xây dựng các báo cáo nghiên cứu và dự báo thị trường; 3) Xây dựng báo cáo phục vụ sản xuất - kinh doanh, hoạt động của tổ chức; 4) Tạo ra các sản phẩm, dịch vụ mới; 5) Xây dựng các báo cáo phục vụ quản lý nhà nước. Sự đa dạng trong mục đích sử dụng dữ liệu cho thấy tiềm năng to lớn của dữ liệu trong việc thúc đẩy sự phát triển và đổi mới của các TCDN cũng như của các cá nhân. (*Xem thêm Báo cáo Khảo sát về nhu cầu sử dụng DLM tại Việt Nam*).

Official Use Only

Bảng 1: Mục đích sử dụng dữ liệu của các TCDN và người dùng cá nhân

	TCDN (n=250)			Người dùng cá nhân (n=91)	Tổng (n=341)
	Viện NC/ Trường ĐH (n=167)	Doanh nghiệp (n=52)	Tổ chức khác (n=31)		
Xây dựng các báo cáo nghiên cứu và dự báo thị trường	58 (34,7%)	30 (57,7%)	16 (51,6%)	37(40,7%)	104 (41,6%)
Xây dựng báo cáo phục vụ sản xuất - kinh doanh hoặc hoạt động của tổ chức	33 (19,8%)	29 (55,8%)	15 (48,4%)	26(28,6%)	77 (30,8%)
Tạo ra các sản phẩm, dịch vụ mới	35 (21%)	29 (55,8%)	6 (19,4%)	17(18,7%)	70 (28%)
Nghiên cứu, giảng dạy	145 (86,8%)	20 (38,5%)	11 (35,5%)	67(73,6%)	176 (70,4%)
Xây dựng các báo cáo phục vụ quản lý nhà nước	-	-	-	16(17,6%)	16(17,6%)
Khác	1 (0,6%)	3 (5,8%)	2 (6,5%)	2(2,2%)	6 (2,4%)

2.2. Nhu cầu về định dạng dữ liệu mở

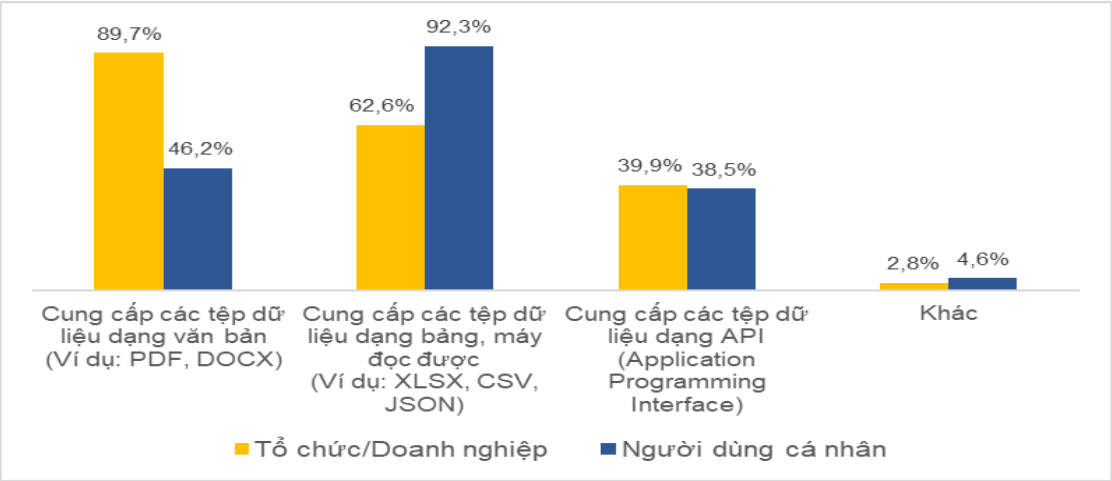
Kết quả cho thấy TCDN, người dùng cá nhân có nhu cầu tiếp cận với DLM ở nhiều định dạng khác nhau ở cả dạng văn bản (PDF, DOCX) và dạng máy đọc tệp dữ liệu máy đọc được (XLSX, CSV, JSON). Trong đó, có 89,7% TCDN ưu tiên sử dụng dữ liệu ở dạng văn bản (PDF, DOCX). Tuy nhiên, việc sử dụng dữ liệu ở định dạng máy không đọc được sẽ gây rất nhiều khó khăn đối với những công việc yêu cầu làm việc chuyên sâu với dữ liệu (ví dụ: sử dụng dữ liệu lớn để phân tích, nghiên cứu và dự báo thị trường…). Ở nhóm người dùng cá nhân, nhu cầu sử dụng dữ liệu dạng bảng, máy đọc được (XLSX, CSV, JSON) cao hơn đáng kể với trên 90% người trả lời muốn được tiếp cận DLM ở các định dạng này. Đây là những định dạng đáp ứng tốt nhu cầu phân tích, xử lý và tái sử dụng dữ liệu, cho thấy khả năng khai thác dữ liệu của người dùng cá nhân là đáng kể hơn. Sự khác biệt giữa hai nhóm đối tượng được khảo sát về định dạng DLM phù hợp với nhận thức của hai nhóm đối tượng này về DLM, theo đó, đa số người dùng cá nhân cho rằng DLM có định dạng máy đọc được, còn đa số TCDN cho rằng ngoài các định dạng XLSX, CSV, DLM có thể ở các định dạng PDF, DOCX.

Đối với dữ liệu cung cấp qua API, nhu cầu từ cả hai nhóm tương đối thấp, với 39,9% TCDN và 38,5% người dùng cá nhân quan tâm. Để sử dụng hiệu quả API, người dùng cần có kĩ năng và phần mềm để kết nối vì vậy sử dụng tệp dữ liệu sử dụng API còn chưa phổ biến.

Như vậy, để đáp ứng tốt nhu cầu của các nhóm người dùng, các cổng cung cấp DLM cần đa dạng hóa định dạng dữ liệu, ưu tiên các định dạng có khả năng phân tích như XLSX, CSV do nhiều người dùng có nhu cầu hơn và hướng dẫn sử dụng API hiệu quả hơn nhằm thúc đẩy người dùng sử dụng các tệp dữ liệu dạng API.

Official Use Only

Hình 1: Nhu cầu về định dạng DLM của các TCDN và người dùng cá nhân (n=420)



**Dựa trên kết quả khảo sát của nhóm Đánh giá về nhu cầu sử dụng DLM tại Việt Nam*

2.3. Nhu cầu đối với các tính năng của cổng dữ liệu mở

Đối với cổng DLM, các tính năng mà TCDN và người dùng cá nhân mong muốn trên cổng DLM chủ yếu tập trung vào khả năng truy cập và sử dụng dữ liệu một cách dễ dàng, hiệu quả. Các tính năng được đánh giá là quan trọng bao gồm: "*Mô tả chi tiết dữ liệu*", "*Giao diện trực quan, dễ dàng sử dụng*", "*Công cụ tìm kiếm nâng cao*", "*Cung cấp API*", "*Liên kết với Cổng DLM quốc gia*", "*Phản hồi (đánh giá, nhận xét, gửi yêu cầu)*" cũng là những tính năng giúp hỗ trợ người dùng, kết nối dữ liệu hiệu quả. Vì vậy, các cổng DLM cần ưu tiên cải thiện việc mô tả chi tiết dữ liệu, thiết kế giao diện thân thiện, dễ sử dụng.

Một số tính năng bổ trợ khác của cổng DLM là "*Hướng dẫn chi tiết*" và "*Xem trước dữ liệu dưới dạng bảng, đồ thị*". Yêu cầu về "*Nhiều phiên bản ngôn ngữ*" tuy không phải ưu tiên hàng đầu, nhưng vẫn cần được xem xét để đáp ứng nhu cầu từ các nhóm người dùng đa dạng nhất là trong bối cảnh hội nhập.

Các tính năng như tìm kiếm nâng cao, có mô tả dữ liệu, liên kết với cổng quốc gia và cung cấp phiên bản đa ngôn ngữ là những tính năng được đề xuất chủ yếu bởi các viện nghiên cứu/trường đại học, nhiều hơn nhóm doanh nghiệp và các tổ chức khác (*Xem thêm Báo cáo Khảo sát về nhu cầu sử dụng DLM tại Việt Nam*).

Bảng 2: Nhu cầu về các tính năng trên cổng DLM (n=372)

Tính năng	TCDN (n=281)	Cá nhân (n=91)	Tổng (n=372)
Mô tả chi tiết dữ liệu	**60,1%**	**73,9%**	**63,5%**
Giao diện trực quan, dễ dàng sử dụng	**65,8%**	**63,1%**	**65,1%**
Công cụ tìm kiếm nâng cao	**67,3%**	**61,5%**	**65,9%**
Cung cấp dịch vụ API	**60,9%**	**55,4%**	**59,6%**
Liên kết với cổng DLM quốc gia	**57,7%**	**64,6%**	**59,4%**
Phản hồi dữ liệu	**56,9%**	**61,5%**	**58,0%**
Miễn phí truy cập và tải về không yêu cầu đăng nhập	54,5%	58,5%	55,5%
Xem trước dữ liệu dưới dạng bảng, biểu đồ hoặc bản đồ	50,9%	58,5%	52,8%

Official Use Only

Tính năng	TCDN (n=281)	Cá nhân (n=91)	Tổng (n=372)
Cung cấp hướng dẫn chi tiết	50,9%	56,9%	52,4%
Cung cấp nhiều phiên bản ngôn ngữ	48,0%	38,5%	45,7%
Khác	0,7%	1,5%	0,9%

3. Hiện trạng cung cấp dữ liệu mở

Nghiên cứu đã thực hiện đánh giá 34 cổng DLM hiện hành ở Việt Nam. Đây là các cổng đã được truy cập thành công và được nhóm nghiên cứu đánh giá dựa trên Bộ tiêu chí đánh giá như đã nêu tại mục 2.2, Phần I của Báo cáo này. Kết quả đánh giá các cổng DLM hiện nay cho thấy các cổng được hình thành một cách rời rạc, không thể chia sẻ, kết nối dữ liệu với nhau, chưa kết nối với cổng DLM quốc gia *(Xem thêm Báo cáo Đánh giá cổng DLM tại Việt Nam).* Bên cạnh đó, các bộ DLM được cung cấp trên các cổng DLM còn hạn chế về tính đa dạng, chất lượng, đa số được cung cấp dưới dạng khó khai thác (như PDF, DOCX). Các bộ dữ liệu ở định dạng thuận tiện cho việc sử dụng và khai thác như XLSX, CSV chỉ chiếm chưa tới 9% số bộ dữ liệu[49]. Số lượng bộ dữ liệu được cung cấp trên các cổng DLM chuyên dụng cũng rất hạn chế. Kết quả đánh giá 30 cổng DLM cấp tỉnh hiện nay tại Việt Nam chỉ cung cấp trung bình khoảng 181 bộ dữ liệu (chủ yếu dưới định dạng PDF, XLSX). Chỉ có 7 địa phương cung cấp trên 200 bộ dữ liệu, 6 địa phương cung cấp 100-200 bộ dữ liệu, 14 địa phương cung cấp dưới 100 bộ dữ liệu. 10/34 cổng được đánh giá (29,4%) đã thực hiện cung cấp dữ liệu đạt 70% danh mục dữ liệu so với kế hoạch, 10 cổng (29,4%) cung cấp số lượng danh mục đạt 30% đến 70% so với kế hoạch. Các cổng còn lại cung cấp dưới 30% so với kế hoạch Điện Biên, Thái Bình, Thanh Hóa không ban hành danh mục DLM.

Cũng theo kết quả khảo sát của nghiên cứu này, trong tổng số 210 CQNN đã từng nhận được yêu cầu cung cấp dữ liệu từ các TCDN, cá nhân, đa số các CQNN đã cung cấp dữ liệu qua các trang thông tin điện tử của cơ quan (59,1%) và phản hồi trực tiếp cho người yêu cầu (52,9%), việc cung cấp qua các cổng DLM còn thấp (41%). Cách thức cung cấp dữ liệu thông qua cổng DLM vẫn chưa phổ biến, điều này có thể khiến việc khai thác dữ liệu trong các hoạt động kinh doanh, sản xuất và nghiên cứu, giảng dạy trở nên kém hiệu quả vì tốn thời gian và khó tìm kiếm và rõ ràng việc cung cấp hiện nay chưa đáp ứng nhu cầu khai thác, sử dụng của các tổ chức, cá nhân. Các mục dưới đây trình bày kết quả đánh giá cụ thể hơn về các cổng DLM hiện nay.

Các cổng được đánh giá và tính điểm theo bộ tiêu chí đánh giá. Điểm số đánh giá cổng tổng hợp được phân bổ thành hai phần với 2 nhóm tiêu chí chính của mỗi phần, tổng cộng là 4 nhóm tiêu chí chính: (i) Yêu cầu chung của Cổng (điểm tối đa: 8), (ii) Đặc điểm kĩ thuật gồm các tính năng hỗ trợ người dùng (điểm tối đa: 8), (iii) Yêu cầu về chất lượng dữ liệu (điểm tối đa: 7), (iv) Yêu cầu về đặc tả dữ liệu (điểm tối đa: 4) *(Xem thêm Báo cáo Đánh giá cổng DLM tại Việt Nam).*

Kết quả đánh giá các cổng cho thấy sự chênh lệch rõ rệt, phản ánh mức độ phát triển không đồng đều trong việc xây dựng và vận hành các cổng DLM. Có 5 tỉnh/thành phố có cổng DLM tốt nhất, đạt trên 50% điểm số là cổng DLM của Hưng Yên (17,3/27 điểm), Đà Nẵng (15,8/27 điểm), Bình Thuận (15,7/27 điểm), Vĩnh Long (15,0/27 điểm), Hậu Giang (14,2/27 điểm). Các địa phương này đạt kết quả tốt ở các tiêu chí về Chất lượng dữ liệu và Đặc điểm kĩ thuật của cổng, cho thấy sự đầu tư cả về công nghệ lẫn quản trị cổng. Bên cạnh đó, nhiều địa phương chưa hoàn thiện các tính năng và yêu cầu về dữ liệu theo các tiêu chí đánh giá như Bắc Ninh (3,3/27 điểm), Bình Định (6,4/27 điểm), Lạng Sơn (6,4/27 điểm).

[49] Bộ Nội vụ và Viện Khoa học tổ chức nhà nước. Thực trạng và giải pháp phát triển dữ liệu mở tại Việt Nam. 2022 [trích dẫn ngày 8 tháng 5 năm 2024]; Tham khảo tại: https://isos.gov.vn/cds/thuc-trang-va-giai-phap-phat-trien-du-lieu-mo-tai-viet-nam-43241.html

Official Use Only

3.1. Một số ưu điểm của các cổng dữ liệu mở

Trong quá trình nghiên cứu xây dựng bộ tiêu chí đánh giá cổng DLM và nghiên cứu kinh nghiệm, thực tiễn tốt của các cổng DLM trên thế giới, nhóm nghiên cứu đã xác định được một số nhóm tính năng chính của một cổng DLM tốt bao gồm:

- Chức năng quản lý danh mục tập dữ liệu: cho phép lọc, tìm kiếm dữ liệu theo nhiều tiêu chí.
- Chức năng quản lý và lưu trữ dữ liệu: cung cấp nhiều định dạng khác nhau, định dạng máy đọc được, cung cấp Dịch vụ API, đảm bảo tính cập nhật của dữ liệu.
- Chức năng trình bày dữ liệu, xem trước dữ liệu: chức năng xem trước dữ liệu dưới dạng bảng, biểu đồ.
- Chức năng cộng đồng: đánh giá dữ liệu, cung cấp thông tin liên quan đến dữ liệu, tính năng gửi yêu cầu, kết nối với các nền tảng xã hội, cung cấp các ứng dụng dữ liệu, cung cấp tin tức.
- Chức năng quản lý tài khoản: cá nhân hóa tài khoản người dùng.

Theo các chức năng này, nhìn chung, các cổng hiện nay của Việt Nam mới chỉ đáp ứng được một phần nhưng cũng đã có một số điểm tích cực đáng chú ý. Thứ nhất, hầu hết các **cổng DLM (32/34 Cổng) đã cho phép người dùng có thể khai thác dữ liệu mọi lúc, mọi nơi và bất kỳ ai đều có thể truy cập và tải về miễn phí**, có khả năng truy cập ổn định trên các thiết bị điện tử.

Thứ hai, về giao diện, **các cổng DLM được thiết kế tối giản, dễ nhìn và thân thiện với người dùng, tương thích với màn hình máy tính**. Một số cổng có giao diện tương đồng (có thể do có cùng nhà cung cấp dịch vụ xây dựng cổng), tạo sự quen thuộc và dễ sử dụng, đặc biệt nhờ các thành phần tiêu chuẩn như tiêu đề đầu trang, menu điều hướng, chân trang và ô tìm kiếm, giúp người dùng nhanh chóng định vị và truy xuất thông tin cần thiết. Các cổng được thiết kế với giao diện dễ nhìn, rõ ràng các tính năng như tìm kiếm DLM, danh mục DLM, văn bản - hướng dẫn, liên hệ. Đồng thời, với giao diện thân thiện, người dùng có thể dễ dàng xem được đầy đủ thông tin cần thiết như tên cổng, cơ quan chủ quản, email và số điện thoại liên hệ. Sự chi tiết và rõ ràng trong việc cung cấp thông tin trên các cổng góp phần nâng cao trải nghiệm người dùng, đồng thời thể hiện sự cam kết của các đơn vị trong việc hỗ trợ và tương tác hiệu quả với cộng đồng.

Thứ ba, đa số cổng DLM **đã tích hợp một số tính năng cơ bản** như quản lý danh mục (tìm kiếm, lọc dữ liệu), quản lý và lưu trữ dữ liệu (tải xuống dữ liệu, truy cập API), chức năng cộng đồng (bình luận, đánh giá, chia sẻ trên mạng xã hội), quản lý tài khoản (cá nhân hóa tài khoản người dùng). Nổi bật nhất là tính năng lọc dữ liệu, người dùng có thể tìm kiếm và lọc dữ liệu theo nhiều tiêu chí khác nhau, giúp tối ưu hóa quá trình truy xuất thông tin cần thiết, hầu hết các cổng đều cho phép người dùng tìm kiếm và lọc dữ liệu theo nhiều tiêu chí khác nhau như tìm kiểu theo chủ đề có sẵn (31/34 cổng), lọc theo đặc tính cụ thể của dữ liệu như theo thứ tự chữ cái từ A-Z hoặc thời gian đăng tải (30/34 cổng), lọc theo cụm từ ngẫu nhiên (30/34 cổng) và có tính năng hiển thị số lượng dữ liệu theo tiêu chí lọc (29/34 cổng). Bên cạnh đó, hầu hết cổng đều cho phép tải về dữ liệu ở ít nhất một định dạng máy đọc được, phổ biến nhất là định dạng XLS, đáp ứng nhu cầu sử dụng dữ liệu cho các mục đích nghiên cứu và phân tích.

3.2. Một số hạn chế của các cổng dữ liệu mở

Chức năng, tính năng kĩ thuật trên các cổng dữ liệu mở chưa đồng đều

Phần lớn các cổng **chưa khai thác hiệu quả tiềm năng của API,** có 61,8% cổng (21/34 cổng) không cung cấp bất kỳ dịch vụ API nào, không cổng nào cung cấp đặc tả bộ dữ liệu bằng API. Điều này hạn chế việc nâng cao khả năng tương tác và tự động hóa trong khai thác dữ liệu. **Nhiều cổng chưa cho phép người dùng gửi yêu cầu dữ liệu** (20/34 cổng, tương ứng 58,8%, không đáp ứng) và 20/34 cổng (58,8%) **không cung cấp bất kỳ cơ chế phản hồi về bộ dữ liệu**, cho thấy sự thiếu hụt trong việc xây dựng môi trường tương tác, tiếp nhận ý kiến người dùng để cải thiện chất lượng dịch vụ. Hiện **không có cổng nào cung cấp ví dụ tái sử dụng dữ liệu,** khiến việc khai thác dữ liệu trở nên phức tạp, đặc biệt đối với người dùng mới hoặc thiếu kinh nghiệm. Một số cổng ghi nhận tình trạng không thể truy cập được trong một vài ngày hoặc không ổn định như cổng DLM

Official Use Only

quốc gia, cổng của Cần Thơ, Thanh Hóa, Lâm Đồng. Các lỗi truy cập thường gặp như báo kết nối không bảo mật, "Lỗi 404 Not Found".

Số lượng và chất lượng dữ liệu được cung cấp trên các cổng dữ liệu mở còn hạn chế

Số lượng dữ liệu cung cấp trên các cổng còn ít và chưa đáp ứng theo đúng kế hoạch: Tại thời điểm đánh giá, nhiều địa phương chưa cung cấp đủ số lượng bộ dữ liệu theo kế hoạch**.** Chỉ có 10/34 cổng, tương đương 29,4%, đã thực hiện cung cấp dữ liệu đạt 70% danh mục dữ liệu so với kế hoạch. Trong các cổng DLM của địa phương, cổng của Đà Nẵng cung cấp số lượng bộ dữ liệu nhiều nhất, với trên 1.200 bộ (vượt chỉ tiêu so với kế hoạch). Nhiều địa phương mới cung cấp rất ít dữ liệu (dưới 30 bộ ở tất cả các định dạng), chưa đạt kế hoạch như Yên Bái (09 trên tổng số 95 bộ cho 15 lĩnh vực theo kế hoạch), Bình Dương (16 bộ trên tổng số 206 bộ cho 14 lĩnh vực theo kế hoạch), Vĩnh Phúc (26 bộ trên tổng số 206 bộ cho 14 lĩnh vực theo kế hoạch). Số lượng bộ dữ liệu trên các cổng DLM theo các lĩnh vực/chủ đề không đồng đều, dẫn đến tình trạng mất cân đối trong cung cấp thông tin ở các lĩnh vực/chủ đề quan trọng. Ví dụ, Hà Giang có tổng cộng 125 bộ dữ liệu nhưng lại không có bất kỳ dữ liệu nào thuộc các lĩnh vực về Nội vụ, Thanh tra, Khoa học và Công nghệ, Tư pháp, Lao động - Thương binh và Xã hội, hay Tài nguyên - Môi trường.

Về chất lượng dữ liệu, **nhiều cổng chủ yếu cung cấp dữ liệu dưới định dạng PDF**, điều này làm giảm khả năng sử dụng của dữ liệu, vì định dạng này không thể dễ dàng đọc hoặc phân tích tự động. Các dữ liệu dưới dạng mà máy có thể đọc được chủ yếu dưới dạng XLSX và chưa phát triển đa dạng các định như CSV, JSON hoặc XML như trên thế giới.

Nhiều cổng DLM hiện nay **vẫn sử dụng dữ liệu cũ và không được cập nhật thường xuyên**, dẫn đến việc chất lượng thông tin không còn chính xác và thiếu tính thời sự, không đáp ứng được nhu cầu sử dụng dữ liệu mới và chính xác để phục vụ cho nghiên cứu, ra quyết định hay các hoạt động phân tích mặc dù theo quy định, DLM được cung cấp phải là dữ liệu được cập nhật mới nhất (Điều 17, Nghị định 47/2020/NĐ-CP). Ví dụ như tại cổng DLM của Bắc Ninh, lần cuối cổng này cập nhật dữ liệu là ngày 05/01/2022 với nội dung thuộc lĩnh vực Tư pháp, còn lại dữ liệu về An ninh - Quốc phòng và Lao động - Thương Binh và xã hội được cập nhập lần cuối lần lượt vào năm 2014 và 2019. Chỉ có một số ít cổng dữ liệu thực hiện cập nhật dữ liệu thường xuyên, như cổng của Đà Nẵng. Dữ liệu giữa các Bộ, ngành, địa phương chưa có sự kết nối, đồng bộ với nhau và với cổng DLM quốc gia.

Nhiều cổng DLM chưa đảm bảo tính bảo mật, làm lộ lọt dữ liệu cá nhân: Điều 20 Nghị định 47/2020/NĐ-CP yêu cầu không tiết lộ thông tin cá nhân khi công bố DLM của CQNN. Việc lộ lọt thông tin cá nhân gây nguy cơ rủi ro bảo mật và vi phạm quyền riêng tư của công dân. Tuy nhiên, theo rà soát, có **18/34 cổng DLM hiện cung cấp dữ liệu chưa đảm bảo tính bảo mật, làm lộ dữ liệu cá nhân** bao gồm những thông tin có thể xác định cá nhân cụ thể hoặc gắn liền với một cá nhân nào đó (như họ tên, số điện thoại, số căn cước công dân, địa chỉ, số chứng chỉ hành nghề), gây nguy cơ rủi ro bảo mật và vi phạm quyền riêng tư của công dân.

Đặc tả dữ liệu chưa đầy đủ

Các cổng DLM hiện nay đã bước đầu cung cấp các mô tả về dữ liệu. Tuy nhiên, các trường dữ liệu trong lớp Dataset chưa đáp ứng được yêu cầu của Tiêu chuẩn DCAT-VN 14171:2024 về Đặc tả tập DLM cho cổng dữ liệu. Tiêu chuẩn đã đặt ra 6 trường dữ liệu bắt buộc phải cung cấp là (i) Tiêu đề; (ii) Tên cơ quan, tổ chức cung cấp; (iii) Mã định danh được tự động sinh ra và duy nhất; (iv) Chủ đề theo danh mục DLM; (v) Ngày cung cấp; (vi) Ngày cập nhật. Tuy nhiên trên thực tế, nhiều cổng chưa đáp ứng các tiêu chuẩn về trường dữ liệu. Có **13/34** cổng có trường dữ liệu về danh mục, **15/34** cổng cung cấp mã định danh cho bộ dữ liệu, **20/34** cổng có thông tin về ngày dữ liệu được cung cấp và **22/34** cổng cung cấp được tên cơ quan tổ chức cung cấp dữ liệu. Tình trạng thiếu thông tin quan trọng về dữ liệu gây khó khăn cho người dùng trong việc truy xuất nguồn gốc, xác định thời điểm cập nhật và tính minh bạch của dữ liệu.

Official Use Only

Hình 2: Kết quả đánh giá cổng DLM cấp Quốc gia và Bộ (n=4)

Tên Cổng	Tiêu chí về yêu cầu chung (Tối đa 8 điểm)	Tiêu chí về tính năng kỹ thuật (Tối đa 8 điểm)	Tiêu chí về dữ liệu (Tối đa 7 điểm)	Tiêu chí về đặc tả dữ liệu (Tối đa 4 điểm)	Tổng điểm 4 tiêu chí (Tối đa 27 điểm)
Bộ TNMT	1	2.9	4.3	2.5	10.7
Quốc gia	3	3.9	3.3	0.5	10.7
Bộ TT&TT	1	0.8	3.7	0.1	5.5
Bộ Tư pháp	2	0.9	2	0	4.9

**Dựa trên kết quả đánh giá cổng DLM tại Việt Nam của nhóm Đánh giá*

Hình 3: Kết quả đánh giá cổng DLM cấp tỉnh (n=30)

Tên Cổng	Tiêu chí về yêu cầu chung (Tối đa 8 điểm)	Tiêu chí về tính năng kỹ thuật (Tối đa 8 điểm)	Tiêu chí về dữ liệu (Tối đa 7 điểm)	Tiêu chí về đặc tả dữ liệu (Tối đa 4 điểm)	Tổng điểm 4 tiêu chí (Tối đa 27 điểm)
Hưng Yên	4	6.4	4.2	2.7	17.2
Đà Nẵng	5.5	5	3.7	1.6	15.8
Bình Thuận	5.5	4.4	3.2	2.6	15.6
Vĩnh Long	4.5	3.9	4.2	2.4	15
Hậu Giang	5	3.6	3.2	2.4	14.2
Đắk Nông	4	4.3	2.8	2.4	13.5
Thừa Thiên Huế	3	4.4	3.2	2.8	13.4
Hải Phòng	3	2.4	4.7	2.6	12.6
Thanh Hoá	4	3.6	2.2	2.6	12.4
Bình Dương	4	3.4	2.5	2.4	12.3
Quảng Ngãi	3	2.9	4.2	2.3	12.3
Gia Lai	3	4.4	2	2.4	11.8
Vĩnh Phúc	5	3.3	1.3	1.4	11
Thái Bình	4	3.5	1.8	1.3	10.6
Sơn La	4	2.1	2.8	1.4	10.4
Điện Biên	3	3.3	2.5	1.6	10.3
Kon Tum	4	1.8	2.3	2.3	10.3
Đồng Tháp	4	3.8	0.8	1.4	10
Lâm Đồng	5	2.1	1.5	1.4	10
TP. Hồ Chí Minh	4	3.1	1	1.4	9.5
Cần Thơ	4	2.3	1.7	1.4	9.3
Long An	2	2.1	2.7	2.4	9.2
Trà Vinh	3	2.8	1.8	1.4	9
An Giang	3	1.4	1.8	2.4	8.6
Hà Giang	3	2.4	1.3	1.3	8
Yên Bái	2	2.6	1.8	1.3	7.7
Ninh Bình	2	3	1.3	1.3	7.6
Lạng Sơn	2	1.4	1.8	1.3	6.5
Bình Định	1	1.3	1.8	1.4	5.5
Bắc Ninh	1	0.1	2	0.2	3.3

**Dựa trên kết quả đánh giá cổng DLM tại Việt Nam của nhóm Đánh giá*

3.3. Đánh giá của người sử dụng về cổng dữ liệu mở tại Việt Nam

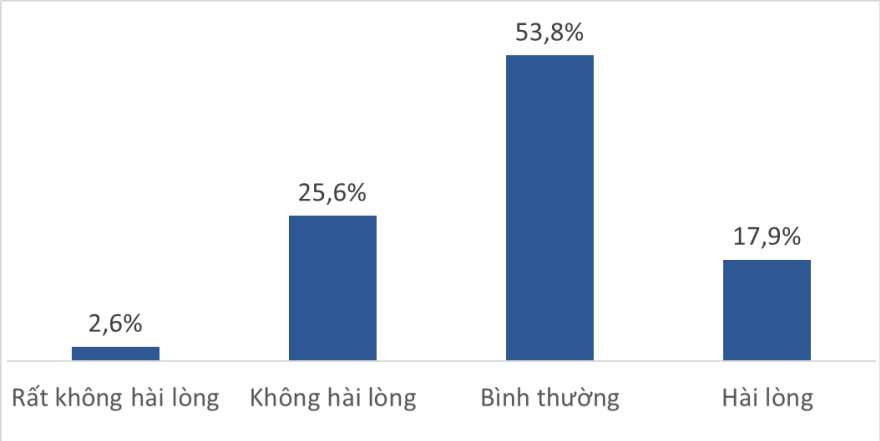
Trong tổng số 95 người sử dụng cá nhân tham gia khảo sát của nghiên cứu này, có 54 người đã từng sử dụng các nguồn DLM tại Việt Nam bao gồm các cổng dữ liệu của các Bộ, ban, ngành cung cấp (Ví dụ: Tổng

Official Use Only

cục Thống kê, Cổng công khai ngân sách nhà nước, Cổng dữ liệu quốc gia,...) và được hỏi về trải nghiệm, mức độ hài lòng đối với dữ liệu đã tiếp cận được.

Về mức độ hài lòng, trong những người đã từng sử dụng đối với cổng DLM có 17,9% "hài lòng" và 53,8% cảm thấy "bình thường". Tỉ lệ không hài lòng và rất không hài lòng là 28,2%.

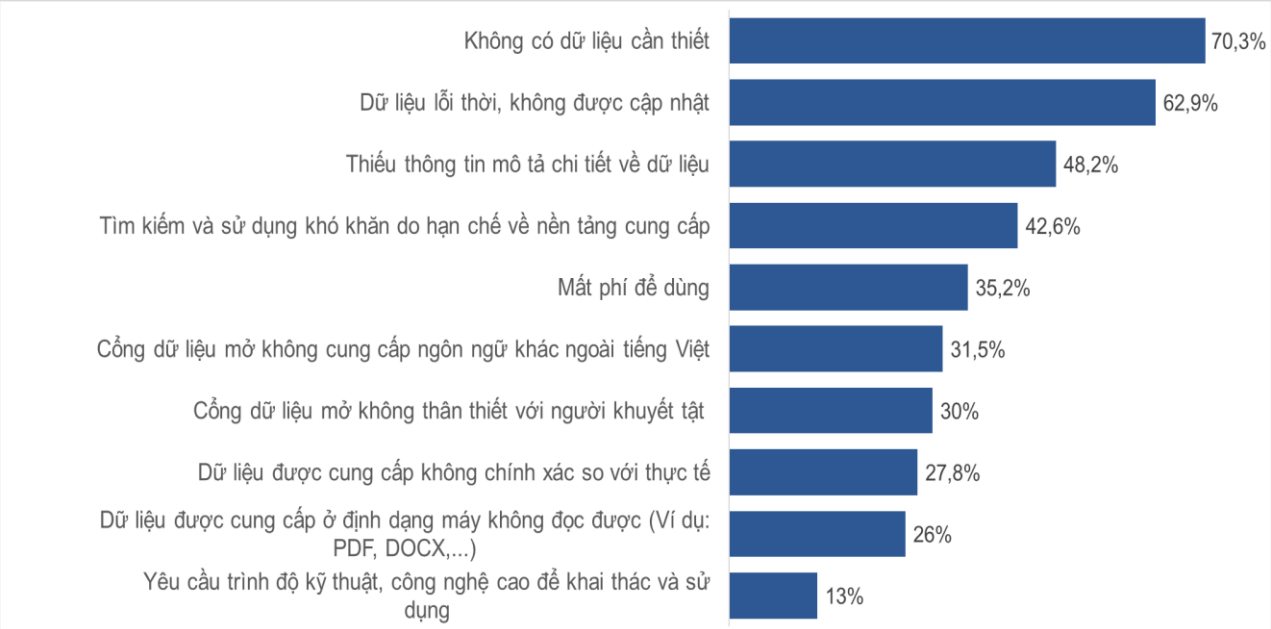
Hình 4: Mức độ hài lòng về các cổng DLM mà người dùng cá nhân đã từng sử dụng (n=54)



**Dựa trên kết quả khảo sát của nhóm Đánh giá về nhu cầu sử dụng DLM tại Việt Nam*

Có 70% người dùng cá nhân cho biết họ thường xuyên và rất thường xuyên gặp khó khăn khi *không tìm kiếm được các dữ liệu cần thiết* trên cổng. Tiếp đến là các khó khăn như các cổng cung cấp *dữ liệu lỗi thời hoặc không được cập nhật* cũng là một hạn chế lớn (62,9%); *thiếu thông tin mô tả chi tiết về dữ liệu (*48,2%*);* việc tìm kiếm và sử dụng khó khăn do hạn chế về nền tảng cung cấp (42,6%). Đáng chú ý, có 26% gặp vấn đề dữ liệu không được cung cấp ở dạng máy đọc được.

Hình 5: Tỷ lệ người dùng cá nhân thường xuyên và rất thường xuyên gặp những khó khăn khi truy cập các cổng DLM (n=54)



**Dựa trên kết quả khảo sát của nhóm Đánh giá về nhu cầu sử dụng DLM tại Việt Nam*

Official Use Only

Kết quả trên cho thấy ưu tiên hàng đầu giúp tăng sự hài lòng của người dùng là liên quan đến chất lượng dữ liệu. Vì vậy, các hoạt động cần ưu tiên là cung cấp đa dạng danh mục các bộ dữ liệu khác nhau với chất lượng cao, được cập nhật thường xuyên, có mô tả chi tiết về dữ liệu, dưới nhiều định dạng khác nhau và được mở miễn phí cho người dùng được tự do truy cập và sử dụng.

Nhóm nghiên cứu đã thực hiện 04 cuộc PVS với các nhà nghiên cứu cá nhân trong lĩnh vực kinh tế, các TCDN sử dụng dữ liệu để tạo ra các sản phẩm dịch vụ, kinh doanh. Các ý kiến thu được cho thấy người sử dụng cảm thấy dễ dàng tiếp cận với các nguồn dữ liệu miễn phí từ các cổng DLM quốc tế hơn như cổng của World Bank, OECD, cổng DLM của Pháp,... do nguồn dữ liệu dồi dào, đa dạng các định dạng (đặc biệt ở các định dạng máy đọc được), trường dữ liệu có cấu trúc nghiêm ngặt, rõ ràng, được cập nhật thường xuyên và được cung cấp trong nhiều năm giúp người sử dụng dễ dàng khai thác và tái sử dụng.

> *"Các cổng dữ liệu trong nước, điển hình như GSO,... càng ngày càng khó sử dụng do cách sắp xếp dữ liệu chưa khoa học. Một trong những vấn đề đáng lo ngại nữa khi sử dụng DLM tại Việt Nam là dữ liệu không chính xác giữa các nguồn khác nhau và tính cập nhật hơi chậm" (PVS nhà nghiên cứu #2)*
>
> *"Cần có một đầu mối để tập trung DLM, như Cổng DLM quốc gia đang thực hiện. Điều này tương tự như đối với GSO là nơi nên tập trung dữ liệu thống kê của cả nước. Nhưng trên thực tế, khi tìm các chỉ số, vẫn rải rác, mỗi nơi ra 1 con số. Vì thế cần phải thống nhất dữ liệu ở các nguồn khác nhau và tập trung vào một đầu mối" (PVS nhà nghiên cứu #3)*
>
> *"Mình sử dụng dữ liệu thứ cấp, GSO cho cái gì thì biết cái đấy. Nhưng rõ ràng những bộ số liệu như VHLSS, về gia đình, giáo dục hoàn toàn có nhưng không bao giờ mình được tiếp cận. Nếu muốn thì mình phải đi mua từ người nọ người kia, không bao giờ được tiếp cận mở. Tất nhiên ở nước ngoài vẫn có những bộ số liệu phải trả tiền nhưng có những bộ số liệu đương nhiên người dân có quyền tiếp cận. Cần có danh mục các bộ dữ liệu cơ bản mà người dân có quyền tiếp cận miễn phí" (PVS nhà nghiên cứu #1)*
>
> *"Một trong những khó khăn khác là về tính sẵn có và tính truy cập của dữ liệu. Nhiều dữ liệu phải xin phép bằng email/văn bản. Thường xuyên bị từ chối cung cấp dữ liệu, ngay cả với thông tin không nhạy cảm. Vì quy trình phê duyệt, có thể chưa chốt được dữ liệu hoặc nhà nước thấy chưa phải thời điểm để công bố" (PVS nhà nghiên cứu #3)*
>
> *"Ở Việt Nam, giao diện của các cổng cung cấp dữ liệu chưa được sáng sủa và khoa học như các nước khác. Điều hướng trang chưa thân thiện và còn rối. Không có hướng dẫn nên khó khăn trong việc tìm kiếm. Hơn nữa, vào không tìm được gì, khả năng truy cập thấp, không có sự tập trung về dữ liệu mà rải rác khắp nơi. Không có một trang tổng hợp mọi thứ cần thiết để doanh nghiệp sản xuất hoặc ngân hàng có thể sử dụng để khai thác thông tin. Và nếu có thì CSDL cũng đã rất lỗi thời." (Tổ chức doanh nghiệp #1)*

4. Một số khó khăn trong quá trình cung cấp dữ liệu mở của cơ quan nhà nước

Một số khó khăn chung trong quá trình cung cấp DLM nhiều CQNN gặp phải (46,1%-57,5% tổng số CQNN tham gia khảo sát) bao gồm: *thiếu nhân sự chuyên môn, thiếu hướng dẫn cụ thể, hạn chế về cơ sở hạ tầng và công nghệ, thiếu kinh phí, thiếu sự phối hợp từ các bên liên quan*. Một số cơ quan cũng gặp vấn đề do *nguồn dữ liệu không sẵn có (40,4%)* và *thiếu nhận thức của các bên liên quan* (39,6%). Một số ý kiến cho rằng *"thiếu các tiêu chuẩn dữ liệu"* cũng là một khó khăn mà các đơn vị cung cấp dữ liệu phải đối mặt.

Trong số những khó khăn điển hình trên, kết quả khảo sát cho thấy rằng các CQNN cấp địa phương gặp nhiều khó khăn hơn về hầu hết các vấn đề. Cụ thể, so với các CQNN cấp TW, các CQNN cấp địa phương gặp khó khăn trong việc thiếu nhân sự chuyên môn (60,9%), thiếu hướng dẫn cụ thể (57,4%), thiếu kinh phí (48,5%), thiếu cơ sở hạ tầng, công nghệ (50,5%). Đặc biệt, các CQNN cấp địa phương gặp khó khăn lớn, so với cấp TW, vì các bên liên quan thiếu nhận thức về DLM (43,6%). Ngoài ra, phân thích theo chức năng của các đơn vị cho thấy rằng các đơn vị, ngoài các đơn vị chuyên trách CNTT, gặp khó khăn do thiếu cơ sở hạ tầng, công nghệ (52,6%).

Official Use Only

Official Use Only

Bảng 3: Những khó khăn của các CQNN trong quá trình cung cấp DLM (n=280)

	Theo cấp		Theo chức năng		
	Trung ương (n =78)	Địa phương (n=202)	Đơn vị chuyên trách CNTT (n=105)	Đơn vị khác (n=175)	Tổng (n=280)
Thiếu nhân sự chuyên môn	**38 (48,7%)**	**123 (60,9%)**	**58 (55,2%)**	**103 (58,9%)**	**161 (57,5%)**
Thiếu hướng dẫn cụ thể	32 (41%)	**116 (57,4%)**	**52 (49,5%)**	**96 (54,9%)**	**148 (52,9%)**
Thiếu kinh phí	31 (39,7%)	**98 (48,5%)**	42 (40%)	**87 (49,7%)**	**129 (46,1%)**
Thiếu cơ sở hạ tầng, công nghệ	29 (37,2%)	**102 (50,5%)**	39 (37,1%)	**92 (52,6%)**	**131 (46,8%)**
Nguồn dữ liệu không sẵn có	26 (33,3%)	87 (43,1%)	45 (42,9%)	68 (38,9%)	113 (40,4%)
Thiếu sự phối hợp của các bên liên quan	25 (32,1%)	90 (44,6%)	49 (46,7%)	66 (37,7%)	115 (41,1%)
Thiếu nhận thức của các bên liên quan về DLM	23 (29,5%)	88 (43,6%)	45 (42,9%)	66 (37,7%)	111 (39,6%)
Khác	2 (2,6%)	5 (2,5%)	3 (2,9%)	4 (2,3%)	7 (2,5%)

Những khó khăn, rào cản về thiếu chính sách, quy định, hướng dẫn cụ thể, hạn chế về cơ sở hạ tầng, kinh phí và sự nhận thức của các bên liên quan về DLM cũng được ghi nhận trong các cuộc PVS với các bên liên quan ở cấp bộ và cấp tỉnh và đã được trình bày tại Mục 1, Phần II ở trên, trong các mục dưới đây tập trung phân tích một số khó khăn, rào cản khác.

4.1. Cơ sở hạ tầng, kĩ thuật và công nghệ thông tin

Một số thách thức khác liên quan đến ***cơ sở hạ tầng CNTT nhằm đáp ứng các yêu cầu phức tạp trong kĩ thuật xử lý dữ liệu đa miền và hạ tầng công nghệ chưa được đầu tư đúng mức*** **được nhiều ý kiến nêu ra**. Cần quan tâm và đầu tư đúng mức vào công nghệ xử lý dữ liệu đa miền do hiện nay dữ liệu được thu thập từ nhiều nguồn khác nhau, đa dạng và phức tạp trong cấu trúc và nội dung dữ liệu. Việc tích hợp, xử lý và chuẩn hóa các loại dữ liệu khác nhau để có thể khai thác hiệu quả đặt ra một thách thức lớn cho các đơn vị thiếu cơ sở hạ tầng CNTT. Trên thực tế, hệ thống lưu trữ và tính toán hiện nay chưa đáp ứng được nhu cầu cung cấp DLM của cả quốc gia.

> *"Chúng tôi gặp khó khăn về kĩ thuật, phần xử lý dữ liệu, đặc biệt là dữ liệu đa chiều. Với những cổng DLM như này, dữ liệu từ rất nhiều nguồn khác nhau và nhiều kiểu dữ liệu khác nhau. Bên cạnh đó là khó khăn liên quan đến hạ tầng lưu trữ, tính toán. Trước đây, chúng ta làm việc với dữ liệu có cấu trúc, đầu tư cho hạ tầng rất nhỏ và tính toán dựa trên các phương pháp truyền thống. Nhưng với những dữ liệu đa miền, tạm gọi là dữ liệu lớn hiện nay thì đây là những khó khăn chung. Ngoài phần cơ sở hạ tầng lưu trữ là phần hạ tầng tính toán, phục vụ cho những nhóm nghiên cứu đang phải làm việc dựa trên dữ liệu lớn. Vì đầu tư cho những cái hạ tầng tính toán rất đắt, hiện nay ở Việt Nam, nói về những lab nghiên cứu mà có hạ tầng tính toán đủ mạnh để thực hiện nghiên cứu những mô hình lớn thì rất ít. Phía Google là doanh nghiệp hàng đầu có tài trợ thử nghiệm và tài trợ hỗ trợ. Còn là triển khai để giải quyết bài toán lớn của quốc gia là rất nhiều tiền." (Cán bộ cấp Bộ #1)*

> *"Hạ tầng kĩ thuật còn chưa đảm bảo. Hạ tầng CNTT trong CQNN được đầu tư khá là ít và thời gian đầu tư đã quá lâu rồi. Có những hệ thống đã được đầu tư cách đây khoảng 20 năm nên rất lộm cộm, dữ liệu không được biên tập, không được xây dựng cấu trúc dữ liệu chặt chẽ, các trường dữ liệu sơ sài, công nghệ thì cũng không có gì, máy chủ, website rất là cũ kĩ, lạc hậu. Những điều này dẫn đến việc tích hợp dữ liệu hoặc chia sẻ vào chung 1 cổng cực kỳ khó khăn." (Cán bộ cấp Bộ #2)*

4.2. Kinh phí và nguồn lực

Các văn bản, quy định đã nêu rõ việc tăng cường chia sẻ DLM ở các cơ quan, đơn vị, tuy nhiên, các chính sách liên quan đến nguồn kinh phí chưa được làm rõ, dẫn đến việc đầu tư cho các hoạt động về DLM tại các địa phương có sự khác biệt đáng kể. Điều này giải thích vì sao, chất lượng cổng DLM và số lượng DLM,

Official Use Only

sự cam kết giữa các Sở, ban, ngành, hoạt động quản lý, giám sát, vận hành cổng DLM ở các địa phương có sự chênh lệch. Hiện nay, các nguồn kinh phí cho hoạt động về DLM toàn bộ được lấy từ ngân sách địa phương vì vậy còn rất hạn chế, nguồn xã hội hóa cho DLM cũng chưa được thúc đẩy. Về lâu dài, cần đẩy mạnh nguồn ngân sách và đầu tư, huy động nguồn vốn xã hội hóa cho xây dựng cơ sở hạ tầng, nền tảng CNTT, nhân lực chất lượng cao đảm bảo cho thúc đẩy và duy trì hoạt động cung cấp DLM một cách hiệu quả.

Quy trình chi tiêu tài chính vẫn chưa rõ ràng. Trong khi thủ tục thực hiện các dự án đầu tư sử dụng ngân sách nhà nước cần theo quy định của pháp luật hiện hành, nhưng một cán bộ địa phương khác cho biết đang gặp một số vướng mắc về kinh phí do những bất cập trong quy định của Nghị định 82/2024/NĐ-CP ngày 10/7/2024 của Chính phủ sửa đổi Nghị định 73/2019/NĐ-CP quy định quản lý đầu tư ứng dụng CNTT. Chính sách mua sắm hiện tại còn nhiều bất cập hạn chế khả năng cung cấp các sản phẩm, dịch vụ có chất lượng tốt hơn nếu áp dụng các tiêu chuẩn về mua sắm hiện hành, vì vậy, cần làm rõ các điều khoản, hướng dẫn, tạo điều kiện cho các đơn vị triển khai hoạt động DLM một cách thuận lợi.

> *"Nghị định mới ban hành có một số nội dung về chi đầu tư CNTT còn vướng mắc khiến các địa phương không thực hiện được. Tỉnh đã có văn bản gửi cho Bộ rồi, hy vọng Bộ sớm phản hồi và hướng dẫn lại cho địa phương. Tại khoản 29 Điều 1, trong trường hợp Giám đốc các Sở được Chủ tịch UBND tỉnh uỷ quyền phê duyệt dự án, thì thẩm quyền quyết định đầu tư thành lập Hội đồng thẩm định thuộc về ai; cơ quan, đơn vị chuyên môn quản lý về tài chính, ngân sách nhà nước có thẩm quyền tổ chức thẩm định dự án thuộc cơ quan nào; và đơn vị chuyên môn về CNTT ở đây được xác định là đơn vị trực thuộc UBND tỉnh hay trực thuộc các Sở được Chủ tịch UBND tỉnh uỷ quyền." (Cán bộ Sở TT&TT cấp tỉnh #3)*

Về nguồn nhân lực, khó khăn lớn nằm ở việc thiếu nhân lực chuyên môn, đặc biệt là các chuyên gia kĩ thuật và chuyên gia về dữ liệu như kĩ sư dữ liệu (data engineer). Đây là những vị trí then chốt trong việc xây dựng và vận hành cổng DLM, nhưng nguồn nhân lực hiện tại chưa đáp ứng được yêu cầu cả về số lượng lẫn chất lượng.

> *"Hiện tại, một số đơn vị khó khăn trong việc xác định DLM vì đơn vị cung cấp không chuyên về CNTT, họ chỉ lấy những dữ liệu họ đang khai thác, thao tác hàng ngày trong công việc, thì họ xác định đó là dữ liệu có thể cung cấp được. Họ tổng hợp bằng file XLSX, PDF hoặc DOCX, cung cấp cho Trung tâm Chuyển đổi số để đưa lên cổng DLM." (Cán bộ Sở TT&TT cấp tỉnh #5)*

4.3. Nhận thức của các bên liên quan

Nhận thức của cán bộ cơ quan nhà nước

Nhận thức về DLM của các cán bộ CQNN còn hạn chế là một thách thức làm chậm tiến trình phát triển DLM tại Việt Nam. Kết quả khảo sát về nhận thức của các đơn vị cung cấp và khai thác DLM cho thấy rằng phần lớn các CQNN, TCDN và người dùng cá nhân đã biết về các lợi ích của DLM nhưng vẫn hiểu sai về DLM, gồm khái niệm, định dạng, khả năng truy cập và cách thức cung cấp.

Một đại diện của địa phương cho biết nhận thức hạn chế về tiêu chuẩn DLM khiến nhiều cán bộ lúng túng trong việc cung cấp và xử lý DLM, dẫn đến việc nhiều DLM vẫn được đăng tải dưới định dạng PDF mà máy không thể đọc được. Mà một trong những nguyên nhân lại là do quy định hiện hành chưa cụ thể, rõ ràng.

> *"Vì Bộ không có hướng dẫn cụ thể mà địa phương lại rất băn khoăn về tính pháp lý của các dữ liệu khi đưa ra. Vì thế, để đảm bảo tính toàn vẹn dữ liệu và tính pháp lý của dữ liệu nên vẫn phải đưa vào dạng có pháp lý là file PDF được các CQNN ký số vào. Còn các dữ liệu để chia sẻ dạng trường thông tin (ví dụ XLSX) cũng có nhưng vì phần hướng dẫn chưa được chặt chẽ, cho nên phần chia sẻ vẫn ở trong dạng PDF. Tuy nhiên, nếu một doanh nghiệp triển khai một nhiệm vụ cần lấy dữ liệu bên tỉnh hoặc của Sở CNTT, thì bên mình sẵn sàng chia sẻ bình thường, không vướng mắc gì". (Cán bộ Sở TT&TT cấp tỉnh #3)*
>
> *Quy định về DLM có thể chưa được phổ biến rộng rãi, dẫn đến tình trạng tổ chức không nắm rõ hoặc không tuân thủ đúng yêu cầu. Công tác hướng dẫn khai thác cổng DLM chưa được triển khai đồng*

Official Use Only

bộ, dẫn đến tổ chức không hiểu rõ cách cung cấp dữ liệu đúng yêu cầu." (Cán bộ Sở TT&TT cấp tỉnh #4)

Văn hóa chia sẻ dữ liệu giữa các cơ quan, TCDN còn chưa được phát triển, ***cán bộ CQNN có nhận thức chưa cao về vấn đề bảo mật dữ liệu cá nhân và chưa năm rõ dữ liệu nào nên được công khai và dữ liệu nào cần bảo mật.*** Điều này dẫn đến tâm lý lo ngại về rủi ro mất an toàn thông tin và vi phạm quyền riêng tư của cá nhân. Kết quả là, ngay cả những thông tin có thể công khai một cách hợp pháp và hữu ích cho cộng đồng cũng bị giữ lại, làm hạn chế khả năng tiếp cận và sử dụng dữ liệu của các bên liên quan, đồng thời cản trở sự phát triển của chính phủ mở và kinh tế dữ liệu.

"Văn hóa chia sẻ còn đang ở rất thấp, cần nâng cao lên, để mọi người biết là khi mình chia sẻ ra thì được lợi gì. Vẫn còn nghi ngại là mình chia sẻ mà các bên khác không chia sẻ. Nên phải có cơ chế ràng buộc hoặc lợi ích, như tuyên dương, đánh giá tốt,..." (Cán bộ cấp Bộ #3)

Trong 105 CQNN chuyên trách về TT&TT tham gia khảo sát[50], có khoảng 60% cơ quan đã tổ chức các sự kiện liên quan đến DLM; 50% CQNN cấp TW và 70% CQNN cấp địa phương đã từng được tập huấn về DLM. Trong 175 cán bộ CQNN cung cấp DLM tham gia khảo sát, tỷ lệ cán bộ cho biết đã tham gia các sự kiện về DLM đạt khoảng 50% và tỷ lệ đã được đào tạo/tập huấn về DLM còn tương đối thấp, đặc biệt ở cấp địa phương chỉ khoảng 30% *(Xem thêm Báo cáo Khảo sát về thực trạng cung cấp DLM tại Việt Nam)*. Trong quá trình cung cấp DLM, các cơ quan cho biết gặp nhiều khó khăn do chưa hiểu rõ về DLM và cách triển khai. Việc cung cấp DLM yêu cầu sự tham gia của tất cả các cơ quan, đơn vị ở tất cả các lĩnh vực. Vì thế, cần phải phổ biến các buổi đào tạo/tập huấn đến cả những cơ quan khác tại mỗi địa phương, để cải thiện khả năng cung cấp DLM ở đa dạng các lĩnh vực, từ đa dạng cơ quan cung cấp, tăng cường sự tham gia của các cơ quan vào các sự kiện về DLM.

Việc tổ chức các chương trình đào tạo, hội thảo và các buổi chia sẻ kinh nghiệm từ các chuyên gia trong và ngoài nước sẽ giúp các cán bộ nâng cao năng lực và tư duy về DLM, từ đó tạo ra sự thay đổi tích cực trong việc áp dụng và triển khai DLM ở các cấp CQNN. Các nước như Estonia, Anh Quốc và Hàn Quốc đã thành công trong việc phát triển DLM nhờ vào các chương trình đào tạo bài bản cho cán bộ CQNN và sự kết hợp chặt chẽ giữa các cấp chính quyền trong việc triển khai các chiến lược DLM[51]. Theo báo cáo của Chính phủ Estonia, việc đào tạo 100% cán bộ công chức về chính phủ điện tử và DLM đã giúp quốc gia này đứng đầu về chỉ số chính phủ điện tử toàn cầu[52].

Nhận thức của tổ chức, cá nhân sử dụng DLM

Nhận thức của các TCDN vẫn còn chưa đầy đủ về định dạng chuẩn của DLM. Theo các thực hành tốt trên thế giới về DLM được cung cấp bởi World Bank[53], DLM phải được cung cấp dưới dạng máy đọc được (machine readable) như XLSX, CSV, JSON hay dạng API. Kết quả khảo sát của nhóm đánh giá cho thấy rằng phần lớn các TCDN cho rằng định dạng của DLM có thể là PDF (82%), DOCX (61%), một phần có thể do các quy định hiện hành chưa đủ rõ ràng về định dạng của DLM. Trong nhóm TCDN, các viện nghiên cứu/trường đại học có kiến thức về định dạng DLM tốt hơn so với các tổ chức khác (tổ chức quốc tế và các cơ quan truyền thông, báo chí). Người dùng cá nhân nhận thức tốt hơn về của các định dạng của DLM và cho rằng DLM nên là các định dạng như XLSX và CSV, với tỷ lệ lần lượt là 55% và 69%.

Ngoài ra, ***cộng đồng còn có nhận thức hạn chế về quyền riêng tư và an ninh mạng***. Tỷ lệ người dùng cá nhân cho rằng DLM không được phép chứa thông tin cá nhân là 73,5%, đã cho thấy rằng người dùng cá nhân

[50] Bao gồm Cục, Vụ, Trung tâm CNTT hoặc tương đương của các Bộ, ngành và Sở TT&TT tại các tỉnh
[51] Bộ Đổi mới Kinh doanh & Kỹ năng của Vương quốc Anh, 2014, https://assets.publishing.service.gov.uk/media/5a7ec24940f0b62305b82ff6/bis-14-946-open-data-strategy-2014-2016.PDF.
[52] Vassil, 2016, https://pubdocs.worldbank.org/en/165711456838073531/WDR16-BP-Estonian-eGov-ecosystem-Vassil.PDF; Quan hệ đối tác Chính phủ mở, 2023, https://www.opengovpartnership.org/stories/estonias-journey-of-open-government-reforms/
[53] Ngân hàng Thế giới, Khóa học về dữ liệu dành cho các nhà cung cấp: *https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/essentials*

Official Use Only

đã có nhận thức đúng về quy định bảo vệ thông tin cá nhân theo pháp luật tại Việt Nam và thông lệ quốc tế về DLM. Mặc dù vậy, kết quả khảo sát cho thấy rằng vẫn có 22,1% người dùng cá nhân cho rằng DLM có thể chứa thông tin cá nhân. Với việc đất nước ngày càng phụ thuộc hơn vào công nghệ số, trong đó có cả chính phủ số, nhu cầu tăng cường kiến thức về an ninh mạng cho người dùng trở nên ngày một cấp thiết. Kinh nghiệm của các quốc gia thành công cho thấy kiến thức về an ninh mạng không chỉ được dạy ở trường học mà còn được bổ sung ở các chiến dịch tăng cường nhận thức về an toàn trực tuyến với chi phí thấp và hiệu quả cao. Một số chiến dịch truyền thông nâng cao nhận thức cho cộng đồng cũng đã được thực hiện tại các nước khác trên thế giới, như Chiến dịch nâng cao nhận thức của Chính phủ Singapore[54], Chiến dịch nâng cao nhận thức của Chính phủ Hoa Kỳ[55].

[54] CSA Singapore. Ra mắt chiến dịch nâng cao nhận thức về an ninh mạng quốc gia lần thứ hai: https://www.csa.gov.sg/news/press-releases/csa-launches-second-national-cybersecurity-awareness-campaign

[55] Cơ quan Phòng thủ mạng của Hoa Kỳ. Tháng nâng cao nhận thức về an ninh mạng quốc gia: https://www.dhs.gov/national-cyber-security-awareness-month

PHẦN III. KẾT LUẬN VÀ ĐỀ XUẤT

1. Kết luận

Kết quả nghiên cứu rà soát, đánh giá các cổng DLM tại Việt Nam cho thấy một số kết luận như sau:

Về nhu cầu sử dụng, khai thác dữ liệu mở

(i) DLM có vai trò quan trọng trong việc thúc đẩy GDĐT, chuyển đổi số, phát triển kinh tế số và xã hội số. Các tổ chức và cá nhân (người dùng) cho thấy có nhu cầu cao trong việc sử dụng DLM phục vụ cho mục đích kinh doanh, sản xuất, phân tích thị trường, nghiên cứu sản phẩm mới, nghiên cứu khoa học...

(ii) Người dùng mong muốn DLM được cung cấp dưới các định dạng dễ sử dụng như Excel, CSV và API, nhằm tối ưu hóa khả năng khai thác và phân tích dữ liệu. Đối với các cổng DLM, người dùng mong muốn có các tính năng quan trọng như mô tả dữ liệu chi tiết, giao diện trực quan, dễ dàng sử dụng, có tính năng phản hồi dữ liệu, có liên kết với cổng DLM quốc gia, có công cụ tìm kiếm nâng cao giúp tìm kiếm và kết nối dữ liệu hiệu quả hơn.

Về hiện trạng cung cấp dữ liệu mở

(i) Một số ưu điểm của các cổng DLM hiện hành ở Việt Nam là đã cho phép người dùng có thể khai thác dữ liệu mọi lúc, mọi nơi và bất kỳ ai đều có thể truy cập và tải về miễn phí, có khả năng truy cập ổn định trên các thiết bị điện tử; về giao diện, các cổng DLM được thiết kế tối giản, dễ nhìn và thân thiện với người dùng, tương thích với màn hình máy tính; đã tích hợp một số tính năng cơ bản như quản lý danh mục (tìm kiếm, lọc dữ liệu), quản lý và lưu trữ dữ liệu (tải xuống dữ liệu, truy cập API), chức năng cộng đồng (bình luận, đánh giá, chia sẻ trên mạng xã hội), quản lý tài khoản (cá nhân hóa tài khoản người dùng).

(ii) Cổng DLM vẫn còn những hạn chế, chưa đáp ứng nhu cầu của người dùng như: *Chức năng, tính năng kĩ thuật trên các cổng DLM chưa đồng đều,* phần lớn các cổng chưa khai thác hiệu quả tiềm năng của API, nhiều cổng chưa cho phép người dùng gửi yêu cầu dữ liệu, không cung cấp bất kỳ cơ chế phản hồi về bộ dữ liệu, hiện không có cổng nào cung cấp ví dụ tái sử dụng dữ liệu.

(iii) *Số lượng và chất lượng dữ liệu được cung cấp trên các cổng DLM còn ít và chưa đáp ứng theo kế hoạch.* Nhiều cổng chủ yếu cung cấp dữ liệu dưới định dạng PDF, điều này làm giảm khả năng sử dụng của dữ liệu, vì định dạng này không thể dễ dàng đọc hoặc phân tích tự động. Nhiều cổng DLM hiện nay vẫn sử dụng dữ liệu cũ và không được cập nhật thường xuyên, dẫn đến việc chất lượng thông tin không còn chính xác và thiếu tính thời sự, không đáp ứng được nhu cầu sử dụng dữ liệu mới và chính xác để phục vụ cho nghiên cứu, ra quyết định hay các hoạt động phân tích mặc dù theo quy định. Các cổng DLM hiện nay đã bước đầu cung cấp các mô tả về dữ liệu. Tuy nhiên, các trường dữ liệu trong lớp Dataset chưa đáp ứng được yêu cầu của Tiêu chuẩn DCAT-VN 14171:2024 về Đặc tả tập DLM cho cổng dữ liệu.

(iv) *Nhiều cổng DLM chưa đảm bảo tính bảo mật, làm lộ lọt dữ liệu cá nhân:* Điều 20 Nghị định 47/2020/NĐ-CP yêu cầu không tiết lộ thông tin cá nhân khi công bố DLM của CQNN. Việc lộ lọt thông tin cá nhân gây nguy cơ rủi ro bảo mật và vi phạm quyền riêng tư của công dân. Tuy nhiên, theo rà soát, có 18/34 cổng DLM hiện cung cấp dữ liệu chưa đảm bảo tính bảo mật, làm lộ dữ liệu cá nhân.

Về các chính sách, quy định hiện hành liên quan đến dữ liệu mở

(i) DLM đã được nhắc đến trong hai văn bản luật là Luật GDĐT năm 2023 và Luật Dữ liệu năm 2024. Tuy nhiên, hiện nay chưa có một văn bản quy phạm dưới luật (Nghị định, Thông tư) quy định riêng đối với DLM. Mặc dù Nghị định 47/2020/NĐ-CP (có hiệu lực thi hành từ ngày 25/5/2020) đã có một số quy định về DLM nhưng chưa đầy đủ, các cơ quan, địa phương vẫn xem cung cấp DLM là một nội dung mới, còn lúng túng trong triển khai thực hiện.

Official Use Only

(ii) Trong các văn bản pháp luật hiện nay có một số khái niệm liên quan đến dữ liệu và DLM như "thông tin", "dữ liệu", "dữ liệu số", "dữ liệu mở", "dữ liệu mở của cơ quan nhà nước", "dữ liệu dùng chung", "dữ liệu dùng riêng", "dữ liệu gốc"… DLM được hiểu là một loại thông tin của CQNN (theo Luật tiếp cận thông tin). Theo quy định, thông tin của CQNN phải được người có thẩm quyền của CQNN ký, đóng dấu hoặc xác nhận bằng văn bản. Điều này lý giải cho hiện trạng trên cổng DLM của nhiều địa phương có nhiều tệp dữ liệu định dạng PDF được chụp từ các văn bản của CQNN với chữ ký, con dấu của người có thẩm quyền.

(iii) Hiện nay có ba văn bản quy phạm pháp luật đưa ra định nghĩa về DLM (Nghị định 47/2020/NĐ-CP, Luật GDĐT năm 2023 và Luật Dữ liệu năm 2024), các định nghĩa không hoàn toàn thống nhất (và chỉ áp dụng trong phạm vi văn bản đó) nhưng có điểm chung là: (i) DLM là "dữ liệu số" (dữ liệu "được biểu diễn bằng tín hiệu số" hay "dùng tín hiệu số" hay "được thể hiện dưới dạng kĩ thuật số") và (ii) mọi cơ quan, tổ chức, cá nhân được tiếp cận, sử dụng miễn phí. Một số đặc điểm quan trọng của DLM về định dạng máy đọc được hay không chứa thông tin cá nhân hiện chưa được nhắc đến khi giải thích thuật ngữ và chưa thể chế hóa các quy định cụ thể hoặc có hướng dẫn để đảm bảo các yêu cầu này được thực hiện.

(iv) Quy định hiện nay về cổng dữ liệu và cổng DLM còn rải rác trong nhiều văn bản (Nghị định 47/2020/NĐ-CP, Nghị định 42/2022/NĐ-CP, Tiêu chuẩn TCVN 14171:2024) mà chưa có hướng dẫn cụ thể về tiêu chuẩn kĩ thuật để xây dựng cổng DLM cấp bộ, cấp tỉnh, dẫn đến tình trạng mỗi cổng thông tin của các địa phương được thiết kế và triển khai theo những cách thức khác nhau, gây khó khăn trong việc tạo dựng hệ thống đồng nhất trên phạm vi quốc gia.

(v) Một số hướng dẫn về quy chuẩn, tiêu chuẩn, quy định kĩ thuật kết nối, chia sẻ DLM hiện được thực hiện thông qua văn bản hành chính (các công văn của Bộ TT&TT), không phải là quy phạm pháp luật do đó tính tuân thủ và hiệu quả thực thi không cao. Mặt khác, nhiều hướng dẫn đối với việc kết nối, chia sẻ dữ liệu nói chung, không dành riêng cho DLM dẫn đến tình trạng DLM được các bộ, địa phương cung cấp hiện nay dưới dạng máy không đọc được hay có chứa thông tin cá nhân tương đối phổ biến.

(vi) Quy trình các bước chuẩn bị dữ liệu cho đến kết nối, chia sẻ DLM chưa được quy định chi tiết trách nhiệm của các bên liên quan, cụ thể là giữa cơ quan chủ quản dữ liệu và cơ quan cung cấp dữ liệu.

(vii) Chưa có quy định về việc kiểm tra, đánh giá, duy trì DLM. Quy định hiện hành áp dụng đối với trường hợp dữ liệu do CQNN chia sẻ nói chung, không phù hợp với DLM. Theo quy định hiện hành, chủ quản CSDL có trách nhiệm tự tổ chức kiểm tra đánh giá định kỳ dữ liệu và gửi báo cáo về đơn vị chuyên trách về CNTT, còn các đơn vị chuyên trách về CNTT có trách nhiệm hướng dẫn chủ quản CSDL tự kiểm tra, đánh giá dữ liệu. Như vậy, trong trường hợp DLM, khi cơ quan chủ quản CSDL cung cấp DLM cho đơn vị chuyên trách về CNTT để chia sẻ cho tổ chức, cá nhân sử dụng trên cổng DLM thì việc cơ quan chủ quản CSDL có trách nhiệm kiểm tra, đánh giá định kỳ theo quy định là không phù hợp, đòi hỏi cần có quy định riêng về trách nhiệm kiểm tra, đánh giá việc cung cấp, chia sẻ DLM của các cơ quan, đơn vị liên quan.

(viii) Chưa có quy định về trình tự xử lý vướng mắc trong cung cấp, khai thác và sử dụng DLM.

Một số khó khăn trong triển khai thực hiện cung cấp dữ liệu mở

Ngoài việc hệ thống quy định pháp luật về DLM chưa đầy đủ khiến việc triển khai thực hiện của các CQNN ở trung ương và địa phương gặp nhiều khó khăn, thách thức, các CQNN còn gặp một số khó khăn, thách thức sau:

(i) Cơ sở hạ tầng CNTT nhằm đáp ứng các yêu cầu phức tạp trong kĩ thuật xử lý dữ liệu đa miền và hạ tầng công nghệ chưa được đầu tư đúng mức. Việc tích hợp, xử lý và chuẩn hóa các loại dữ liệu khác nhau để có thể khai thác hiệu quả đặt ra một thách thức lớn cho các đơn vị thiếu cơ sở hạ tầng CNTT. Trên thực tế, hệ thống lưu trữ và tính toán hiện nay chưa đáp ứng được nhu cầu cung cấp DLM của cả quốc gia.

Official Use Only

(ii) Hiện nay, các nguồn kinh phí cho hoạt động về DLM toàn bộ được lấy từ ngân sách địa phương vì vậy còn rất hạn chế và có sự khác biệt giữa các địa phương khác nhau, nguồn xã hội hóa cho DLM chưa được thúc đẩy. Quy trình chi tiêu tài chính vẫn chưa rõ ràng, chính sách mua sắm hiện tại còn nhiều bất cập hạn chế khả năng cung cấp các sản phẩm, dịch vụ có chất lượng tốt hơn nếu áp dụng các tiêu chuẩn về mua sắm hiện hành.

(iii) Thiếu nhân lực chuyên môn, đặc biệt là các chuyên gia kĩ thuật và chuyên gia về dữ liệu như kĩ sư dữ liệu (data engineer). Đây là những vị trí then chốt trong việc xây dựng và vận hành cổng DLM, nhưng nguồn nhân lực hiện tại chưa đáp ứng được yêu cầu cả về số lượng lẫn chất lượng.

(iv) Nhận thức về DLM còn hạn chế là một thách thức làm chậm tiến trình phát triển DLM tại Việt Nam. Phần lớn các CQNN, TCDN và người dùng cá nhân đã biết về các lợi ích của DLM nhưng vẫn hiểu sai về DLM, gồm khái niệm, định dạng, khả năng truy cập và cách thức cung cấp.

2. Đề xuất

DLM được xác định theo tính chất chia sẻ, trong đó quan trọng nhất là tất cả mọi cơ quan, tổ chức, cá nhân có nhu cầu đều được tiếp cận, chia sẻ, khai thác, sử dụng miễn phí. DLM không chỉ do CQNN mà còn có thể do các tổ chức khác cung cấp, cần khuyến khích các cơ quan, tổ chức tạo lập dữ liệu, chủ sở hữu dữ liệu chia sẻ DLM nhằm phát triển kinh tế số và xã hội số. Trong đó, để thực hiện Luật tiếp cận thông tin, Luật GDĐT, Luật Dữ liệu, các CQNN phải cung cấp DLM theo đó, cần xây dựng và ban hành một văn bản dưới luật riêng cho việc cung cấp DLM của CQNN dưới hình thức Nghị định, khắc phục các vấn đề như đã được nhận diện và kết luận ở trên.

Một số đề xuất cụ thể nhằm thúc đẩy việc cung cấp và sử dụng DLM ở Việt Nam như sau:

Bước 1: Cải thiện khung pháp lý và chính sách liên quan đến DLM

- Các quy định cần phân biệt rõ các loại dữ liệu chia sẻ giữa các CQNN, dữ liệu chia sẻ trong nội bộ một CQNN và DLM, dựa trên quyền truy cập của người dùng và các đặc điểm chia sẻ, kèm theo các tiêu chuẩn kĩ thuật và nội dung tương ứng[56].
- Cần quy định định dạng chuẩn, máy đọc được và đặc tả dữ liệu cho DLM (ví dụ như XML, CSV và JSON).
- Yêu cầu thiết lập cơ cấu quản trị dữ liệu ở cấp địa phương, ví dụ như nhóm công tác dữ liệu mở do lãnh đạo đứng đầu, bao gồm các cán bộ lãnh đạo, cán bộ phụ trách CNTT và cán bộ quản lý dữ liệu. Nhóm này có nhiệm vụ báo cáo cho lãnh đạo cấp cao về lợi ích của DLM và việc chia sẻ dữ liệu công nhằm thúc đẩy sự ủng hộ từ trên xuống.
- Thiết lập cơ chế tài chính rõ ràng: Thực hiện một cơ chế tài chính minh bạch và cụ thể ở cấp trung ương nhằm giảm khoảng cách giữa các tỉnh. Trong đó, cần quy định mức phân bổ ngân sách tối thiểu cho các sáng kiến về DLM ở cấp địa phương, đặc biệt là tại các tỉnh khó khăn, nhằm đảm bảo có đủ nguồn lực thực hiện hiệu quả các nhiệm vụ liên quan đến DLM và thúc đẩy sự phát triển cân bằng.
- Quy trình cung cấp DLM cần được quy định rõ ràng, với việc phân công vai trò và trách nhiệm cụ thể cho từng bên liên quan. Quy trình này cần nêu rõ nghĩa vụ của các đơn vị sở hữu dữ liệu trong việc đảm bảo tính chính xác, quyền riêng tư và tuân thủ quy định pháp luật. Đồng thời, cần xác định nhiệm vụ giám sát của các cơ quan quản lý trong quá trình xử lý dữ liệu, tiếp nhận phản hồi từ người dùng và kiểm tra, giám sát. Cũng cần ban hành hướng dẫn cụ thể cho người dùng và cơ chế xử lý khi có vấn đề phát sinh. Cần thiết lập hệ thống giám sát, kiểm tra và báo cáo nghiêm ngặt để duy trì chất lượng DLM. Cần thành lập đội ngũ chuyên trách để thực hiện đánh giá định kỳ và tổ chức thu thập ý kiến nhằm thúc đẩy cải tiến liên tục.

Bước 2: Kiểm kê và các bộ dữ liệu cần ưu tiên

[56] Tham khảo định nghĩa về DLM của các nước trên thế giới, bao gồm Pháp, Mỹ, Úc, Anh, Hàn Quốc.

Official Use Only

- Cần ban hành danh mục DLM thống nhất của cả nước để đảm bảo tính nhất quán của DLM và khả năng kết nối giữa các hệ thống[57]. Danh mục này phải quy định rõ ràng về cấu trúc dữ liệu, các trường thông tin và tiêu chuẩn định dạng tối thiểu nhằm đảm bảo khả năng kết nối, tính thống nhất và khả năng sử dụng của dữ liệu.
- Danh mục DLM nên được xây dựng, áp dụng và triển khai theo hồ sơ ứng dụng DCAT cho tất cả các cổng DLM tại Việt Nam, đồng thời tận dụng định dạng dữ liệu liên kết mở RDF.
- Như báo cáo đã chỉ ra, nhiều địa phương hiện đã xây dựng danh mục DLM, tuy nhiên điều này là chưa đủ – mà chỉ nên được coi là điểm khởi đầu. Cần rà soát lại danh mục để xác định những tập dữ liệu có thể công khai ngay (dữ liệu không nhạy cảm, được công chúng quan tâm) và những tập dữ liệu chỉ chia sẻ nội bộ (có tính nhạy cảm nhưng cần thiết cho việc phối hợp giữa các cơ quan).
- Tập trung cung cấp các bộ dữ liệu đã được công khai dưới một hình thức nào đó (ví dụ báo cáo trên website) nhưng chưa có định dạng DLM - các dữ liệu này có thể nhanh chóng được chuyển đổi sang định dạng CSV/Excel và công bố.

Ưu tiên các bộ dữ liệu phù hợp với danh mục DLM quốc gia (nếu đã được ban hành) hoặc những bộ dữ liệu được yêu cầu thường xuyên, chẳng hạn như dữ liệu về thủ tục hành chính, thông tin địa chính,... Danh mục DLM nên đề xuất các chủ đề để phân loại những bộ dữ liệu theo các lĩnh vực, xây dựng lịch trình công bố dữ liệu, ví dụ: đặt mục tiêu "Mỗi quý công bố X tập dữ liệu mới". Bắt đầu công bố với số lượng vừa phải, có thể từ 10–20 bộ dữ liệu trọng điểm trong giai đoạn đầu để tạo đà triển khai hiệu quả.**Bước 3: Phát triển cổng DLM (hoặc tích hợp với hệ thống hiện có) kết nối với trung tâm dữ liệu và/hoặc hệ thống trao đổi dữ liệu**

- Đối với các địa phương chưa có cổng DLM, cần lựa chọn nền tảng triển khai và bắt buộc sử dụng phần mềm mã nguồn mở như CKAN hoặc DKAN. Cần xây dựng lộ trình kĩ thuật theo giai đoạn để phát triển cổng theo khung tiêu chuẩn.
- Đảm bảo cổng DLM có các tính năng thiết yếu: tìm kiếm và duyệt các bộ dữ liệu, tải về ở nhiều định dạng, truy cập API cơ bản cho người dùng nâng cao, hiển thị dữ liệu đặc tả và biểu mẫu liên hệ/phản hồi.
- Quy định các tiêu chuẩn kĩ thuật cho cổng DLM ở các cấp, đảm bảo tính thống nhất và khả năng tương tác. Đồng thời, cần phù hợp với tiêu chuẩn kĩ thuật quốc gia để phục vụ việc tích hợp sau này.
- Đối với những địa phương đã có cổng nhưng hoạt động kém hiệu quả (ví dụ: tốc độ chậm, thiếu tính năng), cần có kế hoạch nâng cấp như: cải thiện giao diện người dùng, bổ sung điểm truy cập API và khắc phục các liên kết không hoạt động. Trước khi triển khai chính thức nên thử nghiệm với một nhóm người dùng nhỏ (ví dụ: sinh viên địa phương) để đảm bảo khả năng sử dụng của cổng.
- Với các cơ sở dữ liệu dùng nội bộ trong một CQNN, nếu địa phương đã có hoặc đang xây dựng kho dữ liệu tập trung (như trung tâm tích hợp dữ liệu tỉnh), cần kết nối với cổng DLM để tự động công bố những dữ liệu không nhạy cảm. [58]
- Đề xuất thành lập trung tâm dữ liệu hoặc hệ thống trao đổi dữ liệu ở cấp tỉnh để các sở ban ngành có thể tải dữ liệu lên trước khi đưa ra cổng DLM. Một số địa phương đã có quy trình tương tự, ví dụ Đà Nẵng và Thành phố Hồ Chí Minh, trong đó dữ liệu được đưa vào kho dữ liệu dùng chung trước khi công bố.

Bước 4: Triển khai quy trình quản lý dữ liệu, cụ thể là khi cổng được đưa vào hoạt động, thiết lập quy trình để duy trì vận hành

- *Định kỳ cập nhật dữ liệu*: Phối hợp với các đơn vị để cập nhật định kỳ theo từng loại dữ liệu (hằng tháng, quý, năm). Tự động hóa cập nhật khi có thể (qua API hoặc kịch bản xuất dữ liệu). Với trường hợp cập nhật thủ công, cần gửi thông báo đôn đốc các đơn vị và theo dõi việc thực hiện.

[57] Tại một số nước trên thế giới, gồm Pháp, Ba Lan, Estonia đã ban hành các danh mục dữ liệu mở bắt buộc trong luật pháp quốc gia phù hợp với Chỉ thị về Dữ liệu mở của Liên minh Châu Âu và khái niệm về các bộ dữ liệu có giá trị cao (như đã nêu tại mục 1.3)
[58] Tham khảo kĩ thuật ẩn danh nhằm bảo mật thông tin cá nhân theo Quy định bảo vệ dữ liệu chung 2016/679 của Châu Âu

- *Kiểm soát chất lượng*: Trước khi công bố, đơn vị phụ trách công khai DLM cần kiểm tra định dạng, tính đầy đủ và tính bảo mật của dữ liệu. Áp dụng các kĩ thuật ẩn danh cơ bản: loại bỏ tên cá nhân, tổng hợp dữ liệu cá nhân nếu cần thiết (công bố các kết quả thống kê thay vì dữ liệu cá nhân gốc). Đảm bảo hiển thị đầy đủ dữ liệu đặc tả (ngày, nguồn, liên hệ…).
- *Phản hồi người dùng*: Theo dõi các kênh phản hồi. Khi người dùng yêu cầu một bộ dữ liệu hoặc báo lỗi, phải phản hồi kịp thời. Ngay cả khi chưa thể cung cấp dữ liệu ngay lập tức, vẫn cần ghi nhận và lưu lại các yêu cầu của người dùng, từ đó xây dựng lòng tin từ người dùng. Có thể thiết lập một địa chỉ email riêng hoặc một địa ctrang Facebook để công chúng thảo luận về DLM - việc tiếp cận người dùng sẽ thu được nhiều ý kiến hữu ích để cải thiện hệ thống.
- *Kết nối với cổng dữ liệu quốc gia*: Thường xuyên đồng bộ với cổng quốc gia (data.gov.vn). Tận dụng tính năng tự động tải dữ liệu địa phương lên cổng quốc gia (nếu có). Các cơ quan/địa phương cần phân công cán bộ phụ trách đảm bảo dữ liệu của mình được hiển thị/trích dẫn trên cổng quốc gia để tăng khả năng tiếp cận.

Bước 5: Tăng cường sử dụng cơ sở dữ liệu dùng chung trong nội bộ

- Cải thiện việc chia sẻ dữ liệu nội bộ song song với việc công bố DLM.
- Tiếp cận từng bước: Bắt đầu kết nối một số CSDL chính của các đơn vị (sở, ban ngành) với trung tâm dữ liệu/hệ thống trao đổi dữ liệu. Sau đó kết nối các hệ thống CSDL trung tâm đến trung tâm dữ liệu/hệ thống chia sẻ dữ liệu của tỉnh/thành phố.
- Tập trung vào kết quả cụ thể: nếu việc chia sẻ dữ liệu liên sở, ban, ngành giúp giải quyết một vấn đề cụ thể (chẳng hạn rút ngắn thời gian xử lý một thủ tục xuống vài ngày) cần ghi nhận và báo cáo kết quả này, làm căn cứ để các đơn vị nhận thức được về giá trị của việc giải quyết các vấn đề an toàn kĩ thuật bằng cách phối hợp với bên cung cấp giải pháp hoặc chuyên gia an ninh mạng để đảm bảo rằng việc kết nối dữ liệu có thể thực hiện một cách an toàn. Từng bước xây dựng một môi trường CSDL liên thông theo mô hình phân quyền, trong đó - tùy theo phân quyền - các cơ quan có thể truy xuất thông tin cần thiết từ nhau. Cách làm này cũng sẽ giúp DLM trở nên phong phú và cập nhật hơn (vì dữ liệu trong các hệ thống CSDL dùng chung cũng có thể cung cấp trực tiếp cho các bộ DLM).

Bước 6: Nâng cao năng lực cho cán bộ

- Tổ chức tập huấn cho đội ngũ cán bộ và các đầu mối phụ trách tại các sở, ngành với hỗ trợ chuyên môn từ các bộ ngành trung ương hoặc các dự án tài trợ.
- Nội dung đào tạo có thể bao gồm: cách sử dụng phần mềm cổng DLM, cách chuẩn bị bộ dữ liệu (làm sạch và định dạng dữ liệu), những kiến thức cơ bản về bảo mật và quyền riêng tư (để biết phần nào cần được ẩn danh), cũng như cách sử dụng các công cụ phân tích dữ liệu (giúp cán bộ chủ động khai thác DLM trong công việc).
- Khuyến khích đội ngũ cán bộ CNTT tham gia các diễn đàn cấp quốc gia hoặc cộng đồng trực tuyến về DLM. Bố trí kinh phí cử cán bộ tham dự các hội thảo chuyên đề hoặc chương trình học hỏi quốc tế.
- Về lâu dài, nên xây dựng một nhóm "đại sứ dữ liệu" (data champions) trong nội bộ - có thể là các cán bộ trẻ, nhiệt huyết - để lan tỏa lợi ích của DLM đến đồng nghiệp, khuyến khích học hỏi giữa các đơn vị. Ví dụ, Cục Thống kê, Chi cục thống kê có kĩ năng dữ liệu tốt nên có thể huy động họ hỗ trợ các cơ quan khác trong quản lý và sử dụng dữ liệu.

Bước 7: Tăng cường gắn kết cộng đồng và đối tác

Để khai thác hiệu quả DLM, cần kết nối với hệ sinh thái ngoài nhà nước. Thông báo rộng rãi việc ra mắt cổng DLM qua truyền thông, mời các trường đại học, doanh nghiệp, tổ chức phi chính phủ cùng tham gia sử dụng và phản hồi.

Bước 8: Giám sát, đánh giá và điều chỉnh

- Thiết lập quy trình theo dõi tiến độ và điều chỉnh phù hợp. Theo dõi các chỉ số như: số lượng bộ dữ liệu đang được công bố (và cập nhật), lượng truy cập cổng dữ liệu, số lượt tải xuống hoặc sử dụng

Official Use Only

API cũng như phản hồi nhận được từ người dùng. Đối với những bộ dữ liệu ít được truy cập, cần tìm hiểu nguyên nhân (có thể do không thực sự hữu ích hoặc chưa được quảng bá tốt). Nếu có bộ dữ liệu được yêu cầu nhưng chưa cung cấp, cần lên kế hoạch để công bố.

- Báo cáo định kỳ cho lãnh đạo - ví dụ, có thể gửi báo cáo hàng quý cho UBND tỉnh/thành phố về tình hình cung cấp DLM (thể hiện sự tăng trưởng về số lượng hoặc mức độ sử dụng) nhằm duy trì sự quan tâm và ủng hộ từ cấp trên. So sánh với các địa phương khác (nếu có số liệu hoặc bảng xếp hạng), việc so sánh, đối chiếu giữa các địa phương ở mức độ có chừng mực có thể là động lực để các địa phương nỗ lực hơn.
- Cần phải xem đây là một quá trình cải tiến liên tục: thu thập ý kiến người dùng và rút kinh nghiệm. Nếu người dùng phản ánh chức năng tìm kiếm khó sử dụng, cần phải cải tiến. Bộ, ngành trung ương cần ban hành hướng dẫn mới (về chuẩn kĩ thuật, danh mục, các quy định chi tiết triển khai dữ liệu mở của Luật Dữ liệu v.v.), cụ thể, chi tiết và có cơ chế hướng dẫn, phân bổ kinh phí để hỗ trợ các địa phương cập nhật và áp dụng đồng bộ.

Bước 9: Bảo đảm an toàn và bảo mật dữ liệu xuyên suốt: Đây là bước xuyên suốt toàn bộ quá trình – luôn đảm bảo không có dữ liệu nhạy cảm nào vô tình bị công bố và các hệ thống dữ liệu dùng chung luôn được bảo vệ an toàn.

- Cập nhật quy trình ngay khi có hướng dẫn mới về quyền riêng tư dữ liệu.
- Trước khi mở dữ liệu, cần xác minh và ẩn danh dữ liệu cá nhân, chẳng hạn như số CMND/CCCD hoặc tên trong tập dữ liệu.
- Kiểm tra định kỳ các tập dữ liệu để đảm bảo không có thông tin cá nhân bị lộ lọt (nhất là sau những lần cập nhật bổ sung thông tin mới).
- Xây dựng quy trình phản ứng sự cố: nếu có ai đó phát hiện vấn đề về quyền riêng tư hoặc lỗ hổng bảo mật, cần xử lý nhanh chóng - tạm thời gỡ bộ dữ liệu, khắc phục sự cố rồi mới công bố lại.

Việc chủ động bảo vệ quyền riêng tư không chỉ giúp tăng niềm tin của người dân, mà còn tạo sự yên tâm cho các CQNN khi tham gia chia sẻ dữ liệu qua một quy trình DLM có biện pháp bảo vệ chắc chắn.

Phụ lục: Yêu cầu cơ bản về chức năng, tính năng kĩ thuật của cổng dữ liệu mở

I. CĂN CỨ PHÁP LÝ

- Luật Công nghệ thông tin số 67/2006/QH11 ngày 29/6/2006;

- Luật An toàn thông tin mạng số 86/2015/QH13 ngày 19/11/2015;

- Luật An ninh mạng số 24/2018/QH14 ngày 12/6/2018;

- Luật Giao dịch điện tử số 20/2023/QH15 ngày 22/06/2023;

- Căn cứ Nghị định số 64/2007/NĐ-CP ngày 10 tháng 4 năm 2007 của Chính phủ về ứng dụng công nghệ thông tin trong hoạt động của cơ quan Nhà nước;

- Căn cứ Nghị định số 85/2016/NĐ-CP ngày 01 tháng 7 năm 2016 của Chính phủ về bảo đảm an toàn hệ thống thông tin theo cấp độ;

- Nghị định số 47/2020/NĐ-CP ngày 09/4/2020 của Chính phủ về quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước;

- Nghị định số 42/2022/NĐ-CP ngày 24/6/2022 của Chính phủ quy định về việc cung cấp thông tin và dịch vụ công trực tuyến của cơ quan nhà nước trên môi trường mạng;

- Nghị định số 47/2024/NĐ-CP ngày 09/04/2024 của Chính phủ quy định về danh mục cơ sở dữ liệu quốc gia; việc xây dựng, cập nhật, duy trì, khai thác và sử dụng cơ sở dữ liệu;

- Nghị định số 13/2023/NĐ-CP ban hành ngày 17/04/2023 của Chính phủ về Bảo vệ dữ liệu cá nhân;

- Quyết định của Thủ tướng Chính phủ số 749/QĐ-TTg ngày 3/6/2020 phê duyệt "Chương trình Chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030";

- Quyết định số 942/QĐ-TTg ngày 15/6/2021 phê duyệt Chiến lược phát triển Chính phủ điện tử hướng tới Chính phủ số giai đoạn 2021 - 2025, định hướng đến năm 2030;

- Quyết định số 411/QĐ-TTg ngày 31/3/2022 phê duyệt Chiến lược quốc gia phát triển kinh tế số và xã hội số đến năm 2025, định hướng đến năm 2030;

- Quyết định 142/QĐ-TTg ngày 02/02/2024 phê duyệt Chiến lược dữ liệu quốc gia đến năm 2030.

- Thông tư 22/2023/TT-BTTTT ngày 31/12/2023 của Bộ Thông tin và Truyền thông quy định về cấu trúc, bố cục, yêu cầu kĩ thuật cho cổng thông tin điện tử của cơ quan nhà nước;

- Quyết định số 2323/QĐ-BTTTT ngày 31/12/2019 của Bộ trưởng Bộ Thông tin và Truyền thông ban hành Khung kiến trúc Chính phủ điện tử Việt Nam phiên bản 2.0;

- Quyết định số 186/QĐ-BTTTT ngày 11/2/2022 của Bộ trưởng Bộ Thông tin và Truyền thông phê duyệt Chương trình thúc đẩy phát triển và sử dụng các nền tảng số quốc gia phục vụ chuyển đổi số, phát triển Chính phủ số, kinh tế số, xã hội số.

II. THÔNG TIN CHUNG

1. Mục đích ban hành

Nhằm mục đích thúc đẩy việc công bố dữ liệu mở và để việc xây dựng và phát triển Cổng dữ liệu mở phù hợp với định hướng chung, thuận tiện trong việc tích hợp với Cổng dữ liệu mở quốc gia, cũng như đảm bảo việc cung cấp nhiều tính năng, thuận tiện cho người sử dụng, Bộ Thông tin và Truyền thông ban hành tài liệu hướng dẫn các yêu cầu cơ bản về chức năng, tính năng kĩ thuật của Cổng dữ liệu mở.

Chức năng, tính năng kĩ thuật trong hướng dẫn này là cơ bản và tối thiểu, các bộ, ngành, địa phương căn cứ điều kiện thực tế và các nhu cầu đặc thù có thể xem xét, bổ sung, mở rộng hoặc lựa chọn triển khai các yêu cầu theo mức độ ưu tiên phù hợp.

Official Use Only

2. Đối tượng áp dụng

(i) Các Bộ, cơ quan ngang Bộ, cơ quan thuộc Chính phủ (sau đây gọi là các bộ), Ủy ban nhân dân các tỉnh, thành phố trực thuộc Trung ương (sau đây gọi là các tỉnh)

(ii) Các tổ chức, cá nhân có liên quan trong việc cung cấp giải pháp, xây dựng, phát triển, vận hành Cổng dữ liệu mở tại các Bộ và các tỉnh, thành phố.

3. Giải thích từ ngữ

(i) Cổng dữ liệu mở là một cổng thành phần thuộc Cổng dữ liệu cấp bộ, cấp tỉnh. Theo Khoản 3 Điều 3 Nghị định số 42/2022/NĐ-CP ngày 24/6/2022 của Chính phủ quy định về việc cung cấp thông tin và dịch vụ công trực tuyến của cơ quan nhà nước trên môi trường mạng, Cổng dữ liệu cấp bộ, cấp tỉnh là *"điểm đầu mối truy cập thông tin, dữ liệu trên môi trường mạng phục vụ việc công bố dữ liệu mở, cung cấp thông tin về chia sẻ dữ liệu của cơ quan nhà nước trong nội bộ của bộ, tỉnh; cung cấp các tài liệu, dịch vụ, công cụ, ứng dụng xử lý, khai thác dữ liệu được công bố bởi các cơ quan nhà nước trong nội bộ của bộ, tỉnh"*.

(ii) Dữ liệu mở được giải thích tại Khoản 4 Điều 3 Nghị định số 47/2020/NĐ-CP ngày 09/4/2020 của Chính phủ về quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước: "*Dữ liệu mở của cơ quan nhà nước là dữ liệu được cơ quan nhà nước có thẩm quyền công bố rộng rãi cho cơ quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ.*"

(iii) Các từ ngữ/ viết tắt khác trong tài liệu này được hiểu như sau:

Bảng A.1: Giải thích từ ngữ và chữ viết tắt

TT	Từ ngữ/ viết tắt	Giải thích
1	Người sử dụng	Là bất kỳ người nào truy cập cổng để tìm, truy cập và sử dụng dữ liệu.
2	Người xuất bản	Được hiểu là cán bộ cơ quan nhà nước chịu trách nhiệm cung cấp hoặc cập nhật dữ liệu và nội dung khác trên cổng.
3	Quản trị viên	Được hiểu là cán bộ được giao quản lý các tùy chọn kĩ thuật và cấu hình cho cổng, ví dụ như giao diện trực quan, yêu cầu đối với dữ liệu đặc tả (metadata)…
4	Phải/ Có thể	Phải: Biểu thị một yêu cầu quan trọng đối với việc triển khai thành công cổng dữ liệu mở. Nếu yêu cầu này không được đáp ứng và tuân thủ sẽ làm ảnh hưởng lớn đến hiệu quả của cổng dữ liệu mở. Có thể: Biểu thị một yêu cầu là tùy chọn, cần có, nên có nhưng không bắt buộc.
5	Mức độ ưu tiên 1, 2, 3	Được sắp xếp theo thứ tự giảm dần của mức độ ưu tiên, mức độ ưu tiên 1 là mức độ ưu tiên cao nhất cần triển khai.
6	Danh mục dữ liệu	Là một công cụ được sử dụng để sắp xếp, quản lý và khám phá các dữ liệu trong một tổ chức. Công cụ này như một kho trung tâm lưu trữ dữ liệu đặc tả của các dữ liệu, ví dụ như nguồn, định dạng, chất lượng dữ liệu...
7	Định dạng tập dữ liệu	Mô tả cách một tập dữ liệu được thể hiện trên đĩa máy tính. Định dạng thường tương ứng với phần cuối cùng của tên (gọi là phần mở rộng), ví dụ 1 tập dữ liệu ở định dạng CSV được thể hiện như danhsach.csv.
8	Định dạng độc quyền	Là định dạng mà một công ty sở hữu và kiểm soát. Dữ liệu ở định dạng độc quyền có thể phải cần phần mềm độc quyền để được đọc một cách đáng tin cậy.
9	Định dạng mở	Là định dạng của tập tin hoặc thông điệp dữ liệu được quy định theo công bố kĩ thuật của tổ chức hoạt động về tiêu chuẩn và không bị hạn chế khi sử dụng công bố kĩ thuật này.

Official Use Only

TT	Từ ngữ/ viết tắt	Giải thích
10	Định dạng máy có thể đọc được	Là định dạng của dữ liệu mà máy tính có thể tự động đọc và xử lý mà không cần có thêm phần mềm chuyên nghiệp nào, ví dụ như CSV , JSON , XML , v.v. Dữ liệu máy có thể đọc được phải là dữ liệu có cấu trúc - là dữ liệu có định dạng chuẩn hóa thường ở dạng bảng, bao gồm các hàng và cột xác định rõ ràng các thuộc tính dữ liệu (ví dụ: bảng dữ liệu có cấu trúc về khách hàng bao gồm các cột tên, địa chỉ, tổng số khách hàng, … ngược lại, dữ liệu phi cấu trúc, chẳng hạn như danh sách các bài đăng trên mạng xã hội, sẽ khó phân tích)
11	Dữ liệu đặc tả	"Dữ liệu đặc tả là những thông tin mô tả các đặc tính của dữ liệu như nội dung, định dạng, chất lượng, điều kiện và các đặc tính khác nhằm tạo thuận lợi cho quá trình tìm kiếm, truy cập, quản lý và lưu trữ dữ liệu" (khoản 11 Điều 3 Nghị định 42/2022/NĐ-CP)
12	Giấy phép sử dụng dữ liệu	Là một thỏa thuận nhằm cho phép người sử dụng quyền chia sẻ, sửa đổi và sử dụng dữ liệu. Trong cơ quan nhà nước, việc sử dụng dữ liệu mở phải đáp ứng các điều kiện được quy định về việc sử dụng tại Điều 18 Nghị định 47/2020/NĐ-CP như sau: Được tự do sử dụng, chia sẻ và kết hợp dữ liệu mở, kể cả trong các sản phẩm thương mại hoặc phi thương mại; Phải trích dẫn và ghi nhận khi sử dụng dữ liệu mở trong sản phẩm hoặc tài liệu liên quan; Nguyên trạng, không bao gồm thông tin phát sinh hoặc trình bày khác; Không được bán dữ liệu mở nguyên trạng của cơ quan nhà nước và phải cung cấp miễn phí dữ liệu mở đi kèm sản phẩm thương mại sử dụng dữ liệu này. Cơ quan nhà nước cung cấp dữ liệu mở có thể ban hành các giấy phép sử dụng khác nhưng vẫn phải tuân thủ các quy định tại Điều 18 Nghị định Nghị định 47/2020/NĐ-CP.
13	Giao diện lập trình ứng dụng – API (Application Programming Interface)	Là thể hiện kĩ thuật của một dịch vụ chia sẻ dữ liệu trong hệ thống thông tin.
14	JSON - JavaScript Object Notation	Là một tiêu chuẩn về định dạng dữ liệu trong ngôn ngữ JavaScript.
15	XML - eXtensible Markup Language	Là tiêu chuẩn về định dạng dữ liệu trong máy tính có tên là ngôn ngữ đánh dấu mở rộng.

III. YÊU CẦU CƠ BẢN VỀ CHỨC NĂNG, TÍNH NĂNG KĨ THUẬT CỦA CỔNG DỮ LIỆU MỞ

1. Các yêu cầu chung đối với Cổng dữ liệu mở cấp bộ, cấp tỉnh

Tuân thủ quy định tại Điều 7 Nghị định 47/2024/NĐ-CP xác định mối quan hệ giữa CSDL quốc gia với CSDL của bộ, ngành, địa phương, Cổng dữ liệu mở là một cổng thành phần thuộc Cổng dữ liệu cấp bộ, cấp tỉnh. Mối quan hệ giữa Cổng dữ liệu mở của các Bộ, ngành, địa phương và Cổng dữ liệu mở quốc gia được đề cập tại Mục 5 nhằm đảm bảo khả năng kết nối, chia sẻ dữ liệu giữa các CSDL của các bộ, tỉnh và các CSDL quốc gia

Tuân thủ Nghị định số 42/2022/NĐ-CP ngày 24/6/2022 của Chính phủ quy định về việc cung cấp thông tin và dịch vụ công trực tuyến của cơ quan nhà nước trên môi trường mạng,"Cổng dữ liệu cấp bộ, cấp tỉnh là điểm đầu mối truy cập thông tin, dữ liệu trên môi trường mạng phục vụ việc công bố dữ liệu mở, cung cấp thông tin về chia sẻ dữ liệu của cơ quan nhà nước trong nội bộ của bộ, tỉnh; cung cấp các tài liệu, dịch vụ, công cụ, ứng dụng xử lý, khai thác dữ liệu được công bố bởi các cơ quan nhà nước trong nội bộ của bộ, tỉnh" do vậy Cổng dữ liệu của các Bộ, ngành, địa phương phải có thể truy cập được thông qua tên miền internet của cơ quan nhà nước (ví dụ: opendata.tentinhthanh.gov.vn hoặc data.tintinhthanh.gov.vn). Ngoài ra, Cổng cần cung cấp các thông tin về cơ quan chủ quản, bao gồm các thông tin cơ bản như email, số điện thoại, địa chỉ cơ quan cung cấp.

Official Use Only

Các Cổng dữ liệu mở nên được thiết kế giao diện thân thiện với người dùng và phải truy cập được trên các thiết bị màn hình lớn (máy tính để bàn, máy tính xách tay, ...) và trên các thiết bị màn hình nhỏ (điện thoại di động thông minh, máy tính bảng,…).

Các Cổng dữ liệu mở phải xây dựng được danh mục dữ liệu mở với các đặc điểm sau:

+ Tính dễ dàng trong tiếp cận: Danh mục dữ liệu mở giúp người sử dụng dễ dàng, nhanh chóng và thuận tiện trong việc tiếp cập các dữ liệu. Việc truy cập vào các dữ liệu mở không yêu cầu phải đăng ký hoặc đăng nhập vì những điều này sẽ tạo ra các rào cản và không khuyến khích việc khai phá và sử dụng dữ liệu mở.

+ Yêu cầu hội nhập: Cổng nên cung cấp nhiều phiên bản ngôn ngữ để thu hút nhiều người dùng hơn (bao gồm tiếng Việt, tiếng Anh và/hoặc nhiều ngôn ngữ khác).

+ Tính dễ dàng trong tìm kiếm dữ liệu: Danh mục dữ liệu mở giúp dễ dàng tìm kiếm dữ liệu. Các danh mục dữ liệu phải được sắp xếp theo chủ đề, cơ quan xuất bản hoặc loại dữ liệu cũng như hỗ trợ việc tìm kiếm theo từ khóa.

+ Tiếp cận các dữ liệu có thể đọc được bằng máy: Dữ liệu mở được cung cấp dưới nhiều định dạng khác nhau và khả dụng trong việc tải xuống ở định dạng số không độc quyền, có thể đọc được bằng máy mà không cần phải có phần mềm chuyên nghiệp ví dụ như các định dạng CSV, XLM, JSON.... Trong đó, PDF không phải là định dạng đọc được bằng máy do máy tính sẽ gặp khó khăn khi truy cập thông tin dạng bảng trong tệp PDF.

Đồng thời, cần phải có khả năng tải xuống tất cả các tập dữ liệu của một bộ dữ liệu bằng một lần tải xuống.

+ Trực quan hóa và xem trước các dữ liệu: Có khả năng xem trước dữ liệu trước khi tải xuống hoặc trực quan hóa các dữ liệu thông qua tính năng các đồ thị hoặc bản đồ được tích hợp sẵn.

+ Cung cấp các tin tức và các ứng dụng liên quan đến các bộ dữ liệu mở được công bố trên Cổng.

+ Đảm bảo tính cộng đồng: Cổng kết nối với các nền tảng xã hội, tạo môi trường để người dùng và nhà cung cấp dữ liệu mở có thể trao đổi thông tin.

+ Cá nhân hóa tài khoản người dùng: Cổng cho phép cá nhân hóa tài khoản người dùng, lưu trữ thông tin người dùng đã sử dụng.

+ Dịch vụ Giao diện lập trình ứng dụng (API): Cho phép các nhà phát triển ứng dụng có khả năng truy cập được đến các tập dữ liệu thông qua phần mềm. Đồng thời, tạo điều kiện thuận lợi trong việc khai phá, phân tích, tích hợp, thu thập dữ liệu đặc tả từ các trang web bên ngoài và thông qua các ứng dụng khác.

+ An toàn thông tin: Tiến hành các biện pháp an toàn thông tin để tránh bị thay đổi bởi người dùng trái phép đối với dữ liệu và dữ liệu đặc tả.

+ Đảm bảo không tiết lộ dữ liệu cá nhân và thông tin giúp xác định một con người cụ thể: Đảm bảo không tiết lộ dữ liệu cá nhân và thông tin giúp xác định một con người dưới dạng ký hiệu, chữ viết, chữ số, hình ảnh, âm thanh hoặc dạng tương tự trên môi trường điện tử gắn liền với một con người cụ thể hoặc giúp xác định một con người cụ thể (Điều 20 Nghị định số 47/2020/NĐ-CP ngày 09/4/2020 của Chính phủ về quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước và Điều 2 Nghị định 13/2023/NĐ-CP ngày 17/4/2023 của Chính phủ về Bảo vệ dữ liệu cá nhân).

+ Dữ liệu đặc tả của các tập dữ liệu mở tuân theo ***Tiêu chuẩn quốc gia TCVN 14171:2024 Đặc tả tập dữ liệu mở cho Cổng dữ liệu (DCAT-VN)***

2. Yêu cầu về chức năng

Yêu cầu về chức năng của tài liệu này được phân loại cho 3 đối tượng khác nhau, gồm: người sử dụng, người xuất bản và quản trị viên.

Cổng dữ liệu mở cấp bộ, cấp tỉnh gồm các nhóm chức năng cơ bản dành cho người sử dụng, nhà xuất bản, quản trị viên như sau:

Official Use Only

1. Chức năng quản lý danh mục tập dữ liệu
2. Chức năng quản lý và lưu trữ dữ liệu
3. Chức năng trình bày dữ liệu, xem trước dữ liệu
4. Chức năng cộng đồng
5. Chức năng quản lý tài khoản

Về cơ bản, "người xuất bản" có thể làm tất cả các nội dung mà "người sử dụng" có thể làm, cộng thêm một số nội dung (ghi rõ tại mục 2.2) và "quản trị viên" có thể làm tất cả các nội dung mà "người xuất bản" có thể làm, cộng thêm một số nội dung (ghi rõ tại mục 2.3).

2.1. Đối với người sử dụng

Người sử dụng Cổng dữ liệu mở cấp bộ, cấp tỉnh có thể thực hiện các chức năng cơ bản như sau:

Bảng A.2: Chức năng cơ bản đối với người sử dụng

STT	Nhóm chức năng	Mô tả chức năng	Mức độ ưu tiên
I	**CHỨC NĂNG QUẢN LÝ DANH MỤC TẬP DỮ LIỆU**		
I.1	Chức năng lọc, tìm kiếm dữ liệu	Phải có khả năng lọc được các bộ dữ liệu phân theo từng cơ quan cung cấp.	1
		Phải có khả năng lọc được các bộ dữ liệu phân theo định dạng của dữ liệu.	1
		Phải có khả năng lọc được các bộ dữ liệu phân theo các chủ đề đã được thiết lập.	1
		Phải có khả năng lọc được các bộ dữ liệu phân theo các giấy phép của các bộ dữ liệu.	1
		Phải có khả năng lọc được các bộ dữ liệu có cung cấp API, không có API.	1
		Phải có khả năng sắp xếp các bộ dữ liệu theo thứ tự với các đặc tính như: Tính cập nhật, Chữ cái, Ngày tạo, Mức độ liên quan, Tính phổ biến.	1
		Phải có khả năng hiển thị được tổng số lượng các bộ dữ liệu theo các tiêu chí lọc.	1
		Có thể tìm kiếm được những tập dữ liệu bằng những cụm từ có trong dữ liệu.	3
II	**CHỨC NĂNG QUẢN LÝ VÀ LƯU TRỮ DỮ LIỆU**		
II.1	Miễn phí truy cập và tải về không bắt buộc yêu cầu đăng ký, đăng nhập tài khoản	Cổng cung cấp tất cả các bộ dữ liệu và dịch vụ miễn phí Trong đó không có bất kỳ hạn chế nào theo các giấy phép mở	1
II.2	Cung cấp nhiều định dạng khác nhau	Phải cho phép người dùng tải xuống các tập dữ liệu theo các định dạng khác nhau.	1
II.3	Cung cấp dữ liệu ở định dạng máy đọc được	Phải cho phép người dùng tải xuống dữ liệu ở định dạng có thể đọc được bằng máy (1 trong 3 định dạng phổ biến sau: CSV, XML, JSON)	1
II.4	Cung cấp dịch vụ API	Nên cung cấp dịch vụ API cho người dùng (API được cung cấp để người dùng có thể truy cập vào cổng dữ liệu mở theo chương trình) và để tham chiếu các tập dữ liệu với cổng dữ liệu mở quốc gia (API được cung cấp để cổng dữ liệu mở của các bộ/tỉnh có thể tự động xuất bản lên cổng dữ liệu quốc gia)	2

Official Use Only

STT	Nhóm chức năng	Mô tả chức năng	Mức độ ưu tiên
II.5	Tính cập nhật của dữ liệu	Có thể xem được lịch sử của các tập dữ liệu thông qua các bảng nhật ký các thay đổi đi kèm.	2
		Có thể truy cập các phiên bản khác nhau của tập dữ liệu.	3
III	**CHỨC NĂNG TRÌNH BÀY DỮ LIỆU, XEM TRƯỚC DỮ LIỆU**	Phải có khả năng xem trước các tài nguyên dữ liệu dưới dạng bảng bằng trình duyệt internet.	1
		Có thể trực quan hóa các tài nguyên dữ liệu dưới dạng bản đồ	2
		Có thể trực quan hóa các tài nguyên dữ liệu dưới dạng biểu đồ.	3
		Có thể lựa chọn các thông số trong tập dữ liệu để đưa vào trình bày cho các biểu đồ, bản đồ để hiển thị thông tin phù hợp.	3
IV	**CHỨC NĂNG CỘNG ĐỒNG**		
IV.1	Tính năng đánh giá dữ liệu	Phải có khả năng cung cấp nhận xét, bình luận, phản hồi về các bộ dữ liệu đã được cung cấp.	1
		Có thể xếp hạng, đánh giá theo mức 1 đến 5 sao các bộ dữ liệu đã công bố.	2
IV.2	Tính năng cung cấp thông tin liên quan đến dữ liệu	Có thể sử dụng nguồn cấp tin tức về những cập nhật mới đối với từng tập dữ liệu mà người dùng quan tâm.	2
IV.3	Tính năng gửi yêu cầu	Người dùng có thể gửi yêu cầu đề xuất việc công bố các tập dữ liệu mới tới các cơ quan quản lý	2
		Có thể đề nghị các tài nguyên dữ liệu cần cung cấp công khai.	3
IV.4	Kết nối với các nền tảng xã hội	Cổng kết nối với các nền tảng xã hội, tạo môi trường để người dùng và nhà cung cấp dữ liệu mở có thể trao đổi thông tin.	3
IV.5	Cung cấp các ứng dụng dữ liệu	Có thể thêm đường liên kết đến các ứng dụng có liên quan đến các bộ dữ liệu.	3
		Có thể cung cấp các ví dụ về tái sử dụng dữ liệu mở (thực tế đã triển khai hoặc ý tưởng) được cung cấp.	2
IV.6	Cung cấp tin tức	Có thể xem các tin tức về những thông tin liên quan đến dữ liệu mở.	3
V	**CHỨC NĂNG QUẢN LÝ TÀI KHOẢN**		
V.1	Cá nhân hóa tài khoản người dùng	Phải cho phép cá nhân hóa tài khoản người dùng bằng cách đăng ký tài khoản bằng địa chỉ email.	1
		Người dùng có thể tạo, chỉnh sửa hoặc xóa thông tin hồ sơ mà những người sử dụng khác có thể truy cập được.	2

2.2. Đối với người xuất bản

Người xuất bản của Cổng dữ liệu mở cấp bộ, cấp tỉnh có thể thực hiện được tất cả các chức năng cơ bản của người sử dụng, cộng thêm một số chức năng khác như sau:

Bảng A.3: Chức năng cơ bản đối với người xuất bản

STT	Nhóm chức năng	Mô tả chức năng	Mức độ ưu tiên
I	**CHỨC NĂNG QUẢN LÝ DANH MỤC DỮ LIỆU**	Phải có khả năng thêm, chỉnh sửa hoặc xóa các mục nhập thông tin khi công bố bộ dữ liệu.	1
		Phải có khả năng cung cấp mã định danh duy nhất cho mỗi bộ dữ liệu được phát hành.	1
		Có thể đánh dấu được các mục nhập thông tin là riêng tư để các mục này không hiển thị công khai.	2
II		Phải có khả năng tạo các tài nguyên dữ liệu mới.	1

Official Use Only

STT	Nhóm chức năng	Mô tả chức năng	Mức độ ưu tiên
	CHỨC NĂNG QUẢN LÝ VÀ LƯU TRỮ DỮ LIỆU	Phải có khả năng thay thế hoặc cập nhật các tài nguyên dữ liệu hiện có.	1
III	**CHỨC NĂNG TRÌNH BÀY DỮ LIỆU**	Có thể xem được trình bày dữ liệu khi khởi tạo mới và gán loại trình bày dữ liệu (biểu đồ, bản đồ, bảng biểu) mặc định để tự động lựa chọn hiển thị khi người dùng xem tập dữ liệu này.	2
IV	**CHỨC NĂNG CỘNG ĐỒNG**	Phải có khả năng xem các tài nguyên dữ liệu mà người sử dụng đề nghị cung cấp.	1
		Có thể cập nhật trạng thái xử lý đề nghị của người sử dụng về cung cấp tài nguyên dữ liệu.	2
		Có thể phản hồi các tài nguyên dữ liệu do người sử dụng đề nghị.	2
		Có thể xem và trả lời nhận xét và phản hồi của người sử dụng trong từng tài nguyên dữ liệu.	2
		Có thể xem và phê duyệt việc hiển thị đối với các nhận xét và phản hồi của người sử dụng.	2
V	**CHỨC NĂNG QUẢN LÝ TÀI KHOẢN**	Có thể cấp quyền cho một nhóm người sử dụng cụ thể, hoặc đã đăng ký về khả năng truy cập một số tài nguyên dữ liệu cụ thể thuộc quyền sở hữu của người xuất bản.	2

2.3. Đối với quản trị viên

Quản trị viên của Cổng dữ liệu mở cấp bộ, cấp tỉnh thực hiện được tất cả các chức năng cơ bản của "người xuất bản" cộng thêm chức năng khác như sau:

Bảng A.4: Chức năng cơ bản đối với quản trị viên

STT	Nhóm chức năng	Mô tả chức năng	Mức độ ưu tiên
I	**CHỨC NĂNG QUẢN LÝ DANH MỤC DỮ LIỆU**	Phải có khả năng tạo, chỉnh sửa và loại bỏ các danh mục dữ liệu đặc tả	1
		Phải có khả năng thêm bớt xóa sửa các chủ đề dữ liệu.	1
		Phải có khả năng thêm bớt xóa sửa cơ quan cung cấp dữ liệu.	1
		Phải có khả năng xóa các tập dữ liệu (trong trường hợp người xuất bản cung cấp các tập dữ liệu không phù hợp).	1
		Phải có khả năng chỉnh sửa nội dung dữ liệu đặc tả trên các bộ dữ liệu đã xuất bản (trường hợp các dữ liệu đặc tả của các bộ dữ liệu có nội dung lỗi, sai).	1
		Phải có khả năng cấu hình chế độ sắp xếp mặc định cho danh mục dữ liệu.	1
		Phải có khả năng cấu hình cài đặt chung cho danh mục dữ liệu, bao gồm giao diện, logo, bộ nhận diện, tiêu đề.	1
		Có thể cấu hình các cài đặt bổ sung chung cho danh mục dữ liệu, bao gồm các yếu tố và bố cục hiển thị trang chủ, các thành phần danh mục được hiển thị,…	2
II	**CHỨC NĂNG CỘNG ĐỒNG**	Phải có khả năng theo dõi và phân tích được các phản hồi của người sử dụng.	1
		Phải có khả năng đăng các tin bài lên mục tin tức của Cổng dữ liệu mở.	1
III	**CHỨC NĂNG QUẢN LÝ TÀI KHOẢN**	Phải có khả năng quản lý phân quyền cho các quản trị viên khác và người xuất bản.	1
		Phải có khả năng vô hiệu hóa tài khoản người sử dụng.	1

3. Yêu cầu phi chức năng

Official Use Only

Cổng dữ liệu mở cấp bộ, cấp tỉnh cần đảm bảo các yêu cầu phi chức năng về:

1. Thông tin
2. Giao diện lập trình ứng dụng
3. Đặc tả dữ liệu
4. Định dạng dữ liệu
5. Giấy phép dữ liệu
6. Tính phi cá nhân hóa

Bảng A.5: Các yêu cầu phi chức năng của Cổng dữ liệu mở

STT	Nội dung	Mô tả chức năng	Mức độ ưu tiêu
I	**THÔNG TIN**		
I.1	Thông tin chung về cơ quan quản lý cổng	Cổng cung cấp các thông tin để người dùng nhận biết đây là cổng dữ liệu mở thuộc cơ quan nhà nước *Ví dụ: Cổng sử dụng đường dẫn gov.vn; tên cổng đi kèm tên tỉnh, hoặc cơ quan; cổng cung cấp thông tin cơ quan chủ quản, bao gồm các thông tin cơ bản như email, số điện thoại, địa chỉ cơ quan cung cấp*	1
I.2	Liên kết với Cổng dữ liệu mở Quốc gia	Cổng cung cấp liên kết với Cổng dữ liệu mở Quốc gia	1
I.3	Cung cấp tài liệu hướng dẫn, trợ giúp	Phải cung cấp các khái niệm và những thông tin liên quan đến dữ liệu mở được giải thích để người sử dụng hiểu.	1
I.4	Cung cấp tài liệu hướng dẫn, trợ giúp	Phải cung cấp các khái niệm và những thông tin liên quan đến dữ liệu mở được giải thích để người sử dụng hiểu.	3
		Có thể cung cấp các tập dữ liệu kèm theo các nguồn thông tin liên kết hoặc tài nguyên phục vụ việc hướng dẫn và hỗ trợ người dùng đăng nhập hoặc sử dụng API	2
I.5	Thông tin về Cổng dữ liệu mở	Phải có khả năng xem các chỉ số liên quan đến cổng cập nhật theo thời gian thực, bao gồm: (i) số lượng bộ dữ liệu hiện có, (ii) số cơ quan cung cấp, (iii) số lượt truy cập, (iv) số lượng người dùng	1
I.6	Thông tin về định dạng dữ liệu	Định dạng dữ liệu phải được ghi chú ngay phần bên dưới từng tập dữ liệu trong danh sách giúp người khai thác biết được các định dạng khi khai thác.	1
		Mỗi định dạng có thể kèm theo biểu tượng, màu nền khác nhau để thể hiện sự nổi bật.	2
I.7	Ngôn ngữ	Cổng cung cấp có thể nhiều phiên bản ngôn ngữ để thu hút nhiều người dùng hơn (bao gồm tiếng Anh và/hoặc nhiều ngôn ngữ khác).	3
II	**GIAO DIỆN LẬP TRÌNH ỨNG DỤNG**	Các lệnh gọi API không hợp lệ hoặc không thành công phải trả về mã trạng thái phản hồi HTTP phù hợp (404, 500, ...) và thông báo tình trạng.	1
		Nếu có các điều kiện đối với quyền truy cập vào tập dữ liệu, phương pháp truy xuất dữ liệu cần được chú giải bằng tài liệu.	2
		Các cơ chế về truy cập và tương tác với các bộ dữ liệu được chú giải bằng tài liệu.	2
III	**ĐẶC TẢ DỮ LIỆU**	Các tập dữ liệu cần phải được nhập thông tin về dữ liệu đặc tả, phải đầy đủ các trường bắt buộc theo Tiêu chuẩn Quốc gia TCVN 14171:2024 Đặc tả tập dữ liệu mở cho Cổng dữ liệu (DCAT- VN)	1

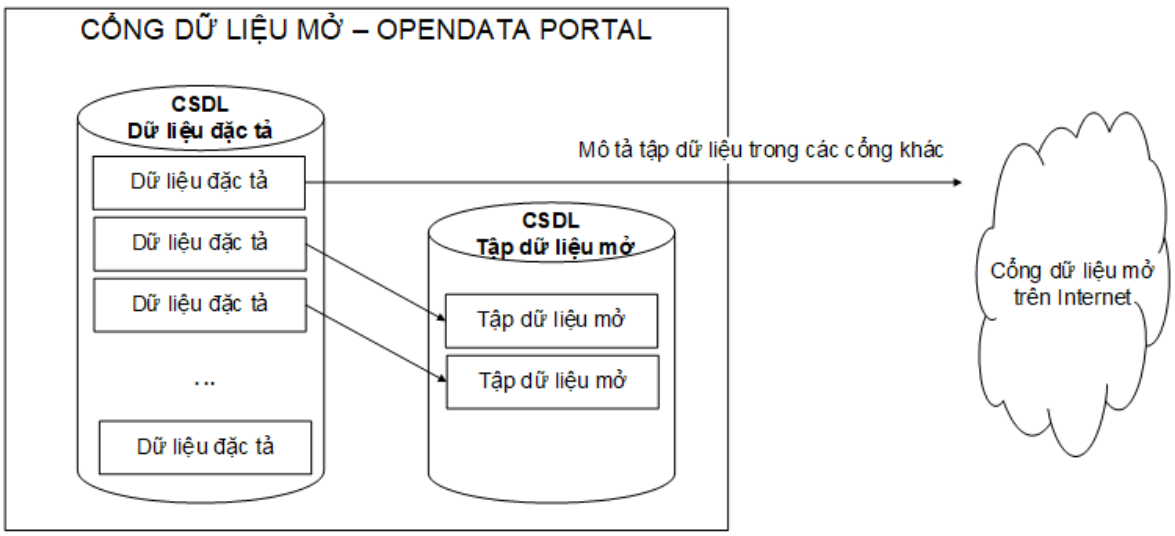
Official Use Only

STT	Nội dung	Mô tả chức năng	Mức độ ưu tiêu
IV	GIẤY PHÉP DỮ LIỆU	Các bộ dữ liệu được kèm theo giấy phép về quyền sử dụng. Trường hợp không phát sinh các quy định khác so với nghị định 47/2020/NĐ-CP thì ghi "Quy định về sử dụng dữ liệu mở của cơ quan nhà nước".	1
		Giấy phép có nêu các điều kiện trong việc phân bổ, tái sử dụng, phân phối lại và thương mại hóa liên quan đến dữ liệu.	3
		Bên xuất bản dữ liệu tuyên bố các chính sách liên quan đến việc phát hành dữ liệu.	3
V	TÍNH MINH BẠCH	Nếu một tập dữ liệu chứa ngày hết hạn thì điều này cần được cung cấp thông tin.	1
		Nếu có sự giới hạn về số lượng truy cập hoặc lượng dữ liệu được truy cập thông qua các dịch vụ web thì các giới hạn này cần được thông tin.	1
		Bất cứ các hạn chế đối với quyền truy cập vào dữ liệu đều được nêu rõ.	2
		Nếu tập dữ liệu có lỗi hoặc không đầy đủ cũng như không đảm bảo thì các vấn đề này cần được thông tin.	2
		Tần suất cập nhật được thể hiện đối với mỗi tập dữ liệu.	2
		Phía xuất bản tập dữ liệu cung cấp thông tin về nguồn gốc của dữ liệu.	3
		Nếu các tập dữ liệu có thông tin về mức độ tin cậy của số liệu hay chất lượng của dữ liệu thì các thông tin này cần được cung cấp.	3
VI	TÍNH PHI CÁ NHÂN HÓA	Cổng cung cấp các bộ dữ liệu đảm bảo bảo mật thông tin cá nhân. Đảm bảo không tiết lộ dữ liệu cá nhân và thông tin giúp xác định một con người cụ thể: Đảm bảo không tiết lộ dữ liệu cá nhân và thông tin giúp xác định một con người dưới dạng ký hiệu, chữ viết, chữ số, hình ảnh, âm thanh hoặc dạng tương tự trên môi trường điện tử gắn liền với một con người cụ thể hoặc giúp xác định một con người cụ thể (Điều 20 Nghị định số 47/2020/NĐ-CP ngày 09/4/2020 của Chính phủ về quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước và Điều 2 Nghị định 13/2023/NĐ-CP ngày 17/4/2023 của Chính phủ về Bảo vệ dữ liệu cá nhân).	1

4. Yêu cầu về quản lý dữ liệu đặc tả và tập dữ liệu

Trong cổng dữ liệu mở, dữ liệu đặc tả tập dữ liệu mở được lưu trữ riêng rẽ với các tập dữ liệu mở.

Dữ liệu đặc tả trong tập dữ liệu có thể mô tả tập dữ liệu lưu trữ trong cùng cổng dữ liệu mở hoặc mô tả tập dữ liệu tại một cổng dữ liệu mở khác trên mạng. Trong trường hợp này, thông tin trong dữ liệu đặc tả sẽ cung cấp đường dẫn để tải về hoặc đường dẫn truy cập để người dùng có thể tiếp cận với nguồn dữ liệu tương ứng.



Official Use Only

5. Yêu cầu về mô hình trao đổi, kết nối dữ liệu

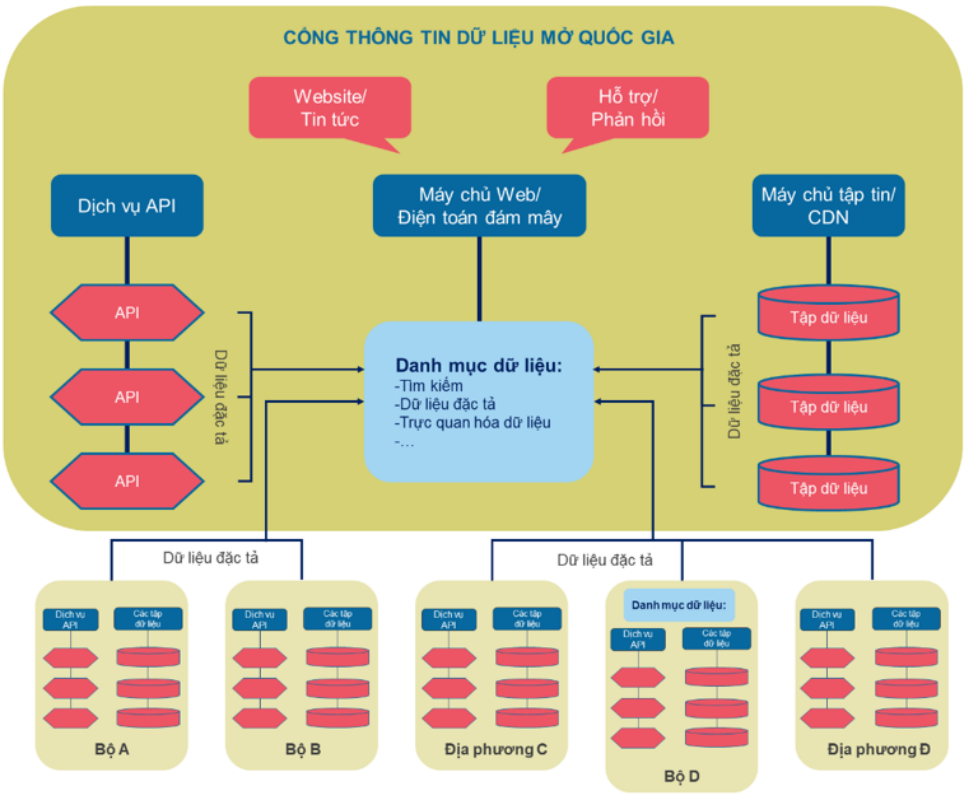
Cổng dữ liệu mở của các Bộ, ngành, địa phương phải được kết nối với Cổng dữ liệu mở quốc gia để tuân thủ theo Điều 21 Nghị định số 47/2020/NĐ-CP ngày 09/4/2020 của Chính phủ về quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước đã quy định: "Dữ liệu mở phải được công bố trên Cổng dữ liệu quốc gia."

Mối quan hệ giữa Cổng dữ liệu mở quốc gia và Cổng dữ liệu mở của các Bộ, ngành, địa phương được tiếp cận theo mô hình danh mục liên kết (hình vẽ dưới đây), theo đó:

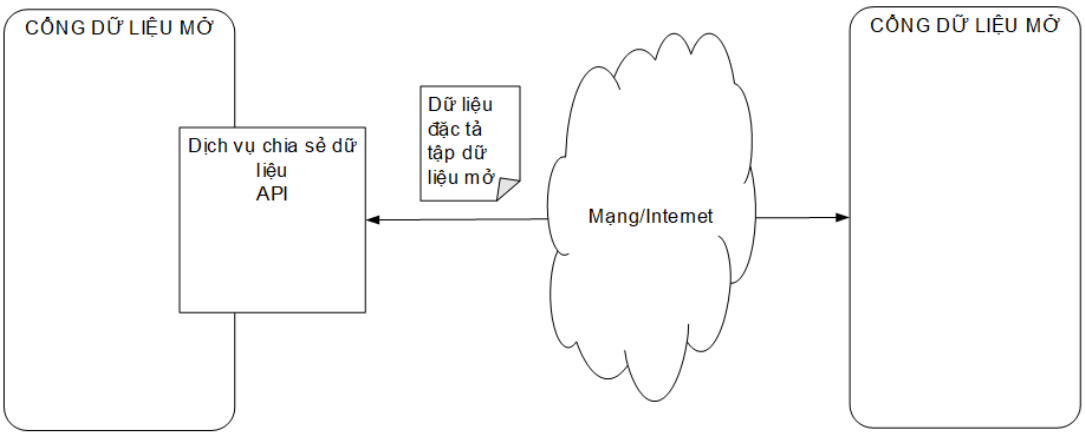
- Việc quản lý danh mục dữ liệu mở cho phép được phân cấp tới nhiều cơ quan khác nhau.

- Các cơ quan có thể vận hành danh mục dữ liệu (ví dụ như dữ liệu không gian địa lý hoặc thống kê giáo dục…) và quản lý các tệp dữ liệu, dịch vụ API của riêng mình.

- Dữ liệu đặc tả được cung cấp cho danh mục trung tâm nhằm cho phép tìm kiếm và truy cập dữ liệu liên cơ quan.



Mô hình kết nối giữa các cổng dữ liệu mở để trao đổi, chia sẻ dữ liệu đặc tả tập dữ liệu mở



Official Use Only

Các cổng dữ liệu mở sẽ chia sẻ các thông điệp dữ liệu đặc tả tập dữ liệu mở.

Các dịch vụ chia sẻ dữ liệu (API) sẽ là đầu mối để cung cấp, tiếp nhận dữ liệu đặc tả. Các dịch vụ chia sẻ dữ liệu bao gồm: truy vấn, tìm kiếm các tập dữ liệu.

Official Use Only

TÀI LIỆU THAM KHẢO

Bộ TT&TT và Cục chuyển đổi số Quốc gia, Thuyết minh Dự thảo tiêu chuẩn Quốc gia: Đặc tả DLM cho Cổng dữ liệu DCAT-VN. 2023

Bộ Nội vụ và Viện Khoa học tổ chức nhà nước. Thực trạng và giải pháp phát triển dữ liệu mở tại Việt Nam. 2022: https://isos.gov.vn/cds/thuc-trang-va-giai-phap-phat-trien-du-lieu-mo-tai-viet-nam-43241.html

Hiến pháp năm 2013

Luật Công nghệ thông tin năm 2006

Luật Bảo vệ bí mật nhà nước năm 2018

Luật Tiếp cận thông tin năm 2016

Luật Giao dịch điện tử năm 2023

Luật Dữ liệu năm 2024

Nghị định 47/2020/NĐ-CP ngày 9 tháng 4 năm 2020 của Chính phủ về quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước

Nghị định số 42/2022/NĐ-CP ngày 24 tháng 6 năm 2022 của Chính phủ quy định về việc cung cấp thông tin và dịch vụ công trực tuyến của cơ quan Nhà nước trên môi trường mạng

Nghị định 13/2023/NĐ-CP ngày 17 tháng 4 năm 2023 của Chính phủ về bảo vệ dữ liệu cá nhân

Quyết định 749/QĐ-TTg ngày 03 tháng 6 năm 2020 của Thủ tướng Chính phủ phê duyệt "Chương trình chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030"

Quyết định 942/QĐ-TTg ngày 15 tháng 6 năm 2021 của Thủ tướng Chính phủ phê duyệt chiến lược phát triển Chính phủ điện tử hướng tới Chính phủ số giai đoạn 2021-2025, định hướng đến năm 2030

Quyết định 411/QĐ-TTg ngày 31 tháng 3 năm 2022 của Thủ tướng Chính phủ phê duyệt chiến lược Quốc gia phát triển kinh tế số và xã hội số đến năm 2025, định hướng đến năm 2030

Quyết định 146/QĐ-TTg ngày 28 tháng 01 năm 2022 của Thủ tướng Chính phủ phê duyệt Đề án "Nâng cao nhận thức, phổ cập kĩ năng và phát triển nguồn nhân lực chuyển đổi số Quốc gia đến năm 2025, định hướng đến năm 2030

Ngân hàng Thế giới, Khóa học về dữ liệu dành cho các nhà cung cấp dữ liệu: https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/essentials

Data.europa.edu. Chọn đúng định dạng cho DLM. https://data.europa.eu/elearning/en/module9/#/id/co-01

Data.gouv.fr, Cấu trúc một tập:

https://guides.data.gouv.fr/guides-open-data/guide-qualite/preparer-un-jeu-de-donnees-de-qualite/structurer-un-jeu-de-donnees

Đạo luật dữ liệu chính phủ MỞ:

https://www.congress.gov/bill/115th-congress/house-bill/4174

Vì sao dữ liệu Chính phủ mở có vai trò quan trọng?: https://www.infrastructure.gov.au/sites/default/files/documents/BCR%20Open%20government%20data%20report%20-%20final%20-%205%20Feb2016.pdf

Bộ trưởng Văn phòng Nội các và Tổng cục trưởng Tài chính. Sách trắng về DLM. 2012

Đạo luật về thúc đẩy việc cung cấp và sử dụng dữ liệu công cộng: https://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=47133&type=part&key=4

Ngân hàng Thế giới. Báo cáo Chính phủ Nước CHXNCN Việt Nam: "Đánh giá Mức độ sẵn sàng về Chính phủ Số và DLM". 2019.

Quy định bảo vệ dữ liệu chung 2016/679 của châu Âu (GDPR)

Công văn số 5752/BTTT-CĐSQG ngày 26/11/2022 của Bộ TT&TT gửi các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, Ủy ban nhân dân các tỉnh, thành phố trực thuộc Trung ương về việc đôn đốc triển khai Nghị định 47/2020/NĐ-CP, thúc đẩy kết nối, chia sẻ dữ liệu, cung cấp DLM.

Nghị định 64/2007/NĐ-CP ngày 10/4/2007 của Chính phủ về Ứng dụng CNTT trong hoạt động của CQNN, bao gồm việc xây dựng, khai thác và duy trì CSDL

Thông tư số 13/2017/TT-BTTTT ngày 23/6/2017 của Bộ trưởng Bộ TT&TT quy định về việc cung cấp thông tin và bảo đảm khả năng truy cập thông tin trên trang thông tin điện tử hoặc cổng thông tin điện tử của cơ quan nhà nước

Thông tư 39/2017/TT-BTTTT ngày 15/12/2017 của Bộ trưởng Bộ TT&TT quy định về Danh mục tiêu chuẩn kĩ thuật về ứng dụng CNTT trong cơ quan nhà nước.

Công văn số 3788/BTTTT-THH ngày 26/12/2014 của Bộ TT&TT hướng dẫn triển khai ứng dụng CNTT trong hoạt động của cơ quan nhà nước quy định về việc liên thông, trao đổi dữ liệu có cấu trúc bằng ngôn ngữ XML giữa các hệ thống thông tin trong cơ quan nhà nước

Công văn số 2051/BTTTT-THH ngày 04/06/2020 của Bộ TT&TT hướng dẫn triển khai một số nhiệm vụ thực thi Nghị định số 47/2020/NĐ-CP

Công văn số 677/BTTTT-THH ngày 3/3/2022 của Bộ TT&TT hướng dẫn kết nối và chia sẻ dữ liệu thông qua Nền tảng tích hợp, chia sẻ dữ liệu quốc gia

Công văn số 1552/BTTTT-THH ngày 26/4/2022 hướng dẫn kĩ thuật triển khai Đề án 06 (phiên bản 1.0)

Văn bản số 761/VPCP-KSTT ngày 29/01/2022 của Văn phòng Chính phủ về việc hướng dẫn kết nối, tích hợp, chia sẻ với CSDLQG về dân cư

Công văn 1016/BTTTT-CĐSQG ngày 22/03/2024 của Bộ TT&TT hướng dẫn xây dựng, áp dụng quy chuẩn, tiêu chuẩn, quy định kĩ thuật kết nối, chia sẻ dữ liệu.

Quyết định 962 QĐ-UBND Quyết định ban hành danh mục dữ liệu mở thành phố Đà Nẵng: https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Quyet-dinh-962-QD-UBND-2023-Danh-muc-du-lieu-mo-thanh-pho-Da-Nang-566456.aspx

Bộ Nội vụ và Viện Khoa học tổ chức nhà nước. Thực trạng và giải pháp phát triển dữ liệu mở tại Việt Nam, 2022:

https://isos.gov.vn/cds/thuc-trang-va-giai-phap-phat-trien-du-lieu-mo-tai-viet-nam-43241.html

Bộ Đổi mới Kinh doanh & Kỹ năng của Vương quốc Anh, 2014:

https://assets.publishing.service.gov.uk/media/5a7ec24940f0b62305b82ff6/bis-14-946-open-data-strategy-2014-2016.PDF.

Vassil, 2016, https://pubdocs.worldbank.org/en/165711456838073531/WDR16-BP-Estonian-eGov-ecosystem-Vassil.PDF; Quan hệ đối tác Chính phủ mở, 2023, https://www.opengovpartnership.org/stories/estonias-journey-of-open-government-reforms/

CSA Singapore. Ra mắt chiến dịch nâng cao nhận thức về an ninh mạng quốc gia lần thứ hai:

https://www.csa.gov.sg/news/press-releases/csa-launches-second-national-cybersecurity-awareness-campaign

Cơ quan Phòng thủ mạng của Hoa Kỳ. Tháng nâng cao nhận thức về an ninh mạng quốc gia : https://www.dhs.gov/national-cyber-security-awareness-month

Australian Aid

Số 8 Đào Tấn, Ba Đình, Hà Nội, Việt Nam
Điện thoại: +84 24 37740100
Fax: +84 24 37740111
Website: www.dfat.gov.au

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

Tầng 8, 63 Lý Thái Tổ, Hoàn Kiếm, Hà Nội, Việt Nam
Điện thoại: +84 24 39346600
Fax: +84 24 39346597
Website: www.worldbank.org/en/country/vietnam

