

**KHUNG KIẾN TRÚC DỮ LIỆU QUỐC GIA,
KHUNG QUẢN TRỊ, QUẢN LÝ DỮ LIỆU QUỐC GIA,
TỪ ĐIỂN DỮ LIỆU DÙNG CHUNG (PHIÊN BẢN 1.0)**
(Kèm theo Quyết định số /QĐ-TTg
ngày tháng năm 2025 của Thủ tướng Chính phủ)

Chương I
KHÁI QUÁT CHUNG

I. MỤC ĐÍCH

Việc xây dựng Khung kiến trúc dữ liệu quốc gia, Khung quản trị, quản lý dữ liệu quốc gia và Từ điển dữ liệu dùng chung quốc gia nhằm thiết lập ba trụ cột cốt lõi trong phát triển và quản trị dữ liệu tạo nền tảng vững chắc cho quá trình xây dựng và vận hành Quốc gia số. Ba thành phần này khi được thiết kế ban hành và triển khai đồng bộ sẽ hình thành nền tảng dữ liệu quốc gia thống nhất, dùng chung, an toàn và hiệu quả, bảo đảm việc kết nối, chia sẻ và khai thác dữ liệu phục vụ công tác chỉ đạo, điều hành, phát triển kinh tế - xã hội và thúc đẩy chuyển đổi số quốc gia.

Khung kiến trúc dữ liệu quốc gia được thiết kế dựa trên Từ điển dữ liệu dùng chung và Khung quản trị, quản lý dữ liệu quốc gia để bảo đảm mô hình thống nhất và tuân thủ trong triển khai. Khung quản trị, quản lý dữ liệu quốc gia là công cụ vận hành, giám sát việc thực thi Khung kiến trúc dữ liệu quốc gia và duy trì Từ điển dữ liệu dùng chung. Từ điển dữ liệu dùng chung đóng vai trò nền tảng, hỗ trợ việc thiết kế Khung kiến trúc dữ liệu quốc gia và triển khai Khung quản trị, quản lý dữ liệu thống nhất.

1. Định hướng thống nhất việc xây dựng và triển khai Khung kiến trúc dữ liệu quốc gia

a) Xây dựng Khung kiến trúc dữ liệu cấp bộ, cấp tỉnh, các cơ quan thuộc hệ thống chính trị từ trung ương đến địa phương, bảo đảm thống nhất về nguyên tắc, cấu trúc, mô hình dữ liệu và phương thức kết nối;

b) Xác định rõ các lớp dữ liệu cơ bản như dữ liệu gốc, dữ liệu đặc tả (siêu dữ liệu), dữ liệu chủ, dữ liệu danh mục, dữ liệu giao dịch, dữ liệu tổng hợp, dữ liệu phân tích và dữ liệu mở, làm cơ sở tổ chức quản lý và khai thác dữ liệu hiệu quả;

c) Thiết lập Nền tảng chia sẻ, điều phối dữ liệu dùng chung giữa các bộ, ngành, địa phương và Trung tâm dữ liệu quốc gia, bảo đảm khả năng tích hợp, kết nối đồng bộ giữa các hệ thống thông tin, các cơ sở dữ liệu quốc gia và chuyên ngành;

d) Chuẩn hóa và mô hình hóa dữ liệu theo Từ điển dữ liệu dùng chung, bảo đảm khả năng đa tương tác, tái sử dụng và tích hợp dữ liệu;

đ) Tăng cường chia sẻ dữ liệu bắt buộc giữa các cơ quan trong hệ thống chính trị, hướng tới hình thành hệ sinh thái dữ liệu quốc gia, kết nối liên ngành và liên vùng, phối hợp hiệu quả trong giám sát và chỉ đạo, điều hành theo thời gian thực góp phần phát triển thị trường dữ liệu, thúc đẩy kinh tế số, đổi mới sáng tạo trong các lĩnh vực công nghệ số, đóng góp tích cực vào tăng trưởng kinh tế và xây dựng quốc gia số toàn diện; nâng cao hiệu quả hoạt động của cán bộ công chức, viên chức, giúp giảm thiểu lỗi thủ công thông qua tự động hóa và tái sử dụng dữ liệu; tạo cơ hội cho các dịch vụ, làm giàu dữ liệu và tạo ra các dữ liệu gia tăng từ các nguồn dữ liệu khác nhau;

e) Nâng cao chất lượng dịch vụ công, giúp tổ chức, cá nhân kê khai thông tin một lần, thúc đẩy cải cách thủ tục hành chính, giảm giấy tờ, tăng tính minh bạch và rút ngắn thời gian xử lý hồ sơ, bảo vệ quyền, lợi ích hợp pháp của người dân, doanh nghiệp;

g) Bảo đảm yêu cầu về an toàn, bảo mật, an ninh mạng, bảo vệ dữ liệu cá nhân và bảo vệ bí mật nhà nước bảo đảm an ninh quốc gia, giám sát truy cập theo quy định của pháp luật trong toàn bộ quá trình thiết kế, triển khai, khai thác và chia sẻ dữ liệu;

h) Bảo đảm hạ tầng dữ liệu và nền tảng dữ liệu quốc gia gồm các trung tâm dữ liệu quốc gia, mạng truyền số liệu quốc gia, nền tảng chia sẻ, điều phối dữ liệu quốc gia đáp ứng nhu cầu lưu trữ và xử lý dữ liệu lớn của các cơ sở dữ liệu quan trọng; bảo đảm triển khai các công nghệ số tiên tiến như Internet vạn vật, dữ liệu lớn, học máy, trí tuệ nhân tạo, chuỗi khối (blockchain), điện toán đám mây;

i) Tăng cường kiểm tra, giám sát, đánh giá hiệu quả đầu tư, vận hành, khai thác hạ tầng dữ liệu, bảo đảm triển khai đồng bộ, tránh trùng lặp, tối ưu hóa hạ tầng, nhân lực và chi phí;

k) Phân công vai trò, trách nhiệm của các chủ thể trong tổ chức dữ liệu: Chủ sở hữu dữ liệu, chủ quản dữ liệu và cơ quan thực hiện điều phối dữ liệu. Bảo đảm tổ chức, vận hành, sử dụng dữ liệu quốc gia thống nhất, dùng chung;

l) Bảo đảm khả năng tương thích quốc tế, thúc đẩy trao đổi dữ liệu xuyên biên giới và tích hợp cơ chế liên kết quốc tế, cho phép Việt Nam kết nối với các nền tảng dữ liệu khu vực và toàn cầu (ASEAN, APEC, EU...), thúc đẩy

thương mại điện tử, nghiên cứu khoa học, cải cách thể chế và nâng cao năng lực tương tác của doanh nghiệp và người dân trong môi trường số toàn cầu.

2. Xây dựng Khung quản trị, quản lý dữ liệu quốc gia bảo đảm dữ liệu “đúng, đủ, sạch, sống, thống nhất, dùng chung” và khai thác tài nguyên dữ liệu

a) Xây dựng Khung quản trị, quản lý dữ liệu cấp bộ, cấp tỉnh, các cơ quan thuộc hệ thống chính trị từ trung ương đến địa phương, tuân thủ các nguyên tắc, quy trình, tiêu chuẩn và công cụ quản trị, quản lý dữ liệu;

b) Xây dựng mô hình tổ chức quản trị, quản lý dữ liệu để bảo đảm rõ vai trò, trách nhiệm trong suốt vòng đời dữ liệu;

c) Thiết lập các tiêu chí đánh giá, bảo đảm chất lượng dữ liệu như: tính đủ, chính xác, hợp lệ, kịp thời, khả năng chia sẻ, liên kết và tái sử dụng; triển khai các quy trình kiểm tra, làm sạch và chuẩn hóa dữ liệu;

d) Thiết lập cơ chế kiểm toán dữ liệu, giám sát an toàn, bảo mật dữ liệu, hạn chế tối đa thất thoát, sai lệch và vi phạm trong quản lý và khai thác dữ liệu;

đ) Yêu cầu về công cụ hỗ trợ quản trị, quản lý dữ liệu tự động như: phân loại dữ liệu, gán nhãn, kiểm soát truy cập và đo lường hiệu quả sử dụng;

e) Thúc đẩy khai thác tài nguyên dữ liệu có kiểm soát phục vụ phát triển kinh tế - xã hội, đổi mới sáng tạo và chuyển đổi số.

3. Xây dựng Từ điển dữ liệu dùng chung làm công cụ tiêu chuẩn hóa dữ liệu, bảo đảm đồng bộ, tương thích và liên kết giữa các cơ sở dữ liệu, hệ thống thông tin trên phạm vi toàn quốc

a) Xây dựng bộ từ vựng cốt lõi, theo lĩnh vực và chuẩn hóa các trường dữ liệu cơ bản, bảo đảm thống nhất về tên gọi, định nghĩa về nghiệp vụ, kỹ thuật và quy tắc áp dụng; bảo đảm mỗi từ vựng có định danh duy nhất, rõ ràng về ngữ nghĩa, phạm vi áp dụng làm cơ sở liên kết dữ liệu giữa các hệ thống;

b) Thiết lập mô hình siêu dữ liệu (metadata) dạng phân cấp, mô tả cấu trúc, mối quan hệ giữa các phần tử dữ liệu, hỗ trợ kết nối, chia sẻ, khai thác dữ liệu hiệu quả, linh hoạt;

c) Cung cấp công cụ trực tuyến để cập nhật, tra cứu, quản lý, đồng bộ Từ điển dữ liệu giữa các cơ quan, tổ chức;

d) Xây dựng quy chế phối hợp liên ngành trong rà soát, bổ sung, chuẩn hoá Từ điển dữ liệu. Quản lý các phiên bản Từ điển dữ liệu bảo đảm cập nhật, kịp thời theo quy định của pháp luật và yêu cầu nghiệp vụ;

đ) Bảo đảm khả năng tương thích với các chuẩn dữ liệu quốc tế phục vụ tích hợp, chia sẻ, khai thác và trao đổi dữ liệu xuyên biên giới.

II. PHẠM VI ÁP DỤNG

Khung kiến trúc dữ liệu quốc gia, Khung quản trị, quản lý dữ liệu quốc gia và Từ điển dữ liệu dùng chung áp dụng thống nhất tại các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan trung ương của các tổ chức chính trị, Viện kiểm sát nhân dân tối cao, Toà án nhân dân tối cao, các cơ quan của Quốc hội, các cơ quan Trung ương thuộc Mặt trận Tổ quốc Việt Nam, Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương, cơ quan, tổ chức, cá nhân trong hoạt động trực tiếp hoặc có liên quan đến các nội dung của Khung kiến trúc dữ liệu quốc gia, Khung quản trị, quản lý dữ liệu quốc gia và Từ điển dữ liệu dùng chung (sau đây gọi là bộ, cơ quan trung ương, địa phương và tổ chức, cá nhân có liên quan).

III. NỘI DUNG CƠ BẢN

1. Khung kiến trúc dữ liệu quốc gia

Khung kiến trúc dữ liệu quốc gia là hệ thống định hướng tổng thể về cấu trúc, phân tầng, tiêu chuẩn kỹ thuật và mô hình chia sẻ dữ liệu giữa các cơ sở dữ liệu quốc gia, chuyên ngành, địa phương và các hệ thống thông tin của các cơ quan trong hệ thống chính trị.

Khung kiến trúc dữ liệu quốc gia mô tả tổng thể kiến trúc các lớp dữ liệu cơ bản như dữ liệu gốc, dữ liệu đặc tả, dữ liệu chủ, dữ liệu danh mục, dữ liệu giao dịch, dữ liệu tổng hợp, dữ liệu phân tích, dữ liệu mở; mô hình hóa các khối dữ liệu theo chức năng, nghiệp vụ: dữ liệu phục vụ người dân, doanh nghiệp, dữ liệu phục vụ hoạt động quản lý nghiệp vụ chuyên ngành, dữ liệu phục vụ công tác chỉ đạo điều hành, xây dựng chính sách; dữ liệu phục vụ quản trị nguồn lực (nội bộ), dữ liệu phục vụ an ninh, quốc phòng, trật tự an toàn xã hội và dữ liệu về thể chế, chính sách; xác định các nền tảng chia sẻ, điều phối dữ liệu, kết nối dữ liệu giữa các hệ thống thông tin và cơ sở dữ liệu cấp bộ, ngành và địa phương với Trung tâm dữ liệu quốc gia và luồng điều phối dữ liệu dùng chung quốc gia.

2. Khung quản trị, quản lý dữ liệu quốc gia

Khung quản trị, quản lý dữ liệu quốc gia là hệ thống các nguyên tắc, cơ chế, quy trình và cấu trúc tổ chức để bảo đảm việc thu thập, lưu trữ, kết nối, chia sẻ, khai thác, mở và bảo vệ dữ liệu quốc gia được thực hiện thống nhất và hiệu quả.

Khung quản trị, quản lý dữ liệu quốc gia xác định rõ vai trò, trách nhiệm của các chủ thể trong quản lý dữ liệu: chủ sở hữu dữ liệu, cơ quan chủ quản dữ liệu, cơ quan khai thác dữ liệu; xác định mô hình giám sát, kiểm toán, đánh giá hiệu quả, an toàn và chất lượng dữ liệu; xây dựng cơ chế quản trị vòng đời dữ liệu, bảo đảm dữ liệu được cập nhật, kiểm định, phân quyền và khai thác hiệu quả.

3. Từ điển dữ liệu dùng chung

Từ điển dữ liệu dùng chung là tập hợp các định nghĩa, cấu trúc dữ liệu, mô hình ngữ nghĩa và quy tắc chuẩn hóa dữ liệu.

Từ điển dữ liệu dùng chung để chuẩn hóa tên gọi, định nghĩa nghiệp vụ, định nghĩa kỹ thuật và quy tắc quản lý khác đối với từng trường dữ liệu để sử dụng trong các hệ thống thông tin, cơ sở dữ liệu của bộ, ngành, địa phương; định danh duy nhất, rõ ràng về ngữ nghĩa, phạm vi áp dụng cho mỗi thuật ngữ; bảo đảm tính mở rộng, khả năng tích hợp và liên thông với các chuẩn dữ liệu quốc tế phổ biến.

IV. NGUYÊN TẮC CHUNG

1. Tuân thủ pháp luật

Bảo đảm tuân thủ pháp luật về dữ liệu, quy định về phân loại dữ liệu, quyền và trách nhiệm của cơ quan nhà nước trong việc quản lý, chia sẻ, khai thác dữ liệu, bảo vệ dữ liệu cá nhân và dữ liệu thuộc bí mật nhà nước; bảo đảm tuân thủ pháp luật về giao dịch điện tử, quy định việc sử dụng dữ liệu điện tử, định danh điện tử, chữ ký số và nền tảng số làm căn cứ pháp lý cho hoạt động kết nối, chia sẻ, sử dụng dữ liệu giữa các cơ quan, tổ chức; các văn bản quy phạm pháp luật liên quan đến an toàn thông tin, chuẩn định danh, chia sẻ dữ liệu, bảo vệ quyền riêng tư; bảo đảm phù hợp thông lệ, điều ước quốc tế.

2. Phù hợp với Khung kiến trúc tổng thể quốc gia số, Khung kiến trúc số cấp bộ, cấp tỉnh

a) Bảo đảm sự nhất quán về nguyên tắc thiết kế, mô hình dữ liệu, nền tảng kỹ thuật và quy trình triển khai giữa Khung kiến trúc dữ liệu quốc gia, Khung kiến trúc tổng thể quốc gia số;

b) Tuân thủ cấu trúc và mối quan hệ tương thích giữa các tầng kiến trúc (nghiệp vụ, ứng dụng, dữ liệu, công nghệ, bảo mật) bảo đảm khả năng tích hợp, mở rộng và tái sử dụng giữa các cấp trung ương và địa phương;

c) Liên thông và kế thừa từ Khung kiến trúc số cấp bộ, cấp tỉnh làm cơ sở để chuẩn hóa việc kết nối, chia sẻ dữ liệu, cập nhật phù hợp với khung kiến trúc số chuyên ngành, bảo đảm phù hợp với đặc thù nghiệp vụ, phạm vi quản lý và hệ thống thông tin theo từng lĩnh vực (y tế, giáo dục, tài chính, tư pháp, v.v.) giữa các bộ, ngành, địa phương và các cơ quan khác thuộc hệ thống chính trị;

d) Thúc đẩy việc hình thành một hệ sinh thái dữ liệu quốc gia đồng bộ, thống nhất, làm nền tảng cho hoạt động chuyển đổi số, ra quyết định dựa trên dữ liệu để phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số trên phạm vi toàn quốc.

3. Ứng dụng nền tảng mở, công nghệ hiện đại, dữ liệu là tài nguyên số chủ đạo để hình thành kinh tế dữ liệu

a) Áp dụng các nền tảng mở, công nghệ số tiên tiến (như điện toán đám mây, trí tuệ nhân tạo, Blockchain, dữ liệu lớn...) nâng cao khả năng tự chủ, linh hoạt và tối ưu hóa chi phí trong khai thác và quản trị dữ liệu;

b) Thúc đẩy phát triển hệ sinh thái số dựa trên dữ liệu, trong đó dữ liệu được coi là tài nguyên chiến lược, đóng vai trò tương tự như đất đai, tài chính hay nhân lực trong phát triển kinh tế;

c) Ưu tiên triển khai các nền tảng số dùng chung cấp quốc gia, vùng, ngành, bảo đảm tính mở, tính liên thông và khả năng mở rộng để hỗ trợ cung cấp dịch vụ số thống nhất, xuyên suốt giữa các cấp hành chính;

d) Khuyến khích doanh nghiệp, tổ chức tham gia vào chuỗi giá trị dữ liệu, bao gồm: thu thập, làm sạch, chuẩn hóa, phân tích và cung cấp dịch vụ dựa trên dữ liệu, tạo điều kiện hình thành thị trường dữ liệu minh bạch, cạnh tranh lành mạnh;

đ) Tạo môi trường pháp lý, kỹ thuật và tài chính thuận lợi để dữ liệu trở thành tài sản số có giá trị kinh tế cụ thể, có khả năng đo lường, định giá và đưa vào giao dịch, thúc đẩy nền kinh tế dữ liệu phát triển bền vững.

4. Bảo đảm an ninh, an toàn dữ liệu, bảo vệ dữ liệu cá nhân và bí mật nhà nước trong quá trình quản trị và khai thác dữ liệu

a) Tuân thủ nghiêm ngặt các quy định pháp luật hiện hành về an toàn thông tin mạng, bảo vệ dữ liệu cá nhân và bảo vệ bí mật nhà nước, Luật An ninh mạng, Luật Bảo vệ bí mật nhà nước, Luật Dữ liệu và các văn bản quy phạm pháp luật khác;

b) Bảo đảm phân loại dữ liệu theo mức độ nhạy cảm và mức độ rủi ro, từ đó áp dụng các biện pháp bảo vệ phù hợp về kỹ thuật và tổ chức;

c) Thống nhất các giải pháp bảo mật theo thiết kế (security by design) trong toàn bộ vòng đời của dữ liệu và hệ thống thông tin, từ giai đoạn xây dựng kiến trúc, phát triển ứng dụng đến vận hành và khai thác;

d) Bảo đảm quyền riêng tư cá nhân và thông tin định danh, thông qua các cơ chế kiểm soát truy cập, xin phép sử dụng dữ liệu, ẩn danh hóa hoặc giả danh dữ liệu trước khi chia sẻ hoặc khai thác;

đ) Bảo đảm hiệu quả giám sát, cảnh báo và ứng phó sự cố an toàn thông tin, bảo đảm dữ liệu quốc gia được kiểm soát toàn diện, tránh rò rỉ, mất mát hoặc bị sử dụng sai mục đích;

e) Nâng cao hiệu quả quy trình kiểm toán, đánh giá an toàn bảo mật dữ liệu định kỳ, gắn với trách nhiệm của từng chủ thể dữ liệu và các đơn vị vận hành kỹ thuật.

5. Chia sẻ dữ liệu theo hướng bắt buộc, minh bạch, kiểm soát được trách nhiệm và hành vi sử dụng dữ liệu

a) Bảo đảm dữ liệu được chia sẻ, thống nhất, dùng chung giữa các cơ quan trong hệ thống chính trị theo cơ chế bắt buộc, không phụ thuộc vào thỏa thuận riêng, phục vụ công tác quản lý nhà nước, cung cấp dịch vụ công và chỉ đạo điều hành;

b) Thống nhất danh mục dữ liệu bắt buộc chia sẻ và quy định rõ trách nhiệm cung cấp, khai thác, cập nhật, bảo vệ và giám sát việc sử dụng dữ liệu đối với từng chủ thể;

c) Bảo đảm hạ tầng kỹ thuật và pháp lý phục vụ chia sẻ dữ liệu, gồm nền tảng chia sẻ, điều phối dữ liệu quốc gia; hệ thống kết nối, giám sát và kiểm soát truy cập dữ liệu theo thời gian thực;

d) Tăng cường giám sát hành vi sử dụng dữ liệu thông qua hệ thống định danh, truy vết và nhật ký hoạt động, phát hiện kịp thời các hành vi vi phạm, sử dụng sai mục đích hoặc gây rủi ro an toàn dữ liệu;

đ) Bảo đảm minh bạch về quyền truy cập, mục đích khai thác, thời gian sử dụng và phạm vi chia sẻ dữ liệu, đồng thời cho phép cơ quan cung cấp dữ liệu được theo dõi, giám sát và quản lý việc khai thác của bên thứ ba.

Chương II

KHUNG KIẾN TRÚC DỮ LIỆU QUỐC GIA

I. MỤC TIÊU, PHẠM VI

1. Mục tiêu

Thống nhất việc xây dựng kiến trúc dữ liệu áp dụng tại các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan trung ương của các tổ chức chính trị, Viện kiểm sát nhân dân tối cao, Tòa án nhân dân tối cao, các cơ quan của Quốc hội, các cơ quan Trung ương thuộc Mặt trận Tổ quốc Việt Nam, Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương.

2. Phạm vi

Khung kiến trúc dữ liệu quốc gia được áp dụng tại các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan trung ương của các tổ chức chính trị, Viện kiểm sát nhân dân tối cao, Tòa án nhân dân tối cao, các cơ quan của Quốc hội, các cơ quan Trung ương thuộc Mặt trận Tổ quốc Việt Nam, Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương, cơ quan, tổ chức, cá nhân trong hoạt động trực tiếp hoặc có liên quan đến các nội dung của Khung kiến trúc dữ liệu quốc gia.

II. NGUYÊN TẮC, YÊU CẦU

1. Nguyên tắc Khung kiến trúc dữ liệu quốc gia

a) Về thu thập, tạo lập, cập nhật dữ liệu

- Thu thập dữ liệu một lần: dữ liệu chỉ được thu thập một lần duy nhất để hạn chế trùng lặp, giảm chi phí xử lý và tăng độ tin cậy;

- Một nguồn dữ liệu gốc duy nhất: dữ liệu được một chủ thể tạo lập, quản lý và chịu trách nhiệm, tạo sự rõ ràng trong kiểm soát, bảo mật và chia sẻ giữa các hệ thống;

- Cập nhật và đồng bộ liên tục: dữ liệu phải được duy trì chính xác, kịp thời, đồng bộ giữa các hệ thống;

- Cơ sở dữ liệu tổng hợp quốc gia có chức năng quản lý thống nhất dữ liệu chủ quốc gia: các quy trình nghiệp vụ, thủ tục hành chính có thực hiện cập nhật, thay đổi dữ liệu chủ thuộc danh mục cơ sở dữ liệu quốc gia phải hoàn thành bước cập nhật vào Cơ sở dữ liệu tổng hợp quốc gia để thông tin chứa bởi dữ liệu chủ đó có giá trị pháp lý.

b) Về lưu trữ và quản lý dữ liệu

- Lưu trữ dữ liệu tối thiểu: các cơ quan, tổ chức có trách nhiệm lưu trữ dữ liệu ở mức tối thiểu cần thiết, phù hợp với chức năng, nhiệm vụ được giao và mục tiêu sử dụng, đồng thời bảo đảm tiết kiệm tài nguyên hệ thống;

- Bảo đảm chất lượng dữ liệu: dữ liệu luôn đúng, đầy đủ, sạch và cập nhật kịp thời, đáp ứng yêu cầu quản lý, khai thác và phục vụ phân tích, ra quyết định.

c) Về chia sẻ dữ liệu

- Minh bạch về dữ liệu chia sẻ: công khai thông tin về nguồn gốc, nội dung và phạm vi sử dụng của dữ liệu chia sẻ, nâng cao mức độ tin cậy và tăng cường giám sát;

- Chia sẻ dữ liệu chủ động: việc chia sẻ cần được triển khai bằng phương thức chủ động, sẵn sàng kết nối và cung cấp dữ liệu, phục vụ dịch vụ công, nhiệm vụ quản lý nhà nước, điều hành và ra quyết định;

- Tối thiểu hóa trung gian kết nối dữ liệu: tối thiểu hóa các khâu trung gian trong quá trình kết nối, bảo đảm dữ liệu được truyền tải nhanh chóng, chính xác và tiết kiệm chi phí hạ tầng.

d) Về khai thác, sử dụng và kết thúc sử dụng dữ liệu

- Sử dụng dữ liệu đúng mục đích: chỉ khai thác dữ liệu phục vụ quản lý nhà nước, cải cách hành chính, phát triển kinh tế - xã hội;

- Minh bạch và trách nhiệm: cơ quan khai thác phải công khai phạm vi sử dụng và chịu trách nhiệm về dữ liệu sử dụng;

- Khi kết thúc sử dụng, dữ liệu cần được lưu trữ lịch sử hoặc hủy dữ liệu theo quy định của pháp luật.

2. Yêu cầu

Khung kiến trúc dữ liệu quốc gia cần bảo đảm một số yêu cầu dưới đây:

a) Khung kiến trúc dữ liệu quốc gia là cơ sở để các đơn vị tham chiếu, triển khai các cơ sở dữ liệu và các hệ thống thông tin có quản lý dữ liệu, bảo đảm tuân thủ luật pháp và chính sách của Đảng và Nhà nước: Mô hình tổ chức thu thập tạo lập dữ liệu; lưu trữ dữ liệu; quản trị, quản lý dữ liệu kết nối, chia sẻ, điều phối dữ liệu; khai thác, sử dụng; phân tích, tổng hợp dữ liệu ... đáp ứng các quy định của pháp luật về dữ liệu, giao dịch điện tử.

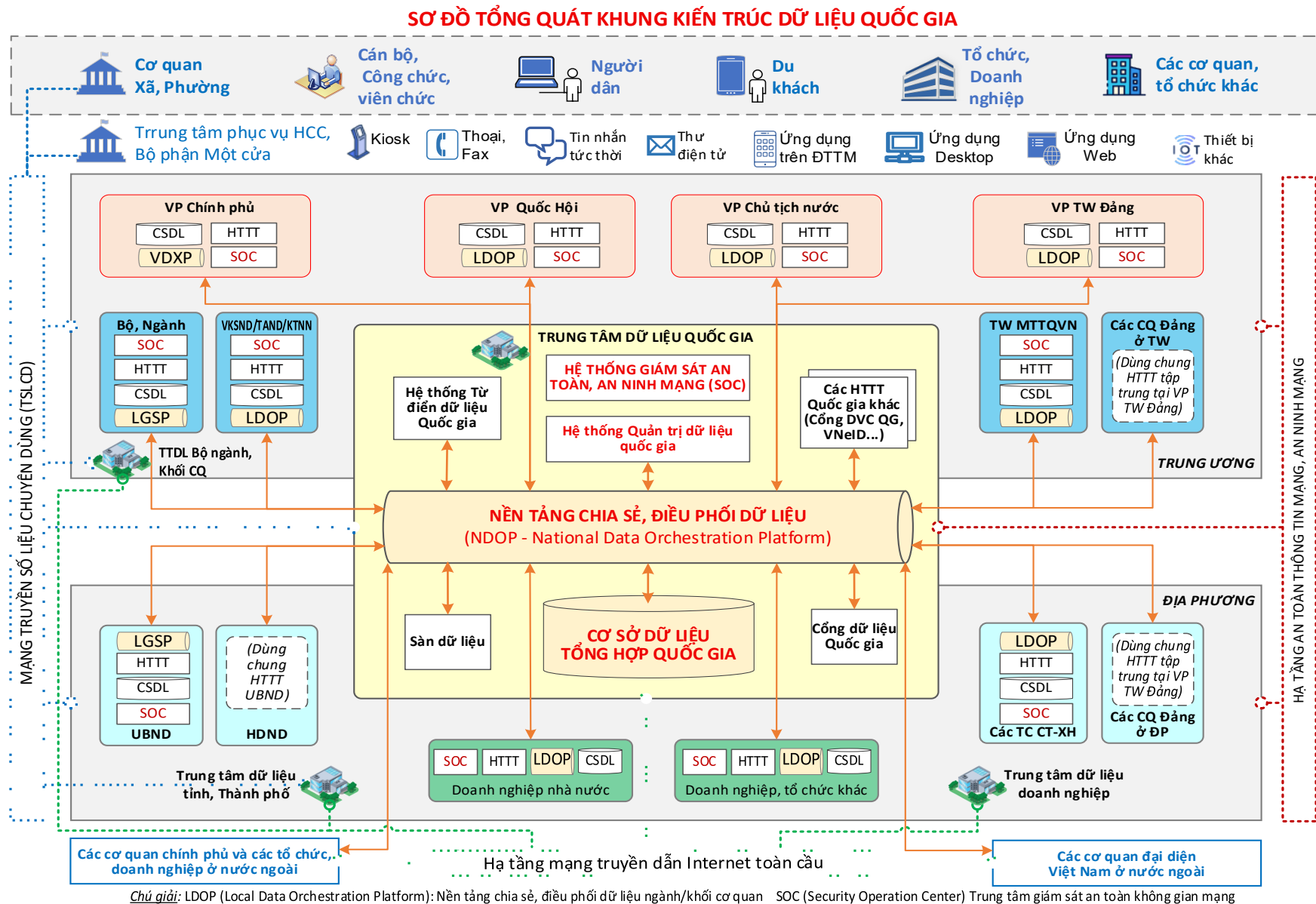
b) Yêu cầu về an toàn, bảo mật và phân quyền tuân thủ các quy định: Luật An ninh mạng, Luật An toàn thông tin, Luật Bảo vệ dữ liệu cá nhân, Luật Bảo vệ bí mật nhà nước, Luật Dữ liệu và các văn bản quy phạm pháp luật khác có liên quan.

c) Khung kiến trúc dữ liệu quốc gia được đánh giá và cập nhật định kỳ theo quy định về quản trị, quản lý dữ liệu.

III. KHUNG KIẾN TRÚC DỮ LIỆU QUỐC GIA

1. Sơ đồ tổng quát khung kiến trúc dữ liệu quốc gia

a) Sơ đồ tổng quát



Hình 1: Sơ đồ khái quát khung kiến trúc dữ liệu quốc gia

b) Mô tả khái quát

Sơ đồ tổng quát khung kiến trúc dữ liệu quốc gia bao gồm các thành phần sau:

- Người sử dụng (người thụ hưởng) bao gồm người dân, cán bộ, công chức, viên chức, du khách; các tổ chức, doanh nghiệp, các cơ quan, tổ chức khác; các cơ quan xã, phường;
- Các kênh giao tiếp bao gồm phương thức giao tiếp tại các Trung tâm phục vụ hành chính công, Bộ phận một cửa và các kênh số như kiosk, điện thoại, tin nhắn, thư điện tử, ứng dụng trên điện thoại thông minh, ứng dụng trên máy tính để bàn (Desktop), ứng dụng web hoặc các thiết bị khác;
- Các thành phần dữ liệu bao gồm: cơ sở dữ liệu (CSDL) tổng hợp quốc gia nằm ở trung tâm; các cơ sở dữ liệu quốc gia, cơ sở dữ liệu dùng chung và cơ sở dữ liệu chuyên ngành, hệ thống thông tin (HTTT) của các cơ quan là các hệ thống vệ tinh để cung cấp dữ liệu cho Cơ sở dữ liệu tổng hợp quốc gia;
- Nền tảng chia sẻ, điều phối dữ liệu (tên tiếng Anh là National Data Orchestration Platform, viết tắt là NDOP) - tất cả hoạt động kết nối, chia sẻ dữ liệu bắt buộc đều phải thực hiện thông qua Nền tảng chia sẻ, điều phối dữ liệu để thực hiện việc giám sát, truy vết và đánh giá hiệu quả;
- Các hệ thống thông tin quốc gia tiêu biểu phục vụ quản trị, quản lý dữ liệu quốc gia bao gồm: hệ thống giám sát an toàn, an ninh mạng (SOC - Security Operation Center) của Trung tâm dữ liệu quốc gia (TTDLQG); hệ thống Quản trị dữ liệu Quốc gia; hệ thống Từ điển dữ liệu dùng chung; hệ thống Cổng dữ liệu Quốc gia; sàn dữ liệu của TTDLQG. Các hệ thống thông tin quốc gia tiêu biểu khác kết nối với Nền tảng Chia sẻ, điều phối dữ liệu như Cổng Dịch vụ công Quốc gia; Nền tảng VNeID/Nền tảng định danh và xác thực điện tử...;
- Các cơ quan ở Trung ương kết nối với Nền tảng NDOP bao gồm các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan trung ương của tổ chức chính trị, Viện Kiểm sát nhân dân tối cao, Tòa án nhân dân tối cao, Kiểm toán Nhà nước, Văn phòng Chủ tịch nước, Văn phòng Quốc hội, Văn Phòng Trung ương Đảng, cơ quan Trung ương thuộc Mặt trận Tổ quốc Việt Nam;
- Các cơ quan địa phương kết nối với Nền tảng NDOP bao gồm UBND và HĐND các tỉnh, thành phố; các cơ quan Đảng ở địa phương, các tổ chức chính trị - xã hội; các doanh nghiệp nhà nước và các tổ chức, doanh nghiệp khác cũng được kết nối vào Nền tảng NDOP;
- Cơ quan đại diện Việt Nam ở nước ngoài kết nối tới Nền tảng NDOP để trao đổi dữ liệu phục vụ thương mại, hợp tác quốc tế;

- Hạ tầng truyền dẫn bao gồm hạ tầng mạng truyền dẫn Internet toàn cầu; hạ tầng mạng truyền số liệu chuyên dùng bảo đảm liên thông, đồng bộ, bí mật trong hoạt động của Đảng, Nhà nước và các cơ quan trong hệ thống chính trị;

- Hạ tầng Trung tâm dữ liệu bao gồm Trung tâm dữ liệu quốc gia, các trung tâm dữ liệu của bộ, ngành, khối cơ quan; các trung tâm dữ liệu của tỉnh/thành phố; các trung tâm dữ liệu của doanh nghiệp;

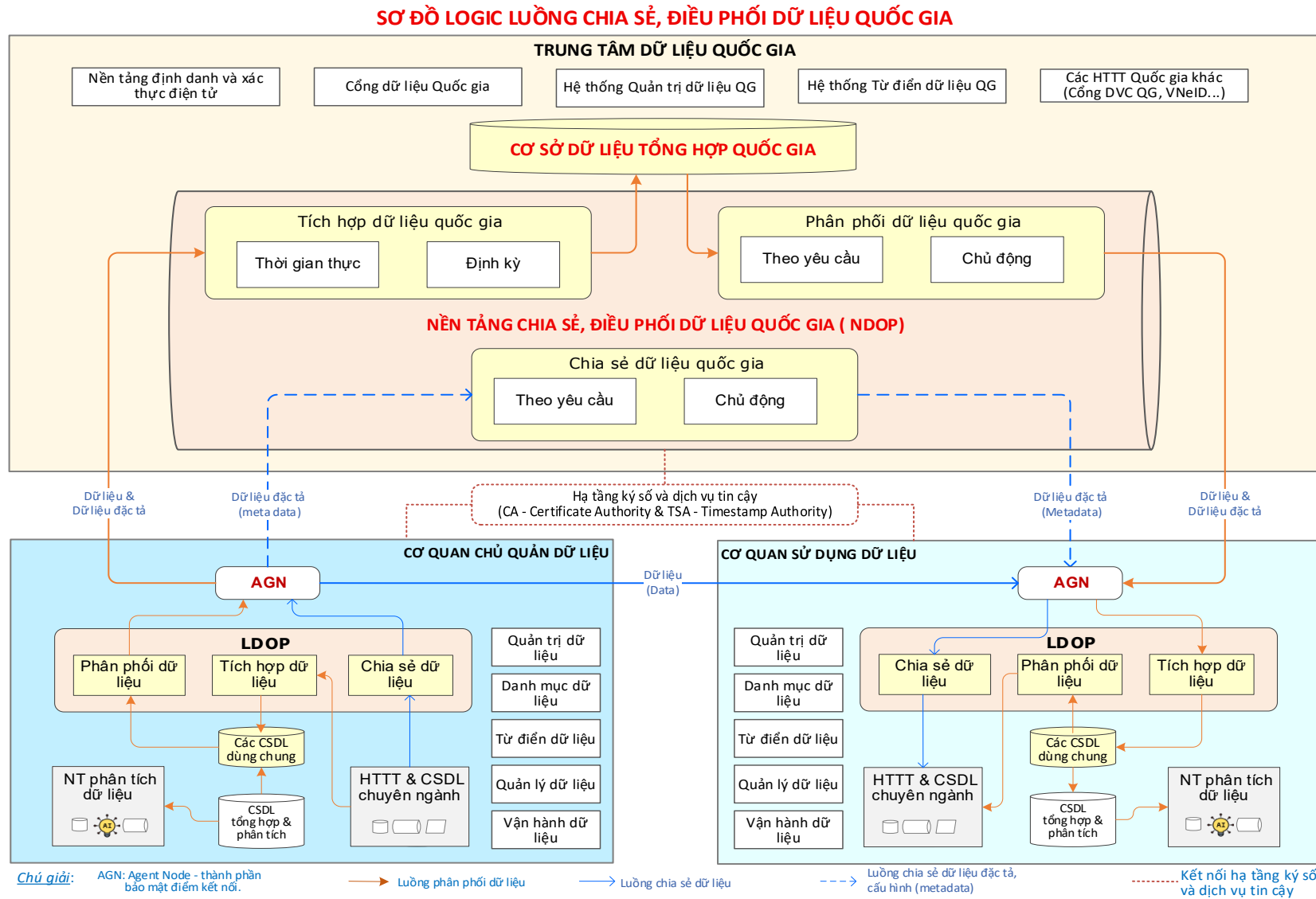
- Hạ tầng An toàn thông tin mạng, an ninh mạng bao gồm các hạ tầng và hệ thống bảo mật, kết nối với Trung tâm giám sát an toàn không gian mạng (SOC) của Trung tâm dữ liệu quốc gia và các SOC của các cơ quan trong hệ thống chính trị, bảo đảm an toàn cho toàn bộ hạ tầng công nghệ thông tin của quốc gia từ trung ương đến địa phương.

c) Các nền tảng tích hợp, chia sẻ dữ liệu của Văn phòng Chính phủ (VDXP - Vietnam Data eXchange Platform), các cơ quan trực thuộc Chính phủ và ủy ban nhân dân các tỉnh, thành phố (LGSP - Local Government Service Platform) cần được nâng cấp về kiến trúc để phù hợp với Khung kiến trúc dữ liệu quốc gia.

2. Sơ đồ logic luồng chia sẻ, điều phối dữ liệu quốc gia

a) Sơ đồ logic

Sơ đồ logic được thể hiện trong Hình 2 sau:



Hình 2: Sơ đồ logic luồng chia sẻ, điều phối dữ liệu quốc gia

b) Mô tả khái quát

Nền tảng chia sẻ, điều phối dữ liệu quốc gia NDOP cung cấp các thành phần chính:

- Thành phần Tích hợp dữ liệu quốc gia: thực hiện chức năng thu thập, tích hợp dữ liệu từ các cơ quan chủ quản dữ liệu về Cơ sở dữ liệu tổng hợp quốc gia theo thời gian thực hoặc định kỳ.

- Thành phần Phân phối dữ liệu quốc gia: thực hiện chức năng phân phối dữ liệu từ Cơ sở dữ liệu tổng hợp quốc gia tới cơ quan sử dụng dữ liệu theo yêu cầu hoặc chủ động.

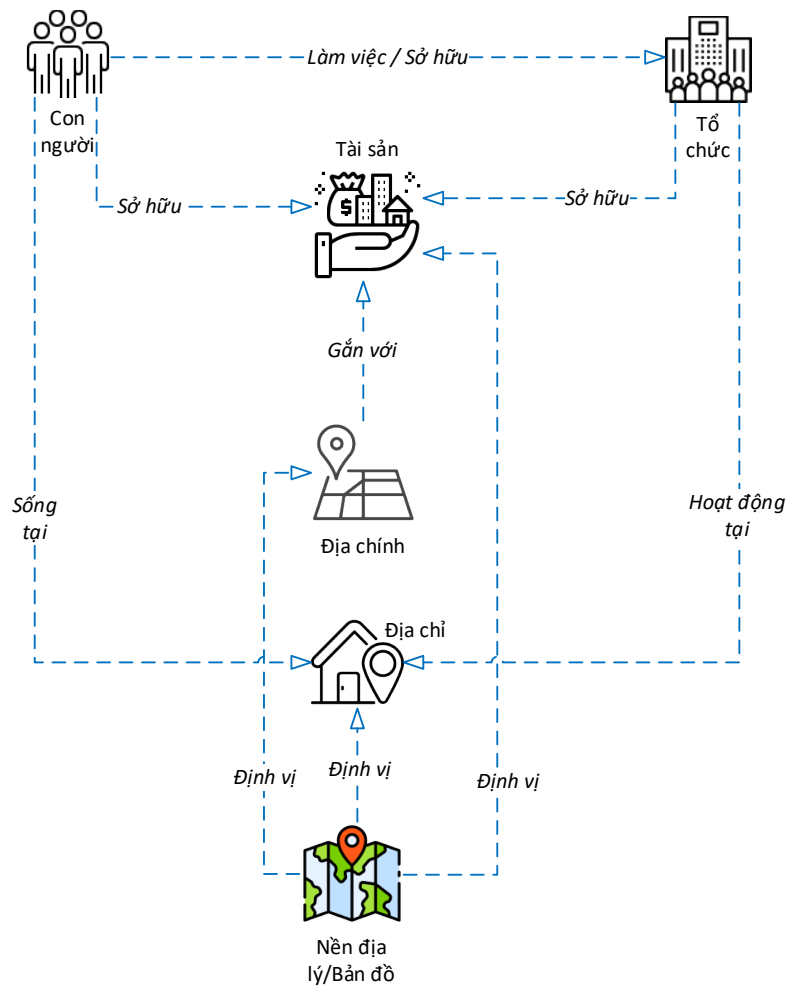
- Thành phần Chia sẻ dữ liệu quốc gia: thực hiện chức năng kết nối, giám sát và trung gian cho phép các cơ quan trao đổi dữ liệu trực tiếp với nhau sau khi xác thực và đăng ký dữ liệu và dịch vụ trao đổi dữ liệu.

- Thành phần bảo mật điểm kết nối AGN (Agent Node) là thành phần của nền tảng NDOP được triển khai tại các cơ quan, tổ chức, bộ ngành, địa phương bảo đảm về an toàn bảo mật điểm kết nối và trao đổi dữ liệu giữa nền tảng LDOP với nền tảng NDOP tại Trung tâm dữ liệu quốc gia để thực hiện điều phối dữ liệu.

Nền tảng LDOP bao gồm các thành phần cơ bản có chức năng tương tự như nền tảng NDOP (tích hợp dữ liệu; phân phối dữ liệu; chia sẻ dữ liệu) và trao đổi dữ liệu với Nền tảng NDOP thông qua thành phần bảo mật điểm kết nối AGN.

3. Mô hình dữ liệu cơ bản

Mô hình dữ liệu cơ bản thể hiện các miền dữ liệu cốt lõi của quốc gia ở mức khái niệm. Trong đó, miền dữ liệu con người là trung tâm và có mối quan hệ với các miền dữ liệu cơ bản khác thể hiện trong lược đồ (Hình 3) sau:



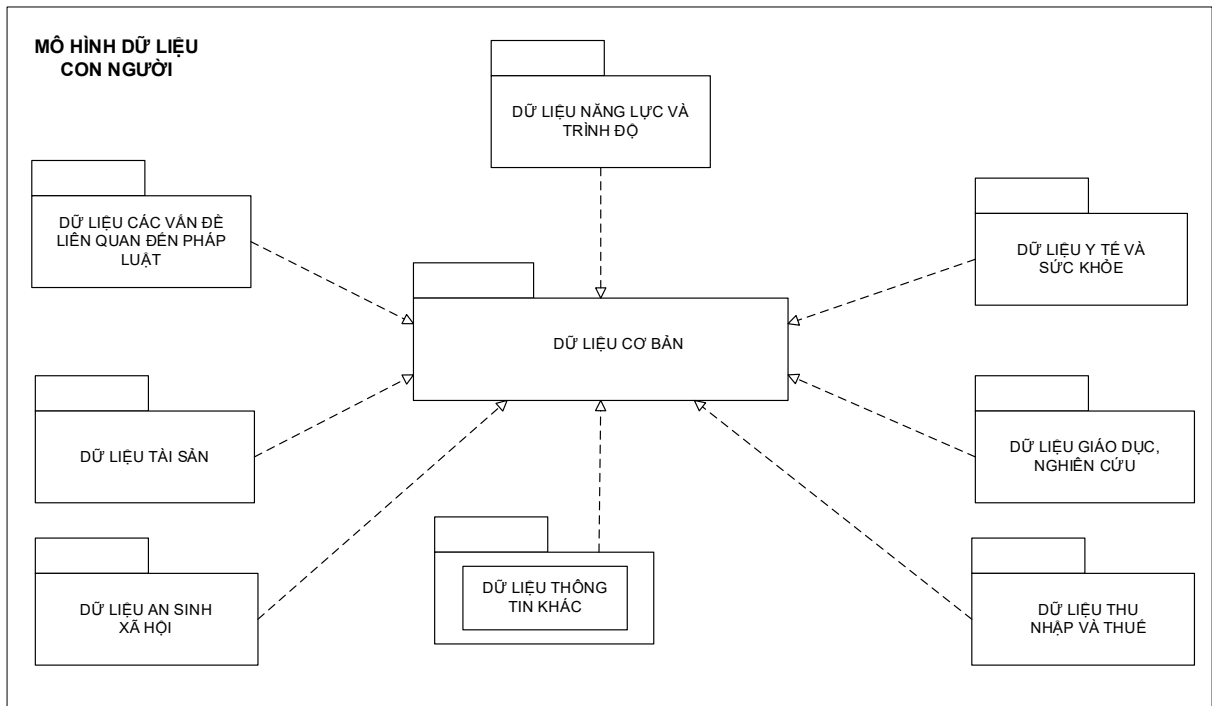
Hình 3: Mô hình dữ liệu cơ bản

Các miền dữ liệu theo từng lĩnh vực được quản lý bởi đơn vị chủ quản của lĩnh vực đó, theo chức năng, nhiệm vụ quản lý nhà nước. Các miền dữ liệu được xem là dữ liệu cơ bản (hoặc dữ liệu lõi) bao gồm các dữ liệu phục vụ đa mục tiêu, không chỉ giúp điều phối dữ liệu giữa các cơ quan quản lý nhà nước, phục vụ các nghiệp vụ chuyên ngành, hỗ trợ xây dựng chính sách, cung cấp dịch vụ công, cung cấp thông tin minh bạch phục vụ phát triển kinh tế - xã hội.

Các miền dữ liệu cơ bản được ưu tiên phát triển đến năm 2030 tại Trung tâm dữ liệu quốc gia (Mô hình logic của dữ liệu được mô tả trong Thông tư số 08/2025/TT-BCA của Bộ Công an về việc ban hành Quy chuẩn kỹ thuật quốc gia về cấu trúc thông điệp dữ liệu trao đổi với Cơ sở dữ liệu tổng hợp quốc gia), bao gồm 06 miền dữ liệu cơ bản sau đây:

a) Dữ liệu về Con người

Dữ liệu về Con người bao gồm: dữ liệu về chủ thể là con người và các dữ liệu liên quan đến chủ thể đó như dữ liệu dân cư, bảo hiểm, y tế và an sinh xã hội, giáo dục và đào tạo, lao động, việc làm, hộ tịch và thu thập tối đa các thông tin lĩnh vực khác như vi phạm hành chính, thuế, thu nhập cá nhân...



Hình 4: Mô hình dữ liệu về con người

Thành phần mô hình dữ liệu về con người (Hình 4):

- Dữ liệu cơ bản: các thông tin cơ bản gắn với một con người, bao gồm các thông tin đăng ký cơ bản như: họ tên, ngày tháng năm sinh, nơi sinh, quê quán, tôn giáo, quốc tịch; các loại giấy tờ định danh như: thẻ căn cước, chứng minh nhân dân 12 số, số chứng minh thư ngoại giao, số chứng minh thư công vụ; thông tin nhân dạng, thông tin về hộ gia đình, các thành viên trong gia đình/người giám hộ, tình trạng hôn nhân, thông tin liên lạc như: địa chỉ liên hệ, email, số điện thoại và các thông tin khác như: chứng sinh, khai sinh, khai tử, thông tin sinh trắc học, nhóm máu, tình trạng khuyết tật.

- Dữ liệu an sinh xã hội: các thông tin về đối tượng người có công, đối tượng nhận trợ giúp xã hội và các chính sách xã hội liên quan, thông tin về quá trình đóng, quá trình hưởng bảo hiểm xã hội, bảo hiểm tự nguyện, bảo hiểm y tế, thông tin về người nghiện ma túy trong phòng, chống tệ nạn xã hội, tình hình bạo lực gia đình, các thành tích trong văn hóa thể thao, tình hình về lao động và việc làm của cá nhân như: thông tin về người lao động, quá trình công tác, thông tin giấy phép lao động, người lao động Việt Nam ở nước ngoài, người lao động nước ngoài ở Việt Nam.

- Dữ liệu tài sản: các thông tin về tình hình sở hữu các loại tài sản của một con người, từ tài sản tài chính như: vốn góp, cổ phần, các khoản tiền gửi ngân hàng; bất động sản như: đất đai, tài sản gắn liền với đất; động sản như: phương tiện giao thông đường bộ, đường thủy, đường sắt; tài sản vô hình như: các quyền sở hữu công nghiệp, quyền với giống cây trồng, quyền tác giả.

- Dữ liệu các vấn đề liên quan đến pháp luật: thông tin về các bản án/quyết định của tòa án như: bản án hình Dân sự/Hành chính/Kinh doanh thương mại/Lao động; quyết định xử phạt vi phạm hành chính, thông tin về các vụ khiếu kiện và các thông tin tư pháp khác như các vụ việc được trợ giúp pháp lý, việc thi hành án dân sự, hành chính, hình sự.

- Dữ liệu năng lực và trình độ: thông tin về trình độ như: trình độ chuyên môn, trình độ học vấn, trình độ lý luận chính trị, trình độ ngoại ngữ; các văn bằng, chứng chỉ của cá nhân, các kết quả đánh giá chuẩn nghề nghiệp và kỹ năng nghề, các danh hiệu được phong tặng, thông tin khen thưởng, kỷ luật.

- Dữ liệu y tế và sức khỏe: các thông tin của các đợt khám, chữa bệnh của cá nhân, thông tin được tiêm chủng, thông tin về người bị phát hiện nhiễm HIV.

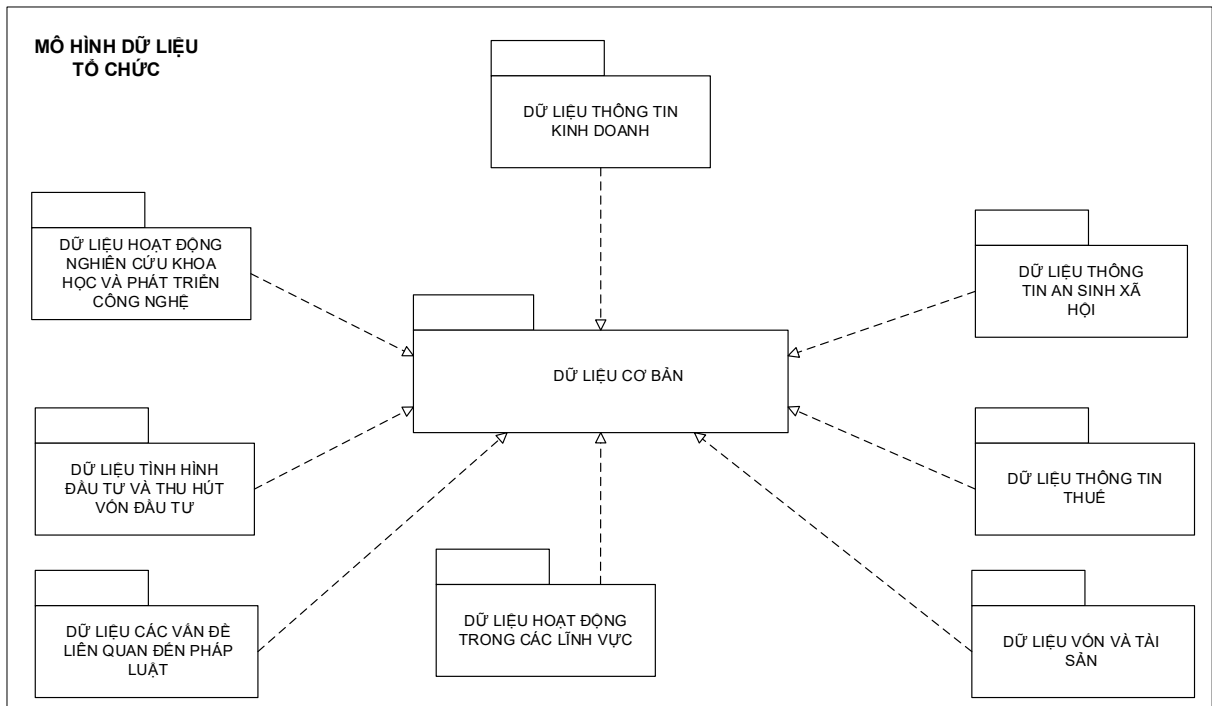
- Dữ liệu giáo dục, nghiên cứu: thông tin về quá trình và kết quả học tập của cá nhân từ mầm non, đến trung học phổ thông và khi là sinh viên, học viên các trường cao đẳng, đại học, cơ sở giáo dục nghề nghiệp, thông tin về hoạt động khoa học công nghệ và các nhiệm vụ khoa học công nghệ đã tham gia, các công bố khoa học công nghệ và các sáng chế, giải pháp hữu ích của cá nhân.

- Dữ liệu thu nhập và thuế: thông tin về thuế của người nộp thuế là cá nhân như: đăng ký thuế, tình hình thu nộp thuế; thu nhập từ hoạt động sản xuất kinh doanh của hộ kinh doanh, thu nhập từ tiền lương, tiền công và các thu nhập khác.

- Dữ liệu thông tin khác về cá nhân như thông tin tham gia hoặc có quan hệ với tổ chức chính trị, kinh tế, xã hội ở nước ngoài.

b) Dữ liệu về tổ chức

Dữ liệu về tổ chức bao gồm dữ liệu về doanh nghiệp, các cơ quan nhà nước, các tổ chức khác hình thành từ dữ liệu về đăng ký doanh nghiệp (Bộ Tài chính), dữ liệu về định danh tổ chức do Bộ Công an triển khai, dữ liệu về doanh nghiệp của Bộ Tài chính và các dữ liệu khác liên quan tới các tổ chức.



Hình 5: Mô hình dữ liệu về tổ chức

Thành phần mô hình dữ liệu về tổ chức (Hình 5):

- Thông tin cơ bản: các thông tin cơ bản gắn với một tổ chức, là các thông tin đăng ký cơ bản như: tên tổ chức, loại hình, lĩnh vực kinh doanh, tình trạng thành lập, định danh điện tử, mã số thuế; các loại giấy tờ định danh như: Giấy chứng nhận đăng ký kinh doanh, Giấy chứng nhận đăng ký thuế, Giấy chứng nhận đăng ký mẫu con dấu, Giấy chứng nhận đầu tư, Quyết định thành lập; các cá nhân, tổ chức có liên quan, các thông tin đăng ký kinh doanh, vốn điều lệ và vốn góp của tổ chức.

- Dữ liệu các vấn đề liên quan đến pháp luật: thông tin về tình hình vi phạm hành chính, các bản án, quyết định của tòa án trên các lĩnh vực hình sự, dân sự, hành chính, kinh doanh thương mại, lao động.

- Dữ liệu tình hình đầu tư và thu hút vốn đầu tư: thông tin về các nhà đầu tư nước ngoài, trong nước và các dự án đầu tư, thông tin về tình hình đầu tư ra nước ngoài.

- Dữ liệu về hoạt động nghiên cứu khoa học và phát triển công nghệ: thông tin về việc tham gia hoạt động nghiên cứu khoa học và phát triển công nghệ, các kết quả có được và tình hình tham gia tổ chức quốc tế, là đối tác quốc tế trong hoạt động khoa học công nghệ.

- Dữ liệu, thông tin kinh doanh: các thông tin trên báo cáo tài chính của tổ chức như: Nguồn vốn chủ sở hữu, Chi phí, Hàng tồn kho, Nợ phải trả, Tài sản cố định và đầu tư dài hạn; thông tin về các giấy phép hoạt động sản xuất, kinh doanh mà tổ chức được cấp.

- Dữ liệu, thông tin về an sinh xã hội: thông tin về số lượng người lao động tại tổ chức, về tình hình đóng và chậm đóng bảo hiểm xã hội.

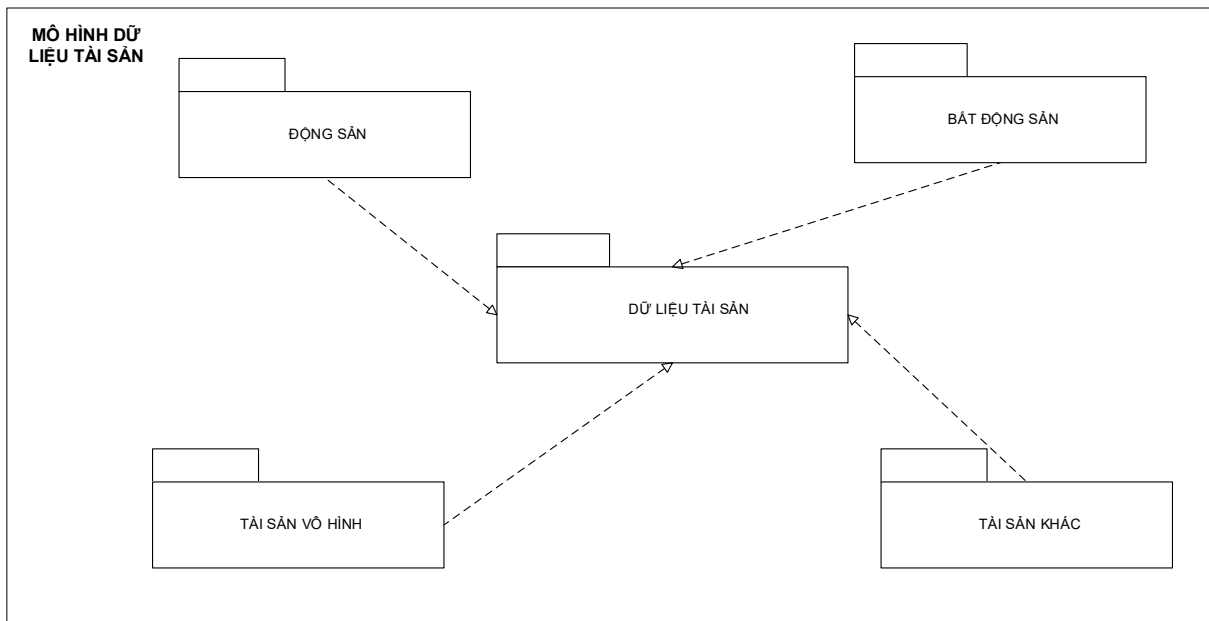
- Dữ liệu thông tin về thuế: các thông tin về đăng ký thuế, tình hình thu nộp thuế như: thuế nội địa, thuế xuất nhập khẩu; tình hình nợ, cường chế nợ thuế.

- Dữ liệu vốn và tài sản: các thông tin về tình hình sở hữu các loại tài sản của một tổ chức, từ tài sản tài chính, bất động sản như: đất đai, tài sản gắn liền với đất; động sản như: phương tiện giao thông đường bộ, đường thủy, đường sắt; tài sản vô hình như: các quyền sở hữu công nghiệp, quyền với giống cây trồng, quyền tác giả; thông tin về tình hình cấp tín dụng.

- Dữ liệu hoạt động trong các lĩnh vực: thông tin về tổ chức trong các lĩnh vực hoạt động khác nhau như: cơ sở y tế, cơ quan báo in, phát thanh truyền hình, công ty cung cấp nước sạch.

c) Dữ liệu về tài sản

Dữ liệu về tài sản bao gồm dữ liệu về các loại tài sản, bao gồm bất động sản như tòa nhà, nhà ở; động sản và các loại tài sản khác.



Hình 6: Mô hình dữ liệu về tài sản

Thành phần mô hình dữ liệu về tài sản (Hình 6):

- Bất động sản: thông tin về nhà, công trình xây dựng gắn liền với đất đai và các tài sản khác gắn liền với đất đai, nhà, công trình xây dựng, công trình ngầm; thông tin về quyền và hạn chế quyền sử dụng đất, quyền quản lý đất, quyền sở hữu tài sản gắn liền với đất; các thông tin về nghĩa vụ tài chính như nợ, miễn giảm nghĩa vụ tài chính.

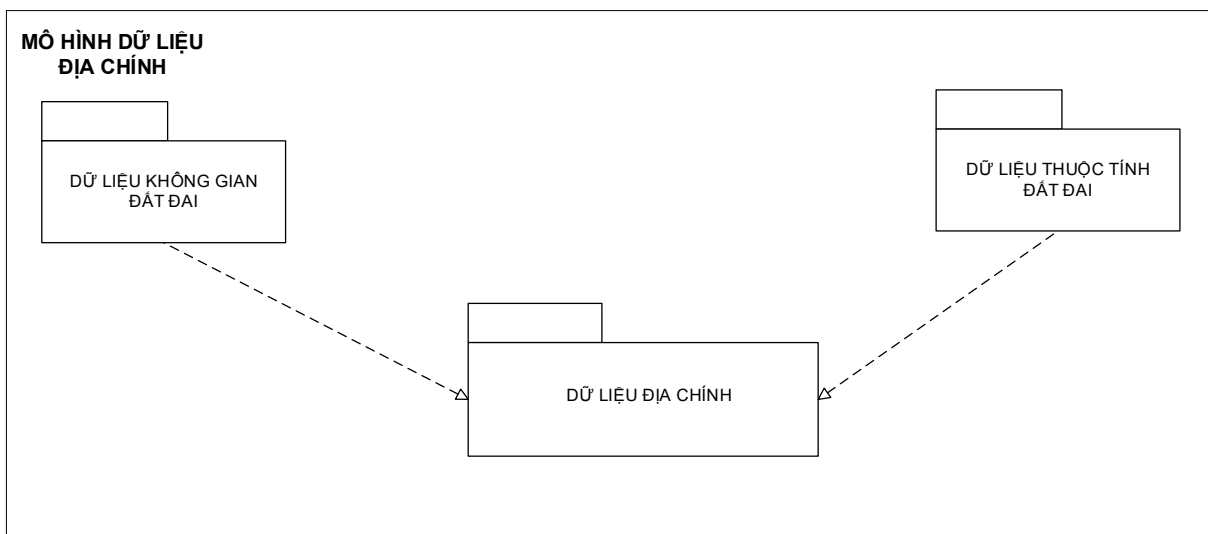
- Động sản: thông tin về các tài sản là động sản như xe mô tô, xe gắn máy, xe ô tô, rơ moóc, sơ mi rơ moóc và xe chở người bốn bánh có gắn động cơ, phương tiện đường sắt, phương tiện đường thủy.

- Tài sản vô hình: thông tin về các tài sản là quyền đối với giống cây trồng; quyền sở hữu công nghiệp (như Bằng độc quyền sáng chế, Bằng độc quyền giải pháp hữu ích, Bằng độc quyền kiểu dáng công nghiệp, Giấy chứng nhận đăng ký thiết kế bố trí mạch tích hợp bán dẫn, Giấy chứng nhận đăng ký nhãn hiệu và Giấy chứng nhận đăng ký chỉ dẫn địa lý); quyền tác giả quyền liên quan (tác phẩm văn học, khoa học, tác phẩm mỹ thuật ứng dụng, chương trình máy tính, tác phẩm âm nhạc, tác phẩm mỹ thuật, tác phẩm nhiếp ảnh, tác phẩm điện ảnh, tác phẩm sân khấu, sách giáo khoa, giáo trình, tác phẩm kiến trúc, bản họa đồ, sơ đồ, bản đồ, bản vẽ liên quan đến địa hình, kiến trúc, công trình khoa học).

- Tài sản khác: thông tin về tài sản khác của cá nhân, tổ chức như di sản văn hóa được xác lập sở hữu riêng, sở hữu chung mà cá nhân, tổ chức, gia đình, dòng họ, cộng đồng sở hữu.

d) Dữ liệu về địa chính

Dữ liệu về địa chính bao gồm các dữ liệu về các ô, thửa đất, ranh giới, đặc điểm không gian.



Hình 7. Mô hình dữ liệu về địa chính

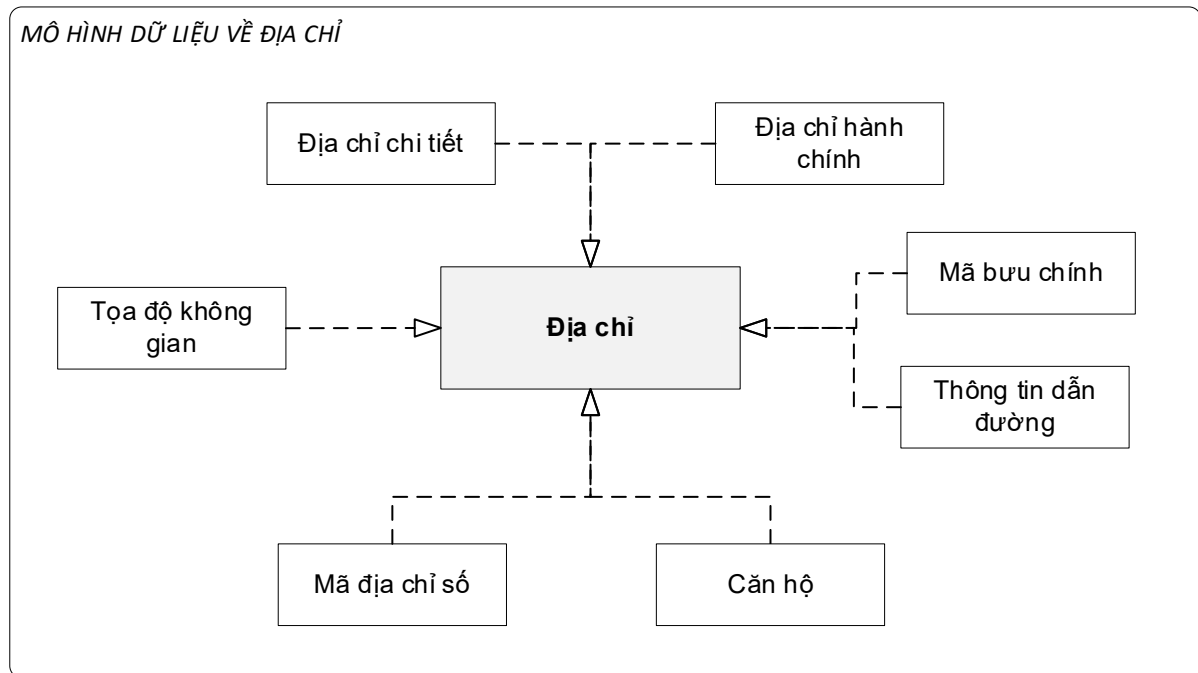
Thành phần Mô hình dữ liệu về địa chính (Hình 7):

- Dữ liệu thuộc tính đất đai là các thông tin cơ bản về thửa đất như Số hiệu tờ bản đồ, Số thửa đất, nguồn gốc sử dụng, loại đất, thông tin về hồ sơ địa chính, giá thửa đất...

- Dữ liệu không gian đất đai: gồm các Thông tin không gian địa chính; thông tin không gian giá đất; dữ liệu về chất lượng đất, tiềm năng đất đai cấp vùng kinh tế - xã hội; dữ liệu về thoái hóa đất cấp vùng kinh tế - xã hội; dữ liệu về ô nhiễm đất cấp vùng kinh tế - xã hội; dữ liệu về xử lý, bảo vệ, cải tạo, phục hồi đất cấp vùng kinh tế - xã hội.

đ) Dữ liệu về địa chỉ

Dữ liệu về địa chỉ bao gồm dữ liệu về địa chỉ của đường xá, khu vực..., thông tin tọa độ không gian.



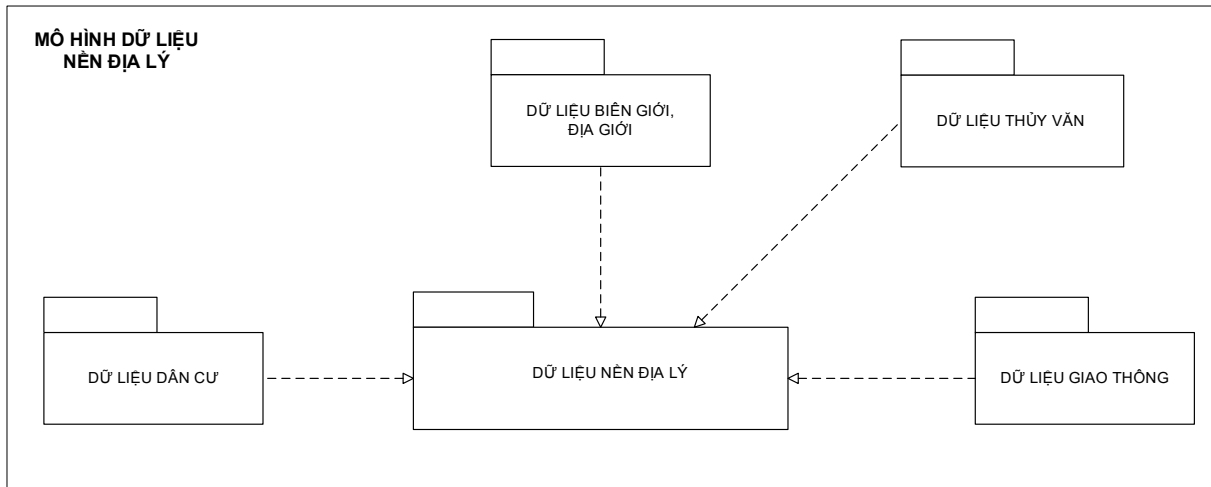
Hình 8: Mô hình dữ liệu về địa chỉ

Thành phần mô hình dữ liệu về địa chỉ (Hình 8):

- Mã địa chỉ số: mã địa chỉ số.
- Địa chỉ hành chính: thông tin về Thôn/Tổ, Phường/Xã, Tỉnh/Thành phố hành chính của địa chỉ.
- Địa chỉ chi tiết: thông tin về Phân loại địa điểm (nhà ở, cơ quan, ...), Số nhà, tên địa điểm, Tên đường.
- Mã bưu chính: cấp, Mã, Tên.
- Căn hộ: thông tin về tên Khu chung cư, Phòng/Cửa, Tầng, Tòa nhà.
- Thông tin dẫn đường: thông tin về Biển báo cho từng làn, Độ rộng, Giới hạn loại xe, Giới hạn thời gian, Giới hạn tốc độ, Mã làn đường, Vạch kẻ đường, Vạch làn chỉ dẫn từng đoạn, Vị trí (số thứ tự so với mép trái).
- Tọa độ không gian: thông tin về Loại điểm tính tọa độ không gian, Kinh độ của điểm, vĩ độ của điểm.

e) Dữ liệu nền địa lý

Dữ liệu nền địa lý bao gồm các dữ liệu không gian và địa lý như vị trí, đặc điểm địa lý và các yếu tố môi trường, địa hình.



Hình 9: Mô hình dữ liệu về nền địa lý

Thành phần mô hình dữ liệu về nền địa lý (Hình 9):

- Dữ liệu dân cư: là dữ liệu về các khu dân cư và các công trình liên quan đến dân cư, gồm khu dân cư đô thị, khu dân cư nông thôn, khu công nghiệp, khu chức năng đặc thù, khu dân cư khác, các công trình dân sinh, cơ sở kinh tế, văn hóa, xã hội khác;

- Dữ liệu biên giới địa giới: bao gồm dữ liệu biên giới quốc gia và dữ liệu địa giới hành chính. Dữ liệu biên giới quốc gia là dữ liệu về đường biên giới quốc gia, hệ thống mốc quốc giới; điểm cơ sở, đường cơ sở dùng để tính chiều rộng lãnh hải Việt Nam, điểm đặc trưng biên giới quốc gia, các đối tượng địa lý liên quan đến việc thể hiện đường biên giới quốc gia trên đất liền, trên biển. Dữ liệu địa giới hành chính là dữ liệu về đường địa giới hành chính các cấp, hệ thống mốc địa giới hành chính; dữ liệu về các đối tượng địa lý liên quan đến việc thể hiện đường địa giới hành chính các cấp;

- Dữ liệu thủy văn: là dữ liệu về hệ thống sông, suối, kênh, mương, biển, hồ, ao, đầm, phá, nguồn nước, đường bờ nước, các đối tượng thủy văn khác.

- Dữ liệu giao thông: là dữ liệu về mạng lưới giao thông và các công trình có liên quan đến giao thông, gồm hệ thống đường bộ, đường sắt, đường hàng không, đường thủy, cầu, hầm giao thông, bến cảng, nhà ga, các công trình giao thông khác.

4. Danh mục dữ liệu quốc gia

Danh mục dữ liệu quốc gia (National Data Catalog) được công bố tại Cổng dữ liệu quốc gia bao gồm:

a) Dữ liệu chủ dùng chung quốc gia;

- b) Dữ liệu danh mục dùng chung quốc gia;
- c) Dữ liệu chuyên ngành dùng chung quốc gia;
- d) Dữ liệu tổng hợp quốc gia bao gồm dữ liệu lớn, dữ liệu tổng hợp, dữ liệu phân tích;
- đ) Dữ liệu mở quốc gia;
- e) Các dữ liệu khác theo quy định của pháp luật.

Từ điển dữ liệu dùng chung cung cấp dữ liệu đặc tả (siêu dữ liệu) về các dữ liệu được công bố trên Cổng dữ liệu quốc gia.

Danh mục dữ liệu dùng chung quốc gia được cập nhật thường xuyên cùng với quá trình xây dựng các cơ sở dữ liệu, các hệ thống thông tin có tạo lập, thu thập các dữ liệu mới.

Danh mục dữ liệu dùng chung quốc gia phiên bản 1.0 được quy định trong Phụ lục của Quyết định này.

5. Tổ chức các cơ sở dữ liệu dùng chung

a) Tổ chức các cơ sở dữ liệu dùng chung

Các cơ sở dữ liệu dùng chung được sắp xếp, tổ chức ở các cấp theo mô hình sau (Hình 10):

	Metadata (Từ điển dữ liệu)	Dữ liệu danh mục	Dữ liệu chủ	Dữ liệu giao dịch (ngh nghiệp vụ)	Dữ liệu lớn, DL tổng hợp, DL phân tích	Dữ liệu mở
Quốc tế	 Từ điển dữ liệu quốc tế	 DL danh mục quốc tế	 DL chủ chuẩn quốc tế	 DL chuyên ngành quốc tế	 Dữ liệu Tổng hợp quốc tế	 Dữ liệu mở quốc tế
Quốc gia	 Từ điển dữ liệu quốc gia	 Dữ liệu danh mục DC QG	 Dữ liệu chủ dùng chung QG	 Dữ liệu chuyên ngành dùng chung QG	 Dữ liệu Tổng hợp QG	 Dữ liệu mở QG
Bộ, ngành, Khối cơ quan	 Từ điển dữ liệu ngành	 DL DMDC ngành	 DL chủ dùng chung ngành	 Dữ liệu chuyên ngành	 DLTổng hợp ngành	 DL mở ngành
Tỉnh, thành phố	 Từ điển dữ liệu địa phương	 DL DMDC địa phương	 DL chủ dùng chung địa phương	 Dữ liệu địa phương	 Dữ liệu Tổng hợp địa phương	 Dữ liệu mở địa phương
Xã, phường						
Nội bộ						
Hệ thống thông tin						

Hình 10: Các cơ sở dữ liệu dùng chung ở các cấp

(1) Cơ sở dữ liệu tổng hợp quốc gia

Cơ sở dữ liệu tổng hợp quốc gia bao gồm 08 lớp dữ liệu dùng chung ở phạm vi quốc gia và quốc tế được tổ chức thành các kho dữ liệu dùng chung gồm:

- Kho dữ liệu Từ điển dữ liệu dùng chung;
- Kho dữ liệu danh mục dùng chung quốc gia;
- Kho dữ liệu chủ dùng chung quốc gia;
- Kho dữ liệu chuyên ngành dùng chung quốc gia;
- Kho dữ liệu tổng hợp quốc gia bao gồm dữ liệu lớn, dữ liệu tổng hợp, dữ liệu phân tích;
- Kho dữ liệu mở quốc gia.

(2) Các cơ sở dữ liệu dùng chung của mỗi bộ, ngành, địa phương, khối cơ quan

Các cơ sở dữ liệu dùng chung của mỗi bộ, ngành, địa phương, khối cơ quan liên quan bao gồm đủ 08 lớp dữ liệu tương ứng để đồng bộ lên Cơ sở dữ liệu tổng hợp quốc gia. Với mỗi lớp dữ liệu đó, bộ, ngành, địa phương, khối cơ quan cần quy định cụ thể các nhóm dữ liệu dùng chung ở mức quốc gia và nhóm dữ liệu đặc thù riêng của bộ, ngành và địa phương, khối cơ quan.

b) Các cơ sở dữ liệu cơ bản/cơ sở dữ liệu lõi

Các dữ liệu cơ bản được tổ chức thành các cơ sở dữ liệu cơ bản thuộc Cơ sở dữ liệu tổng hợp quốc gia để chia sẻ, điều phối trên phạm vi quốc gia cho các khối cơ quan trong hệ thống chính trị.

Sáu cơ sở dữ liệu cơ bản ưu tiên xây dựng bao gồm:

- Cơ sở dữ liệu về con người;
- Cơ sở dữ liệu về tổ chức;
- Cơ sở dữ liệu về tài sản;
- Cơ sở dữ liệu về địa chính;
- Cơ sở dữ liệu về địa chỉ;
- Cơ sở dữ liệu nền địa lý.

c) Các cơ sở dữ liệu quốc gia

Cơ sở dữ liệu quốc gia là cơ sở dữ liệu quản lý dữ liệu dùng chung quốc gia của một hoặc một số lĩnh vực kinh tế - xã hội được xây dựng, cập nhật và duy trì đáp ứng yêu cầu truy nhập và sử dụng thông tin các cơ quan trong hệ thống chính trị cũng như đáp ứng phát triển chính phủ số, kinh tế số, xã hội số theo quy định của pháp luật.

IV. CÁC MÔ HÌNH THAM CHIẾU

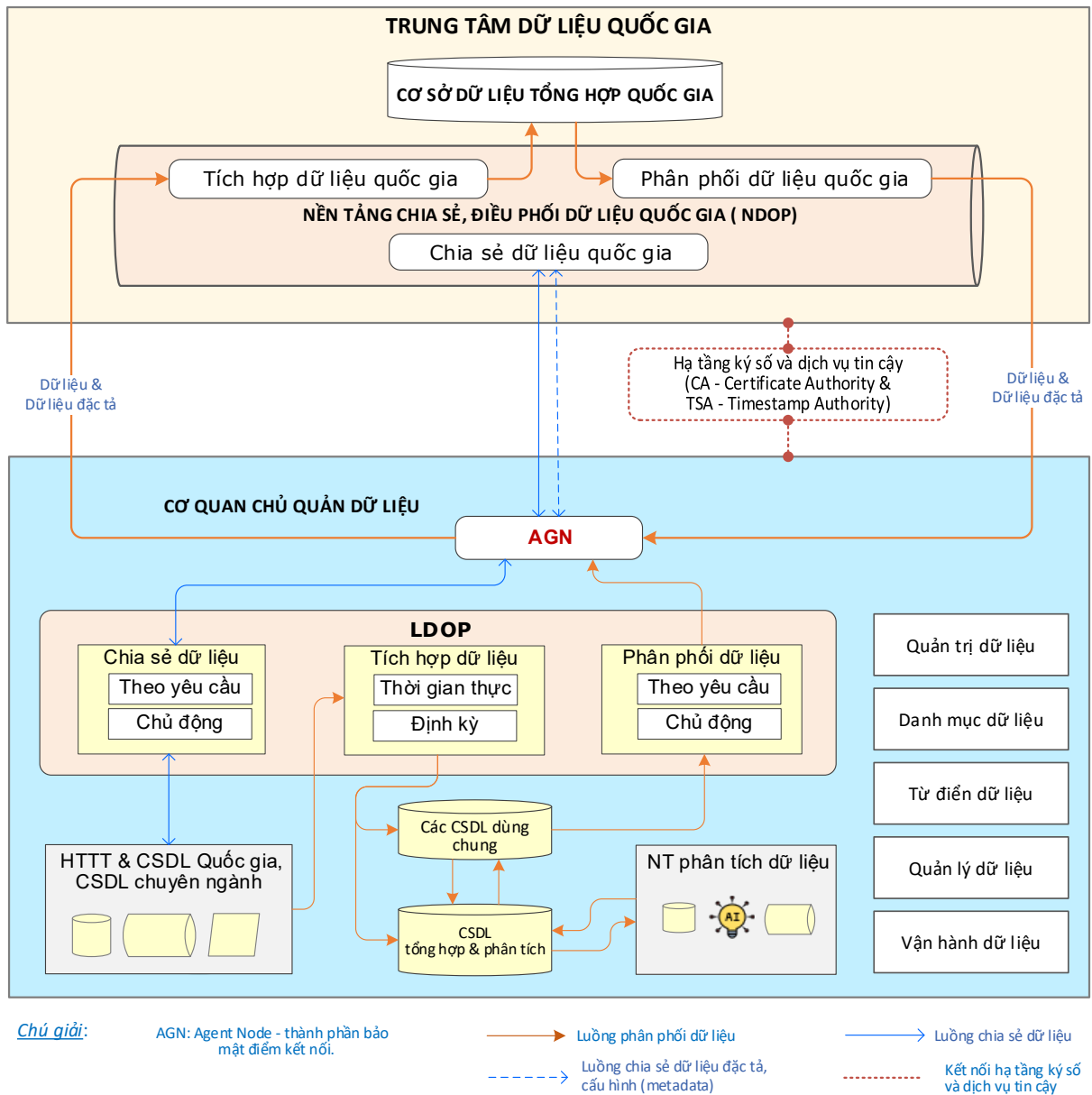
1. Mục đích các mô hình tham chiếu

Các cơ quan, các bộ, ngành, địa phương tham chiếu trong quá trình xây dựng Khung kiến trúc dữ liệu của khối cơ quan; Khung kiến trúc dữ liệu cấp bộ, ngành; Khung kiến trúc dữ liệu cấp tỉnh.

2. Sơ đồ tổng quát khung kiến trúc dữ liệu một khối cơ quan, bộ ngành, địa phương

a) Sơ đồ tham chiếu

**SƠ ĐỒ TỔNG QUÁT KHUNG KIẾN TRÚC DỮ LIỆU THAM CHIẾU
KHỐI CƠ QUAN, BỘ NGÀNH, ĐỊA PHƯƠNG**



**Hình 11: Sơ đồ tổng quát khung kiến trúc dữ liệu tham chiếu
khối cơ quan, bộ ngành, địa phương**

b) Mô tả

Các thành phần chính của Khung kiến trúc dữ liệu tham chiếu các cơ quan, bộ, ngành, địa phương:

- Nền tảng LDOP cung cấp các chức năng về chia sẻ dữ liệu, tích hợp dữ liệu và phân phối dữ liệu. Nền tảng LDOP có khả năng cung cấp phương thức chia sẻ dữ liệu/phân phối dữ liệu theo yêu cầu hoặc chủ động phù hợp với yêu cầu nghiệp vụ. Nền tảng LDOP cũng thực hiện chức năng tích hợp dữ liệu, thu thập dữ liệu từ các hệ thống thông tin, cơ sở dữ liệu quốc gia, cơ sở dữ liệu chuyên ngành về các cơ sở dữ liệu dùng chung, cơ sở dữ liệu/kho dữ liệu tổng hợp và phân tích.

- Thành phần bảo mật điểm kết nối AGN (Agent Node) bảo đảm về an toàn bảo mật, kết nối giữa nền tảng LDOP với nền tảng NDOP tại TTDLQG để thực hiện điều phối dữ liệu quốc gia.

- Các cơ sở dữ liệu dùng chung là nơi lưu trữ, quản lý dữ liệu để cung cấp, phân phối trong nội bộ cơ quan, bộ ngành, địa phương và cung cấp cho Cơ sở dữ liệu tổng hợp quốc gia, chia sẻ với các cơ quan, tổ chức khác.

- Cơ sở dữ liệu tổng hợp và phân tích lưu trữ, quản lý dữ liệu lớn, dữ liệu tổng hợp, dữ liệu phân tích phục vụ các nhu cầu về báo cáo, thống kê, phân tích dữ liệu.

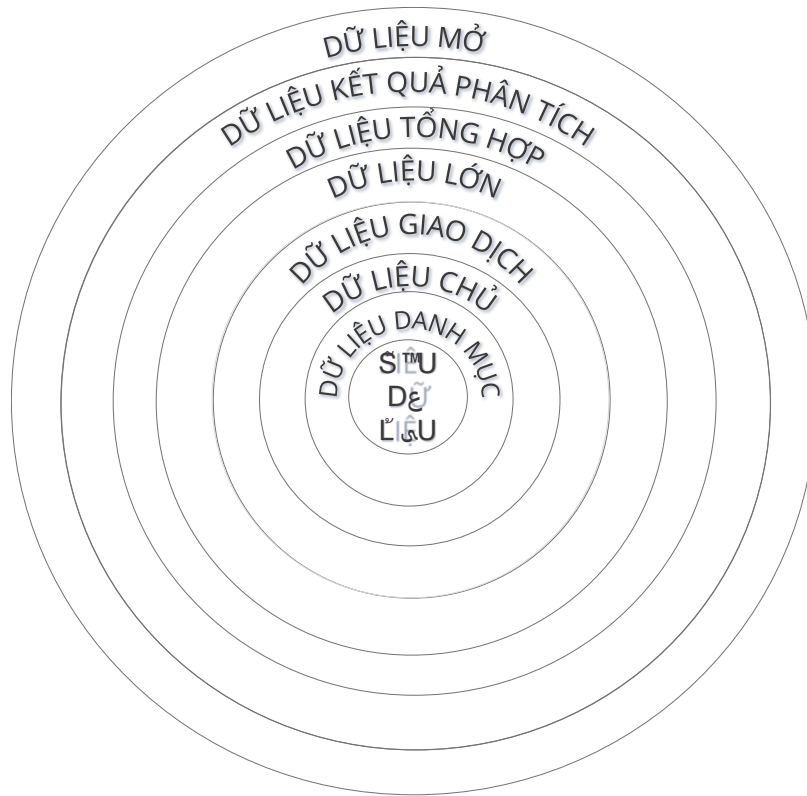
- Các thành phần chức năng đáp ứng yêu cầu về quản trị dữ liệu, quản lý danh mục dữ liệu, quản lý từ điển dữ liệu, quản lý dữ liệu và vận hành dữ liệu.

3. Mô hình dữ liệu khái niệm

a) Phân lớp dữ liệu dùng chung

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.

Mô hình dưới đây minh họa vai trò và sự phụ thuộc của các lớp dữ liệu trong mỗi cơ quan, tổ chức, làm cơ sở xác định các loại dữ liệu dùng chung:



Hình 12: Mô hình các lớp dữ liệu

Theo đó, mô hình các lớp dữ liệu thể hiện tính kế thừa, mở rộng và khả năng tích hợp dữ liệu: các lớp dữ liệu phía trong đóng vai trò nền tảng, là cơ sở gốc để thiết lập các lớp dữ liệu phía ngoài. Quá trình hình thành dữ liệu phía ngoài được thực hiện thông qua tham chiếu hoặc khai thác trực tiếp các dữ liệu phía trong, đồng thời bổ sung, làm phong phú và mở rộng phạm vi lớp dữ liệu phía trong, phục vụ các mục tiêu phân tích, quản lý và ra quyết định. Nguyên lý này giúp tăng cường hiệu quả vận hành, giảm trùng lặp thông tin và tạo tính nhất quán trong toàn bộ hệ sinh thái dữ liệu.

Theo mô hình các lớp dữ liệu, các loại dữ liệu dùng chung cần được xây dựng bao gồm:

- Siêu dữ liệu (metadata) hay dữ liệu đặc tả dữ liệu đóng vai trò cốt lõi vì được sử dụng để định nghĩa các dữ liệu khác. Siêu dữ liệu gồm:

- + Siêu dữ liệu về kỹ thuật (technical metadata), mô tả các tài nguyên dữ liệu tĩnh như bảng cơ sở dữ liệu, cột, kiểu dữ liệu và cấu trúc chỉ mục. Siêu dữ liệu kỹ thuật tồn tại trong từ điển dữ liệu kỹ thuật.

- + Siêu dữ liệu về hoạt động dữ liệu (operational metadata), mô tả luồng dữ liệu, di chuyển của dữ liệu và hoạt động điều phối luồng dữ liệu đó. Siêu dữ liệu hoạt động tồn tại dưới dạng luồng xử lý dữ liệu.

+ Siêu dữ liệu nghiệp vụ, mô tả từ vựng được sử dụng để tham khảo về dữ liệu. Siêu dữ liệu nghiệp vụ (business metadata) bao gồm các thuật ngữ riêng lẻ và các cấu trúc phức tạp hơn như hệ thống phân cấp và phân loại do người dùng công nhận. Siêu dữ liệu nghiệp vụ trùng lặp đáng kể với nội dung của bảng từ vựng nghiệp vụ (business glossary), được quy định trong từ điển dữ liệu nghiệp vụ; siêu dữ liệu xã hội là các loại dữ liệu mang tính chủ thích, bình luận...

- Dữ liệu danh mục dùng chung (reference data) 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. Dữ liệu danh mục dùng chung được sử dụng bảo đảm sự thống nhất, đồng bộ và tương thích giữa các hệ thống thông tin của cơ quan từ trung ương đến địa phương.

- Dữ liệu chủ (master data) là dữ liệu chứa thông tin cơ bản nhất để mô tả một đối tượng cụ thể, làm cơ sở để tham chiếu, đồng bộ giữa các cơ sở dữ liệu hoặc các tập dữ liệu khác nhau. Đây là tập hợp các thông tin cốt lõi, ổn định và có tính tham chiếu cao như dữ liệu về con người, tổ chức, địa điểm,... được sử dụng xuyên suốt trong các hệ thống thông tin và cơ sở dữ liệu chuyên ngành. Việc quản lý và duy trì dữ liệu chủ một cách tập trung, thống nhất giúp bảo đảm tính chính xác, đồng bộ và khả năng tích hợp giữa các hệ thống, đồng thời giảm thiểu trùng lặp, sai lệch thông tin.

- Dữ liệu giao dịch (transaction data) là loại dữ liệu phát sinh trong quá trình thực hiện các nghiệp vụ, giao dịch hành chính hoặc cung cấp dịch vụ công của chủ thể dữ liệu (được biểu diễn bởi dữ liệu chủ). Dữ liệu này phản ánh hoạt động của chủ thể dữ liệu, ví dụ như thông tin cụ thể về thời gian, địa điểm, đối tượng, nội dung và kết quả của từng giao dịch, đóng vai trò quan trọng trong việc theo dõi, kiểm tra, đánh giá hiệu quả hoạt động và hỗ trợ ra quyết định.

- Dữ liệu lớn (big data) là tập hợp các dữ liệu có dung lượng lớn, đa dạng và được tạo ra với tốc độ cao từ nhiều nguồn khác nhau, bên cạnh các dữ liệu nội bộ, dữ liệu lớn bao gồm dữ liệu thu thập từ các nguồn dữ liệu bên ngoài mỗi cơ quan, tổ chức như mạng xã hội, thiết bị Internet vạn vật (Internet of Things - IoT), mạng Internet.

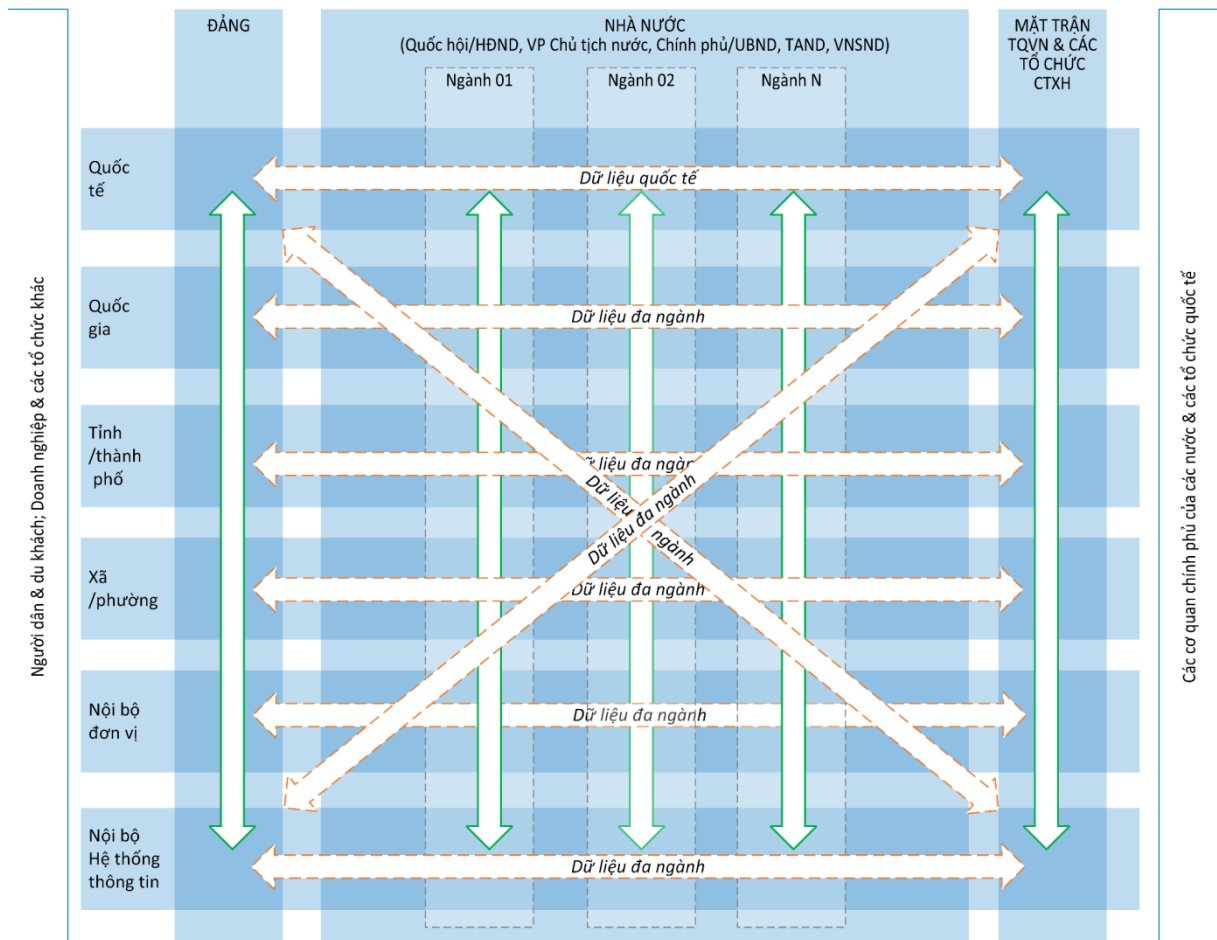
- Dữ liệu tổng hợp (aggregate data) là kết quả của quá trình xử lý, tập hợp và thống kê từ các dữ liệu chi tiết tạo ra thông tin có tính khái quát cao, phục vụ phân tích và ra quyết định.

- Dữ liệu kết quả phân tích (hay dữ liệu suy diễn - inferred data) là kết quả áp dụng các mô hình học máy (ML-machine learning), trí tuệ nhân tạo (AI – artificial intelligence) để phân tích nâng cao các lớp dữ liệu tổng hợp, dữ liệu lớn và các lớp dữ liệu gốc.

- Dữ liệu mở (open data) là tập hợp dữ liệu được các 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.

b) Phạm vi dữ liệu dùng chung

Mô hình dưới đây minh họa việc phân loại phạm vi dữ liệu, làm cơ sở quy hoạch quản trị, quản lý dữ liệu dùng chung ở các cấp và theo các khối cơ quan của bộ, ngành, địa phương.



Hình 13: Phân loại phạm vi dữ liệu

Phạm vi dữ liệu được quản lý và khai thác tại các cơ quan ở các cấp theo chiều dọc từ bên trong nội bộ hệ thống thông tin đến nội bộ một cơ quan/đơn vị ra bên ngoài, từ các cấp hành chính xã/phường tới cấp quốc gia và quốc tế. Dữ liệu cũng được quản lý và khai thác theo chiều ngang: dữ liệu được quản lý và khai thác tại các cơ quan khác nhau trong hệ thống chính trị: các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan trung ương của tổ chức chính trị, Viện Kiểm sát nhân dân tối cao, Tòa án nhân dân tối cao, Kiểm toán Nhà nước, Văn phòng Chủ tịch nước, Văn phòng Quốc hội, Văn phòng Trung ương Đảng, cơ quan trung ương thuộc Mặt trận Tổ quốc Việt Nam.

Dữ liệu được phân loại theo phạm vi sử dụng:

- Dữ liệu nội bộ là dữ liệu chỉ được sử dụng, dùng chung trong phạm vi một hệ thống thông tin, một cơ quan, đơn vị.

- Dữ liệu dùng chung địa phương là dữ liệu được sử dụng, dùng chung trong phạm vi cấp xã/phường hoặc cấp tỉnh/thành phố.

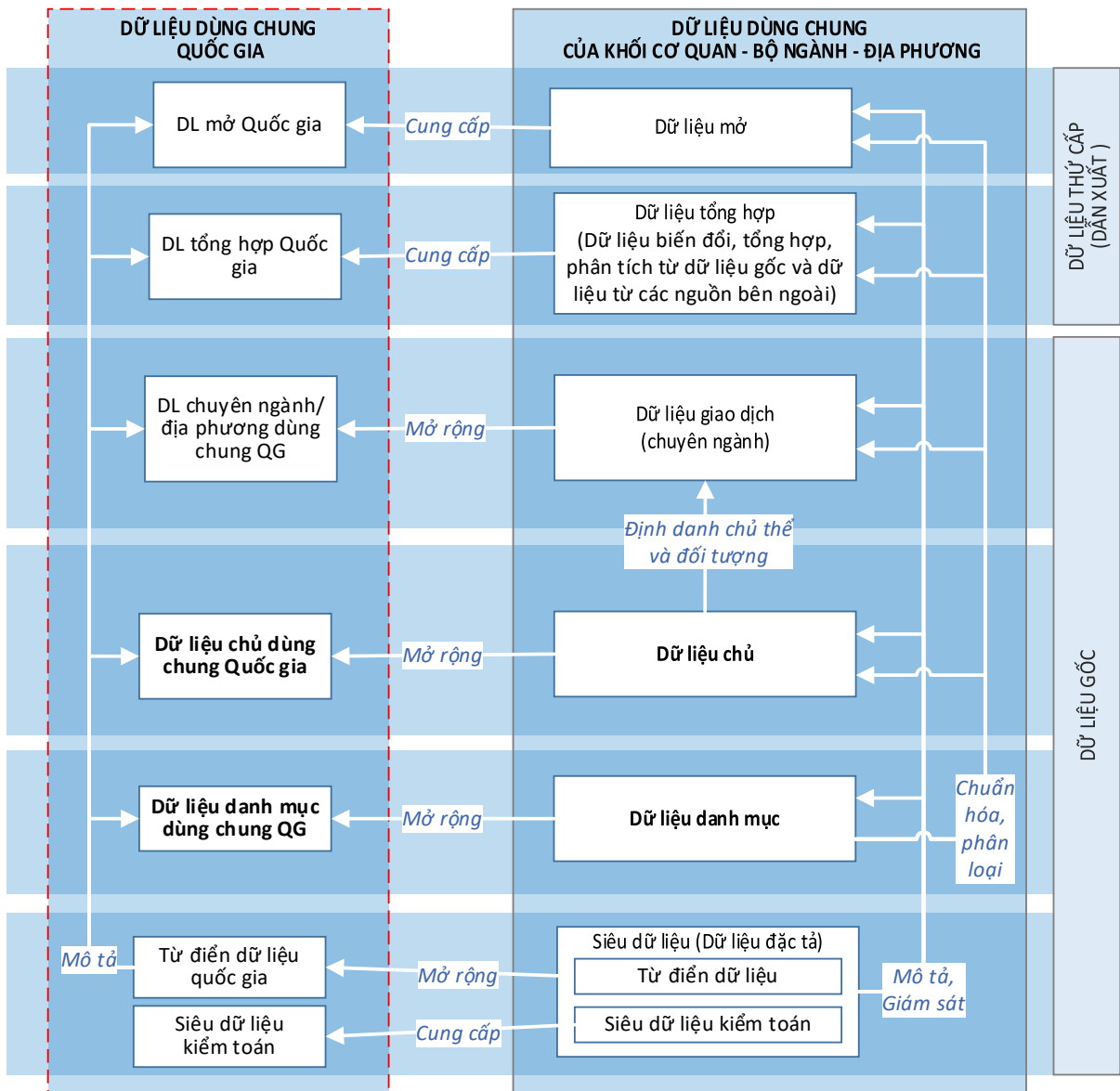
- Dữ liệu dùng chung đa ngành là dữ liệu được sử dụng, dùng chung bởi nhiều ngành, có thể cùng hoặc khác cấp quản lý hành chính.

- Dữ liệu dùng chung quốc gia là dữ liệu được sử dụng, dùng chung trong phạm vi toàn quốc, phục vụ nhiều địa phương, nhiều ngành, nhiều cơ quan, đơn vị.

- Dữ liệu dùng chung quốc tế là dữ liệu được trao đổi giữa các nước với Việt Nam theo tiêu chuẩn, quy chuẩn kỹ thuật được cơ quan có thẩm quyền của Việt Nam chấp thuận.

c) Mô hình đặc tả dữ liệu

Mô hình đặc tả (metamodel) các lớp dữ liệu là mô hình đặc tả mức khái niệm các nhóm dữ liệu tổng quát cần được xác định trong mỗi khối cơ quan, làm cơ sở để xác định các nhóm dữ liệu chia sẻ, dùng chung cũng như hỗ trợ xây dựng kiến trúc logic dữ liệu, các luồng dữ liệu và các hệ thống/nền tảng quản trị, quản lý dữ liệu.



Hình 14: Mô hình đặc tả các lớp dữ liệu

Trong mô hình đặc tả các lớp dữ liệu, các lớp dữ liệu gốc dùng chung tại các cơ quan, bộ, ngành, địa phương bổ sung, mở rộng các lớp dữ liệu dùng chung quốc gia.

d) Mô hình dữ liệu tham chiếu

Mô hình dữ liệu tham chiếu thể hiện các miền dữ liệu được khái quát hóa từ nghiệp vụ, các hoạt động và các lớp dữ liệu.

Mô hình dữ liệu tham chiếu của một khối cơ quan, bộ ngành, địa phương được thể hiện trong hình sau. Trong đó các miền dữ liệu thành phần mang tính chất tham khảo. Mỗi khối cơ quan, bộ ngành, địa phương cần xác định cụ thể các miền dữ liệu và danh mục dữ liệu thành phần ở mỗi lớp:

DỮ LIỆU (DÙNG CHUNG) QUỐC GIA TẠI TTDLQG		DỮ LIỆU CỦA KHỐI CƠ QUAN - BỘ NGÀNH - ĐỊA PHƯƠNG						
		Dữ liệu phục vụ người dân, doanh nghiệp (TTHC, DVC)	Dữ liệu phục vụ hoạt động nghiệp vụ chuyên ngành / địa phương	Dữ liệu phục vụ công tác quản lý, chỉ đạo điều hành, xây dựng chính sách	Dữ liệu phục vụ quản trị nguồn lực nội bộ	Dữ liệu phục vụ an ninh, quốc phòng, đối ngoại, trật tự an toàn xã hội	Dữ liệu về thể chế, chính sách	
		DL mở về TTHC	DL mở (chi tiết)	DL mở các lĩnh vực (tổng hợp)	DL mở về nguồn lực	DL mở ANQP, ĐN, TTATXH	DL mở về thể chế, CS	Dữ liệu mở
DL tổng hợp quốc gia khác DL phục vụ CT Chỉ đạo ĐH của CP, TTg CP		Dữ liệu quá trình DVC khác DL tổng hợp về DVC, TTHC	Dữ liệu báo cáo tác nghiệp, báo cáo nhanh chuyên ngành/ địa phương	Dữ liệu tổng hợp về chỉ tiêu báo cáo, chỉ tiêu thống kê, chỉ số CĐĐH...	Dữ liệu báo cáo quản trị	DL tổng hợp về ANQP, Đối ngoại, TTATXH	Dữ liệu tổng hợp về thể chế, chính sách	Dữ liệu tổng hợp (biến đổi, tổng hợp phân tích)
DL phục vụ KTS, XHS... DL về điều phối & quản trị dữ liệu Dữ liệu tiếp nhận & trả KQ TTHC		Dữ liệu quá trình DVC khác Dữ liệu xử lý TTHC	Dữ liệu hoạt động nghiệp vụ chuyên ngành /địa phương	Dữ liệu hoạt động chỉ đạo, điều hành	DL hoạt động quản trị nguồn lực	DL hoạt động về ANQP, Đối ngoại, TTATXH	Dữ liệu hoạt động về thể chế, chính sách	Dữ liệu giao dịch (nghiệp vụ chuyên ngành)
DLQG khác (TTHC...) DLQG về tổ chức DLQG về dân cư, căn cước		Quy định về xử lý TTHC Dữ liệu doanh nghiệp Dữ liệu người dân	DL chủ chuyên ngành/địa phương	Dữ liệu chủ về các nguồn lực phục vụ chỉ đạo, điều hành	DL tài sản DL cơ cấu tổ chức DL về nhân sự	DL sự kiện QP, ANTT DL trang thiết bị DL địa lý, hạ tầng trọng yếu	Văn bản pháp lý, chính sách	Dữ liệu Chủ
Dữ liệu danh mục dùng chung QG		Dữ liệu DMDC liên quan DVC, TTHC	Dữ liệu DMDC chuyên ngành/ địa phương	DMDC về chỉ tiêu báo cáo, chỉ tiêu thống kê, chỉ số CĐĐH...	DMDC về quản trị nội bộ	DMDC về ANQP, ĐN, Trật tự ATXH	DMDC về thể chế, chính sách	Dữ liệu danh mục
Từ điển dữ liệu quốc gia Siêu dữ liệu (SDL) kiểm toán dữ liệu		Từ điển DL về người dân, DN, TTHC SDL kiểm toán DL TTHC...	Từ điển DL chuyên ngành/ địa phương SDL kiểm toán DL nghiệp vụ	Từ điển DL CĐĐH SDL kiểm toán DL QL, CĐĐH...	Từ điển dữ liệu quản trị nội bộ SDL kiểm toán DL QTNB	Từ điển DL An ninh QP, ĐN, TTATXH SDL kiểm toán DL ANQP...	Từ điển DL về thể chế, chính sách SDL kiểm toán DL thể chế	Siêu dữ liệu (Dữ liệu đặc tả)

Hình 15: Mô hình dữ liệu tham chiếu của một khối cơ quan, bộ ngành, địa phương

4. Mô hình tham chiếu về đặc tả danh mục dữ liệu

Danh mục dữ liệu mô tả cụ thể các tài nguyên dữ liệu của mỗi miền dữ liệu trong mô hình dữ liệu tham chiếu của khối cơ quan, bộ ngành, địa phương. Danh mục dữ liệu cần cung cấp các thông tin đặc tả về các tài nguyên dữ liệu: thông tin phân loại và các thông tin về ngữ cảnh và ngữ nghĩa của dữ liệu.

a) Phân loại dữ liệu

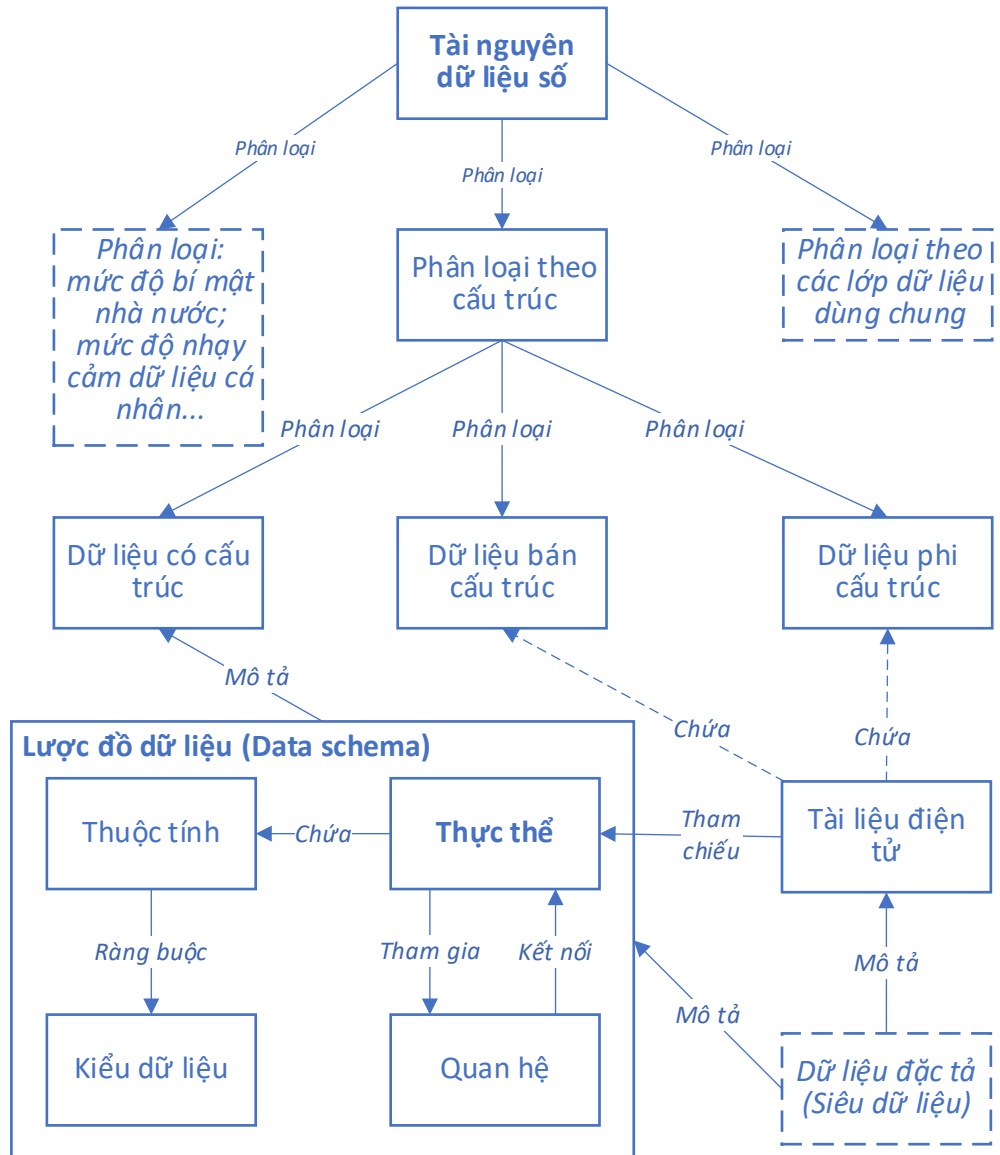
Khi xây dựng, cập nhật, chuẩn hóa dữ liệu, các cơ quan, đơn vị thực hiện phân loại dữ liệu theo tiêu chí dưới đây:

STT	Tiêu chí phân loại	Phân loại
1	Phân loại theo lớp dữ liệu gốc	
1.1		Dữ liệu đặc tả/Siêu dữ liệu
1.2		Dữ liệu danh mục
1.3		Dữ liệu chủ
1.4		Dữ liệu giao dịch chuyên ngành
1.5		Dữ liệu khác
2	Phân loại dữ liệu dẫn xuất, dữ liệu thứ cấp	
2.1		Dữ liệu thu thập nguồn mở
2.2		Dữ liệu tổng hợp (từ dữ liệu gốc và từ các cơ quan bên ngoài, các nguồn mở)
2.3		Dữ liệu kết quả phân tích nâng cao
3	Phân loại theo tính chất chia sẻ	
3.1		Dữ liệu dùng riêng
3.2		Dữ liệu dùng chung
3.3		Dữ liệu mở
3.4		Dữ liệu dịch vụ có thu phí
4	Phân loại theo tính chất quan trọng của dữ liệu	

STT	Tiêu chí phân loại	Phân loại
4.1		Dữ liệu chưa phân loại mức độ quan trọng
4.2		Dữ liệu quan trọng
4.3		Dữ liệu cốt lõi
5	Phân loại theo quy định về Bảo vệ dữ liệu cá nhân	
5.1		Dữ liệu cá nhân cơ bản
5.2		Dữ liệu cá nhân nhạy cảm
5.3		Dữ liệu cá nhân khác
6	Phân loại theo quy định về Bảo vệ bí mật nhà nước	
6.1		Không mật
6.2		Mật
6.3		Tối mật
6.4		Tuyệt mật
7	Bảo vệ bí mật kinh doanh	
7.1		Dữ liệu bí mật kinh doanh
7.2		Dữ liệu không phải bí mật kinh doanh
8	Phân loại theo cấu trúc dữ liệu	
8.1		Dữ liệu có cấu trúc
8.2		Dữ liệu bán cấu trúc
8.3		Dữ liệu phi cấu trúc

b) Đặc tả các tài nguyên dữ liệu

Tài nguyên dữ liệu là các dữ liệu có cấu trúc hoặc phi cấu trúc có khả năng tái sử dụng. Tài nguyên dữ liệu cần được mô tả với các thông tin trong mô hình đặc tả dưới đây:



Chú giải:



Thành phần bắt buộc



Thành phần không bắt buộc



Được mô tả cụ thể trong phần khác

Hình 16: Mô hình đặc tả tài nguyên dữ liệu

Theo mô hình trên, mỗi tài nguyên dữ liệu được đặc tả theo các bảng sau:

(1) Dữ liệu có cấu trúc

Thành phần	Thuộc tính	Mô tả	Ví dụ
Thực thể	<i>Định danh</i>	Mô tả định danh	“001309F”
	<i>Tên</i>	Tên của thực thể	“Con người”
	<i>Mô tả</i>	Mô tả về thực thể	
Kiểu dữ liệu	<i>Tên</i>	Tên kiểu dữ liệu	“string”
	<i>Mô tả</i>	Mô tả kiểu dữ liệu	
Thuộc tính	<i>Tên</i>	Tên của thuộc tính	“Ngày sinh”
	<i>Mô tả</i>	Mô tả thuộc tính	
Quan hệ	<i>Tên</i>	Tên của quan hệ	“Làm việc cho”
	<i>Từ</i>	Tên thực thể chủ - là thực thể xuất phát của quan hệ.	“Nhân viên”
	<i>Đến</i>	Tên thực thể phụ thuộc - là thực thể đích đến của quan hệ.	“Công ty”

(2) Dữ liệu phi cấu trúc:

Thành phần	Thuộc tính	Mô tả	Ví dụ
Tài nguyên dữ liệu phi cấu trúc	<i>Định danh</i>	Định danh của tài nguyên dữ liệu phi cấu trúc	
	<i>Tên</i>	Tên của tài nguyên dữ liệu	
	<i>Ngày</i>	Ngày khởi tạo	
	<i>Định dạng</i>	Định dạng của dữ liệu phi cấu trúc.	“text/plain”
	<i>Mô tả</i>	Mô tả nội dung của tài nguyên dữ liệu.	
	<i>Nguồn</i>	Tham chiếu tới nguồn giúp tạo ra nội dung.	
	<i>Chủ đề</i>	Chủ đề của nội dung	
	<i>Đơn vị phát hành</i>	Đơn vị công bố nội dung của tài nguyên	
	<i>Đơn vị tham gia</i>	Đơn vị có tham gia đóng góp vào nội dung.	
	<i>Ngôn ngữ</i>	Ngôn ngữ của nội dung.	
	<i>Quan hệ</i>	Tham chiếu tới tài nguyên liên quan.	
	<i>Phạm vi</i>	Phạm vi địa lý của nội dung	
	Phạm vi địa lý	Thông tin địa lý của nội dung liên quan	

Ghi chú: Để mô tả các tài nguyên phi cấu trúc, các cơ quan, tổ chức có thể tham khảo tiêu chuẩn Dublin Core Metadata Initiative (DCMI), phiên bản 1.1 hoặc Tiêu chuẩn TCVN 7980-1:2024.

5. Mô hình tham chiếu luồng chia sẻ, điều phối dữ liệu

a) Các mô hình khái niệm về chia sẻ, điều phối dữ liệu

- Hai mô hình nghiệp vụ cơ bản về chia sẻ dữ liệu xác định rõ vai trò và trách nhiệm của các bên tham gia. Khung kiến trúc dữ liệu cũng hướng dẫn lựa chọn giữa hai mô hình chia sẻ dữ liệu cơ bản: chia sẻ dữ liệu theo yêu cầu và chia sẻ dữ liệu chủ động, cụ thể:

+ Chia sẻ dữ liệu theo yêu cầu là phương thức trong đó việc chia sẻ dữ liệu bắt đầu bằng việc bên sử dụng dữ liệu yêu cầu chủ quản dữ liệu (hay bên kiểm soát dữ liệu) cho phép truy cập vào dữ liệu. Do đó, người sử dụng dữ liệu là bên chủ động. Chủ quản dữ liệu - người chịu trách nhiệm kiểm soát quyền truy cập vào dữ liệu - đánh giá xem việc chia sẻ có hợp pháp hay không theo cơ sở pháp lý của việc truy cập đó. Chủ quản dữ liệu cần bảo đảm kiểm soát truy cập chính xác và ghi nhật ký truy cập và sử dụng dữ liệu nếu cần thiết.

+ Chia sẻ dữ liệu chủ động là phương thức trong đó bên chủ quản dữ liệu (hay bên kiểm soát dữ liệu) chủ động chia sẻ dữ liệu với bên sử dụng. Theo mô hình này, bên chủ quản dữ liệu (bên kiểm soát dữ liệu) đóng gói dữ liệu dưới dạng thông điệp, gán địa chỉ bên nhận dữ liệu và gửi thông điệp tới bên sử dụng. Mô hình chia sẻ dữ liệu chủ động thường dựa trên các sự kiện: khi có một sự kiện liên quan tới chủ thể dữ liệu phát sinh dữ liệu mới về chủ thể, chủ quản dữ liệu sẽ gửi dữ liệu tới bên sử dụng.

- Sáu mô hình tham chiếu logic về chia sẻ dữ liệu được thể hiện trong bảng dưới đây:

Mô hình nghiệp vụ	Mô hình tham chiếu	Mô tả
Chia sẻ dữ liệu theo yêu cầu	Truy cập trực tiếp	Dữ liệu được chia sẻ khi bên sử dụng dữ liệu gửi yêu cầu truy vấn dữ liệu trực tiếp đến hệ thống thông tin của bên chủ quản dữ liệu.

Mô hình nghiệp vụ	Mô hình tham chiếu	Mô tả
	Phân phối dữ liệu	Bên chủ quản dữ liệu đã ủy quyền chuyển dữ liệu cho <i>bên trung gian dữ liệu</i> (<i>bên phân phối dữ liệu</i>) để phục vụ các yêu cầu của bên sử dụng dữ liệu. Bên chủ quản chịu trách nhiệm về danh mục và chất lượng dữ liệu. Bên phân phối (đóng vai trò bên xử lý dữ liệu) chịu trách nhiệm cung cấp dữ liệu tới bên sử dụng dữ liệu. Mỗi khi có dữ liệu mới, bên chủ quản dữ liệu có trách nhiệm cập nhật <i>bản sao dữ liệu</i> tại bên phân phối dữ liệu.
	Nền tảng dữ liệu dùng chung	Mô hình này sử dụng một nền tảng chung để phân phối dữ liệu và cho phép chủ quản dữ liệu và người dùng dữ liệu hoạt động trên cùng một nền tảng, các bộ dữ liệu của mỗi bên chủ quản dữ liệu tập trung về mô hình vật lý nhưng tách biệt về mô hình logic.
Chia sẻ dữ liệu chủ động	Truyền/nhận dữ liệu trực tiếp	Theo mô hình này, dữ liệu được chuyển dưới hình thức các thông điệp từ bên gửi (chủ quản dữ liệu) đến bên nhận (bên sử dụng dữ liệu) thông qua kết nối điểm - điểm.
	Truyền/nhận dữ liệu qua trung gian	Mô hình này sử dụng một hệ thống trung gian để gửi các thông điệp/gói dữ liệu từ chủ quản dữ liệu tới bên nhận (bên sử dụng dữ liệu). Mô hình này giúp tách rời người gửi (chủ quản dữ liệu) và người nhận (bên sử dụng dữ liệu) về hạ tầng kỹ thuật và cung cấp các dịch vụ gia tăng như thông báo và địa chỉ...
	Cộng đồng truyền/nhận dữ liệu	Mô hình này kết nối các hệ thống truyền nhận dữ liệu với nhau trong một cộng đồng dịch vụ (được cung cấp trong hệ sinh thái các nhà cung cấp dịch vụ) cung cấp trao đổi thông điệp chứa dữ liệu. Mỗi nhà cung cấp dịch vụ có thể phục vụ một số người gửi/người nhận.

b) Các mô hình tham chiếu về chia sẻ, điều phối dữ liệu dùng chung

- Chia sẻ, điều phối dữ liệu từ điển, dữ liệu danh mục dùng chung: với các dữ liệu dùng chung có tính chất là dữ liệu nền tảng, ổn định, ít thay đổi hoặc tần suất thay đổi thấp nên được quản lý tập trung ở mức quốc gia theo mô hình “Nền tảng dữ liệu dùng chung” tại Trung tâm dữ liệu quốc gia và chia sẻ dữ liệu theo yêu cầu cho các hệ thống sử dụng dữ liệu của các cơ quan thông qua Hệ thống từ điển dữ liệu dùng chung và Hệ thống quản lý danh mục dùng chung quốc gia.

- Chia sẻ, điều phối dữ liệu chủ: dữ liệu chủ có tính pháp lý và gắn với hệ thống nghiệp vụ chuyên ngành và các thủ tục hành chính. Do vậy, việc tạo lập, cập nhật dữ liệu được thực hiện trên các hệ thống thông tin và cơ sở dữ liệu chuyên ngành. Để bảo đảm tính nhất quán và đồng bộ của dữ liệu chủ trong Cơ sở dữ liệu tổng hợp quốc gia, dữ liệu chủ cần được cập nhật ngay vào Cơ sở dữ liệu tổng hợp quốc gia như một bước trong quy trình nghiệp vụ chuyên ngành hoặc quy trình giải quyết thủ tục hành chính. Việc này bảo đảm tính duy nhất và tính tức thời của các dữ liệu chủ đã được quy định bởi Luật Dữ liệu. Việc chia sẻ dữ liệu chủ sẽ kết hợp hai mô hình:

+ Các cơ quan chủ quản dữ liệu chủ đồng bộ dữ liệu chủ cho Cơ sở dữ liệu tổng hợp quốc gia theo phương thức chủ động: tích hợp vào quy trình nghiệp vụ/quy trình giải quyết TTHC có thay đổi, tạo lập dữ liệu chủ, để thực hiện cập nhật vào Cơ sở dữ liệu tổng hợp quốc gia ngay trong quy trình thông qua Nền tảng chia sẻ, điều phối dữ liệu;

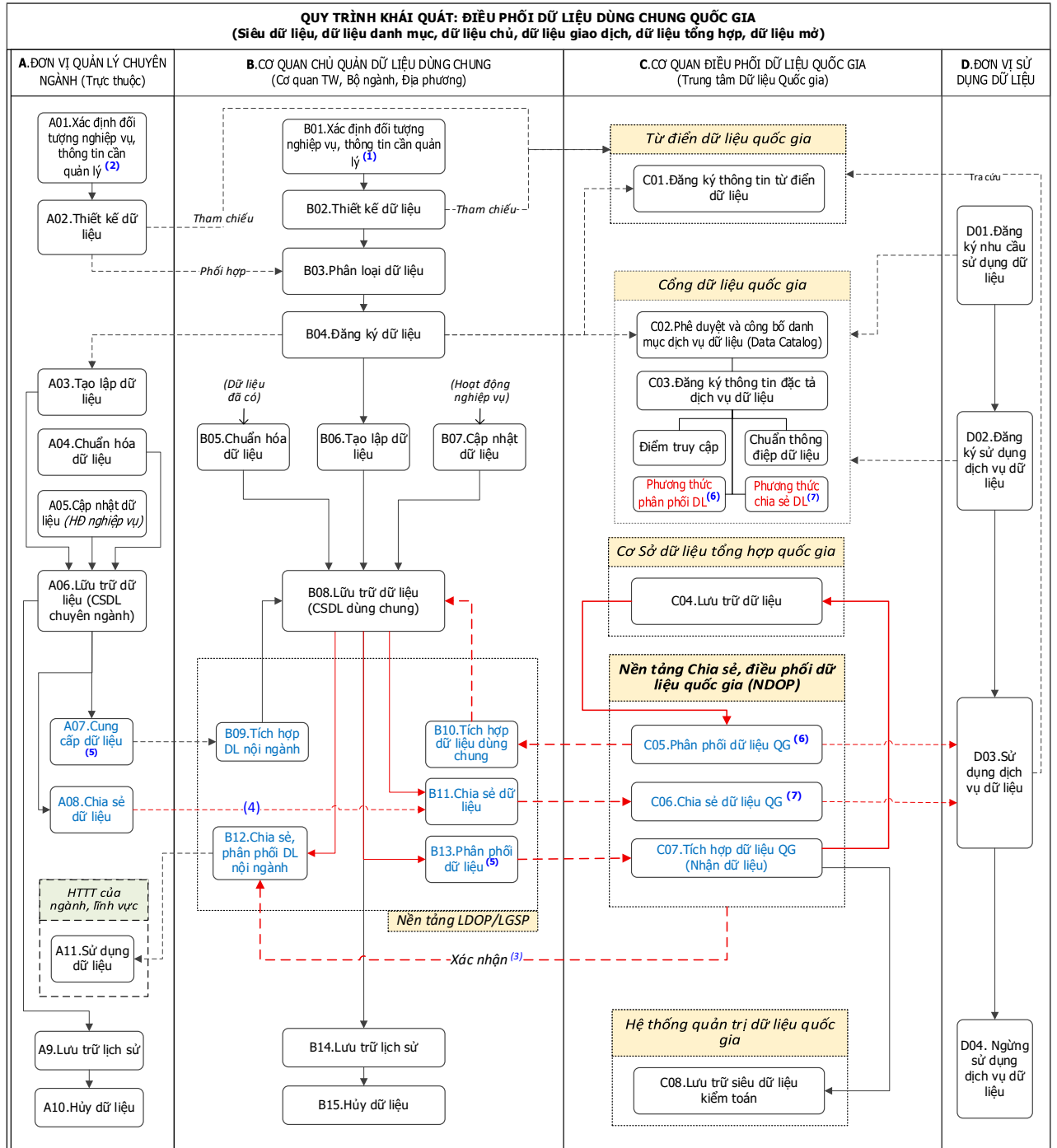
+ Trung tâm dữ liệu quốc gia chia sẻ dữ liệu chủ từ Cơ sở dữ liệu tổng hợp quốc gia cho các cơ quan theo yêu cầu thông qua Nền tảng chia sẻ, điều phối dữ liệu.

- Chia sẻ, điều phối dữ liệu giao dịch/dữ liệu nghiệp vụ chuyên ngành, dữ liệu lớn, dữ liệu tổng hợp và dữ liệu kết quả phân tích: dữ liệu chuyên ngành, dữ liệu tổng hợp và dữ liệu kết quả phân tích được chia sẻ, đồng bộ vào Cơ sở dữ liệu tổng hợp quốc gia để phục vụ công tác chỉ đạo điều hành.

Do vậy, dữ liệu chuyên ngành và dữ liệu tổng hợp phải được đồng bộ theo danh mục đáp ứng các nhiệm vụ của các cơ quan Trung ương theo từng giai đoạn, theo biến động của dữ liệu và yêu cầu chỉ đạo điều hành.

Việc chia sẻ dữ liệu giao dịch/dữ liệu nghiệp vụ chuyên ngành, dữ liệu tổng hợp và dữ liệu lớn sẽ được thực hiện theo phương thức chia sẻ dữ liệu chủ động: các cơ quan chủ quản dữ liệu truyền dữ liệu có thay đổi, cập nhật cho Cơ sở dữ liệu tổng hợp quốc gia: theo thời gian thực - đối với các dữ liệu có phục vụ dịch vụ công; chỉ đạo điều hành và theo định kỳ hàng ngày - đối với các dữ liệu phục vụ thống kê, phân tích.

c) Quy trình tham chiếu về chia sẻ, điều phối dữ liệu



Hình 17: Quy trình khái quát Điều phối dữ liệu dùng chung quốc gia

(1) Quy trình tại đơn vị quản lý chuyên ngành

Bước	Thành phần	Mô tả
A.01	Xác định đối tượng nghiệp vụ, thông tin cần quản lý	Xác định rõ các loại dữ liệu cần thu thập và quản lý
A.02	Thiết kế dữ liệu	Xây dựng cấu trúc dữ liệu, xác định các thực thể, mô hình hóa quan hệ, kiểu dữ liệu và định nghĩa rõ các trường thông tin cần thiết.
A.03	Tạo lập dữ liệu	Tạo, thu thập hoặc tổng hợp dữ liệu từ các hệ thống, biểu mẫu, hoặc nguồn tài liệu chuyên ngành khác.
A.04	Chuẩn hóa dữ liệu	Tổ chức và cấu trúc dữ liệu trong cơ sở dữ liệu để loại bỏ dữ liệu trùng lặp, bảo đảm tính nhất quán, tính toàn vẹn và giảm thiểu sai sót của dữ liệu
A.05	Cập nhật dữ liệu (hoạt động nghiệp vụ)	Cập nhật dữ liệu theo quy trình nghiệp vụ định kỳ hoặc theo nhu cầu phát sinh như cập nhật thông tin dân cư sau đăng ký, bổ sung thông tin,...
A.06	Lưu trữ dữ liệu (cơ sở dữ liệu chuyên ngành)	Sau khi dữ liệu đã được tạo lập, chuẩn hóa và cập nhật từ các quy trình nghiệp vụ, đơn vị quản lý chuyên ngành sẽ thực hiện lưu trữ dữ liệu vào Cơ sở dữ liệu chuyên ngành
A.07	Cung cấp dữ liệu	Cung cấp dữ liệu cho hệ thống dùng chung (cơ sở dữ liệu dùng chung), chủ động đẩy dữ liệu
A.08	Chia sẻ dữ liệu	Cho phép các đơn vị khác truy vấn hoặc chia sẻ dữ liệu theo đúng quyền, phạm vi được cho phép.
A.09	Lưu trữ lịch sử	Sau khi dữ liệu đã được sử dụng hoặc cập nhật, thực hiện lưu trữ dữ liệu lịch sử bảo đảm khả năng truy xuất, tra cứu, đối soát hoặc phục vụ kiểm tra, rà soát về sau.
A.10	Hủy dữ liệu	Hủy dữ liệu khi hết hạn lưu trữ hoặc không còn giá trị cần lưu trữ (theo quy định), không cho phép khôi phục và cần lưu lại lịch sử hủy dữ liệu.
A.11	Sử dụng dữ liệu	Dữ liệu được dùng phục vụ để tra cứu, phân tích, quản lý, báo cáo theo nghiệp vụ từng chuyên ngành, hỗ trợ ra quyết định và các chính sách.

(2) Quy trình tại cơ quan chủ quản quản lý dữ liệu

Bước	Thành phần	Mô tả
B.01	Xác định đối tượng nghiệp vụ, thông tin cần quản lý	Phân tích các nghiệp vụ quản lý thuộc phạm vi của bộ, ngành, địa phương để xác định đâu là dữ liệu có khả năng sử dụng chung hoặc cần tổng hợp. Xác định phạm vi dữ liệu dùng chung từ nhiều đơn vị chuyên ngành khác nhau
B.02	Thiết kế dữ liệu	Xây dựng mô hình dữ liệu tổng hợp, liên ngành. Thiết kế kiến trúc, định dạng chuẩn, danh mục mã để chuẩn hóa dữ liệu đầu vào từ nhiều nguồn. Bảo đảm dữ liệu đáp ứng các chuẩn tích hợp quốc gia.
B.03	Phân loại dữ liệu	Phân loại dữ liệu theo các nhóm: Mức độ bảo mật: Dữ liệu thường, dữ liệu mật Mức độ chia sẻ: Dữ liệu mở, dữ liệu nội bộ
B.04	Đăng ký dữ liệu	Thực hiện đăng ký các tập dữ liệu, danh mục dữ liệu dùng chung (Kết quả của B02 và B03) với cơ quan điều phối Dữ liệu quốc gia Đầy đủ đặc tả kỹ thuật, mô tả nghiệp vụ và các tài liệu liên quan
B.05	Chuẩn hóa dữ liệu	Tổ chức và cấu trúc dữ liệu trong cơ sở dữ liệu để loại bỏ dữ liệu trùng lặp, bảo đảm tính nhất quán, tính toàn vẹn và giảm thiểu sai sót của dữ liệu
B.06	Tạo lập dữ liệu	Tổ chức xây dựng dữ liệu tổng hợp, tạo lập trực tiếp, nhập biểu mẫu, thu thập bằng phần mềm, hoặc tiếp nhận từ hệ thống khác
B.07	Cập nhật dữ liệu	Cập nhật dữ liệu theo quy trình nghiệp vụ định kỳ hoặc theo nhu cầu phát sinh như cập nhật thông tin dân cư sau đăng ký, bổ sung thông tin,...
B.08	Lưu trữ dữ liệu (cơ sở dữ liệu dùng chung)	Lưu trữ toàn bộ dữ liệu đã chuẩn hóa vào cơ sở dữ liệu dùng chung, có khả năng phục vụ: Tra cứu, quản lý, thống kê, phân tích, phân phối, chia sẻ
B.09	Tích hợp dữ liệu ngành	Nhận dữ liệu từ nhiều cơ sở dữ liệu chuyên ngành (A07) để lưu vào cơ sở dữ liệu dùng chung
B.10	Tích hợp dữ liệu dùng chung	Nhận dữ liệu từ Cơ sở dữ liệu tổng hợp quốc gia lưu vào cơ sở dữ liệu dùng chung
B.11	Chia sẻ dữ liệu	Chia sẻ dữ liệu với NDOP (C06) để cung cấp dữ liệu ra bên ngoài theo quyền truy cập và phạm vi dữ liệu được xác định: các cơ quan khác, tổ chức, doanh nghiệp, người dân.
B.12	Chia sẻ, phân phối DL nội ngành	Chia sẻ dữ liệu với HTTT của ngành, lĩnh vực theo quyền truy cập được xác định và phạm vi dữ liệu của chuyên ngành tương ứng.

Bước	Thành phần	Mô tả
B.13	Phân phối dữ liệu	Phân phối, đẩy các dữ liệu cần thiết sang Cơ sở dữ liệu tổng hợp quốc gia
B.14	Lưu trữ lịch sử	Sau khi dữ liệu đã được sử dụng hoặc cập nhật, thực hiện lưu trữ dữ liệu lịch sử bảo đảm khả năng truy xuất, tra cứu, đối soát hoặc phục vụ kiểm tra, rà soát về sau.
B.15	Hủy dữ liệu	Hủy dữ liệu khi hết hạn lưu trữ hoặc không còn giá trị cần lưu trữ (Theo quy định), không cho phép khôi phục, lưu lại lịch sử hủy tài liệu

(3) Quy trình tại cơ quan điều phối dữ liệu quốc gia

Bước	Thành phần	Mô tả
C.01	Đăng ký thông tin từ điển dữ liệu	Thiết lập từ điển dữ liệu dùng chung bao gồm: các danh mục chuẩn, định nghĩa thuật ngữ, định dạng dữ liệu, mã, phân loại ... Mỗi cơ quan (ở Mục A và B) phải đồng bộ và tham chiếu từ điển dữ liệu này khi thiết kế, chia sẻ, tích hợp dữ liệu.
C.02	Phê duyệt và công bố danh mục dữ liệu quốc gia	Rà soát, kiểm duyệt và công bố danh mục dữ liệu dùng chung, dữ liệu mở, dữ liệu trọng yếu. Cung cấp thông tin như: tên tập dữ liệu, đơn vị sở hữu, phạm vi sử dụng, hình thức chia sẻ.
C.03	Đăng ký thông tin đặc tả dịch vụ dữ liệu	Đăng ký, chuẩn hóa thông tin kỹ thuật về dịch vụ chia sẻ dữ liệu. Các thông tin cơ bản: - Điểm truy cập - Chuẩn thông điệp dữ liệu - Phương thức phân phối dữ liệu - Phương thức chia sẻ dữ liệu
C.04	Lưu trữ dữ liệu (Cơ sở dữ liệu tổng hợp quốc gia)	Tiếp nhận, lưu trữ tập trung các loại dữ liệu dùng chung từ các đơn vị mục B (cơ sở dữ liệu dùng chung) và các dữ liệu khác
C.05	Phân phối dữ liệu QG	Phân phối dữ liệu từ Cơ sở dữ liệu tổng hợp quốc gia đến cơ sở dữ liệu dùng chung và các đơn vị sử dụng (Mục D) thông qua nền tảng NDOP. Dữ liệu được lọc, đúng theo phân quyền, mục đích sử dụng, tính chất dữ liệu, phân loại dữ liệu
C.06	Chia sẻ dữ liệu	Cho phép các đơn vị sử dụng dữ liệu (Mục D) tra cứu, kết nối, tải về dữ liệu phục vụ nhu cầu Việc chia sẻ phải có kiểm soát quyền truy cập, nhật ký sử dụng, phạm vi tính chất dữ liệu và hạn chế theo quy định bảo mật

Bước	Thành phần	Mô tả
C.07	Tích hợp dữ liệu (Nhận dữ liệu)	Nhận dữ liệu được phân phối từ cơ sở dữ liệu dùng chung để lưu trữ vào Cơ sở dữ liệu tổng hợp quốc gia phục vụ phân phối và chia sẻ (C05 và C06)
C.08	Lưu trữ siêu dữ liệu kiểm toán	Lưu trữ thông tin dữ liệu, nhật ký sử dụng, lịch sử truy cập, lịch sử chia sẻ để phục vụ kiểm tra, giám sát và bảo đảm tuân thủ

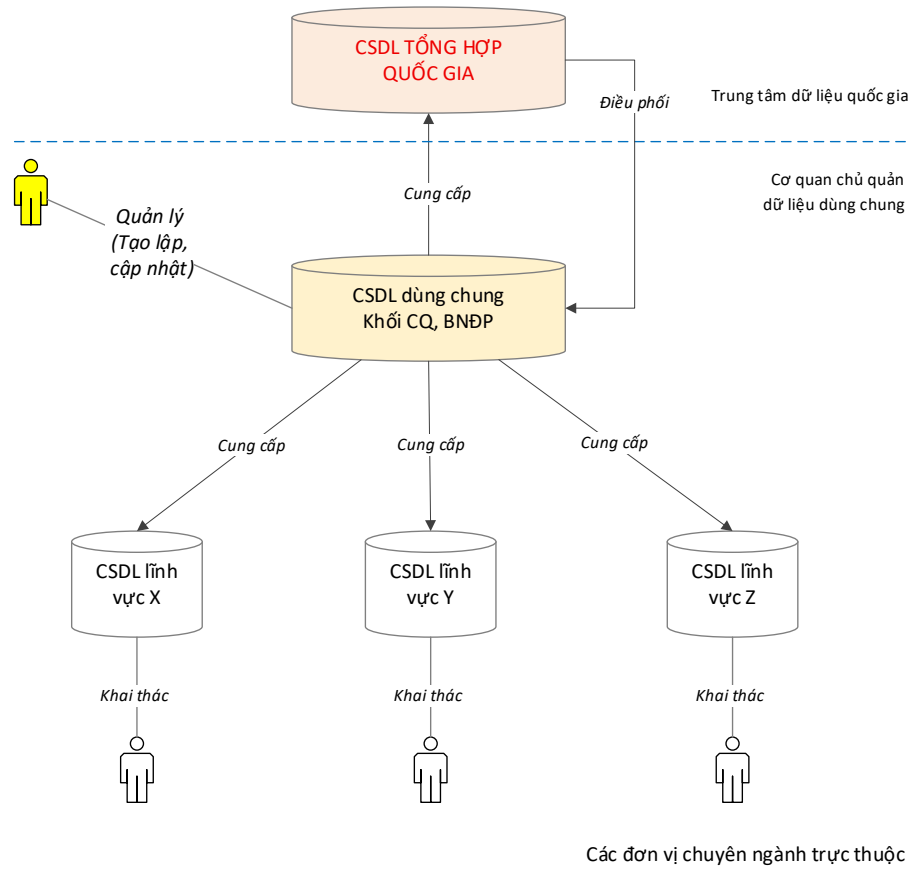
(4) Quy trình tại đơn vị sử dụng dữ liệu

Bước	Thành phần	Mô tả
D.01	Đăng ký nhu cầu sử dụng dữ liệu	Đơn vị sử dụng xác định nhu cầu sử dụng dữ liệu và đăng ký trên cổng dữ liệu quốc gia
D.02	Đăng ký sử dụng dịch vụ dữ liệu	Đơn vị sử dụng gửi đề xuất sử dụng dịch vụ qua cổng dữ liệu quốc gia: Cần cung cấp, chỉ rõ mục đích sử dụng, phạm vi khai thác và cam kết (nếu cần)
D.03	Sử dụng dịch vụ dữ liệu	Cơ quan sử dụng có thể sử dụng dịch vụ dữ liệu trên NDOP và tra cứu trên từ điển dữ liệu dùng chung
D.04	Ngừng sử dụng dịch vụ dữ liệu	Cơ quan sử dụng dịch vụ dữ liệu kết thúc sử dụng dịch vụ dữ liệu.

6. Mô hình tham chiếu về các cơ sở dữ liệu, kho dữ liệu

Các mô hình triển khai (Implementation Model) có thể áp dụng để quản lý dữ liệu dùng chung tại các khối cơ quan, bộ ngành, địa phương và đáp ứng chia sẻ, cung cấp dữ liệu cho Cơ sở dữ liệu tổng hợp quốc gia. Các cơ quan, đơn vị lựa chọn mô hình phù hợp cho từng nhóm dữ liệu dùng chung và đặc thù của các hệ thống nghiệp vụ chuyên ngành, cơ sở dữ liệu chuyên ngành.

a) Mô hình 01: Quản lý tập trung

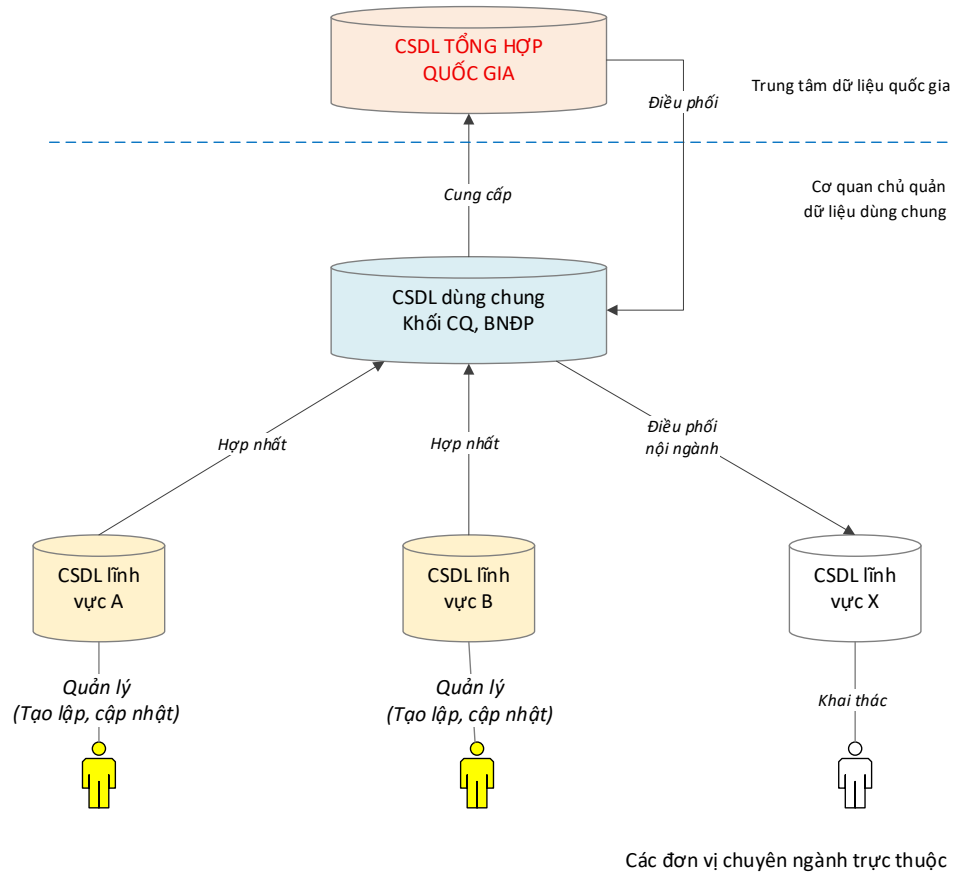


Hình 18: Mô hình tập trung về quản lý dữ liệu dùng chung

Trong mô hình tập trung, dữ liệu dùng chung (ví dụ: dữ liệu danh mục dùng chung, từ điển dữ liệu ...) được tạo lập, cập nhật, duy trì tại một cơ sở dữ liệu dùng chung của toàn bộ khối cơ quan, toàn bộ một ngành hoặc một địa phương.

Các hệ thống chuyên ngành và các cơ sở dữ liệu lĩnh vực của ngành sử dụng sử liệu cung cấp từ cơ sở dữ liệu dùng chung, không thực hiện tạo lập, cập nhật.

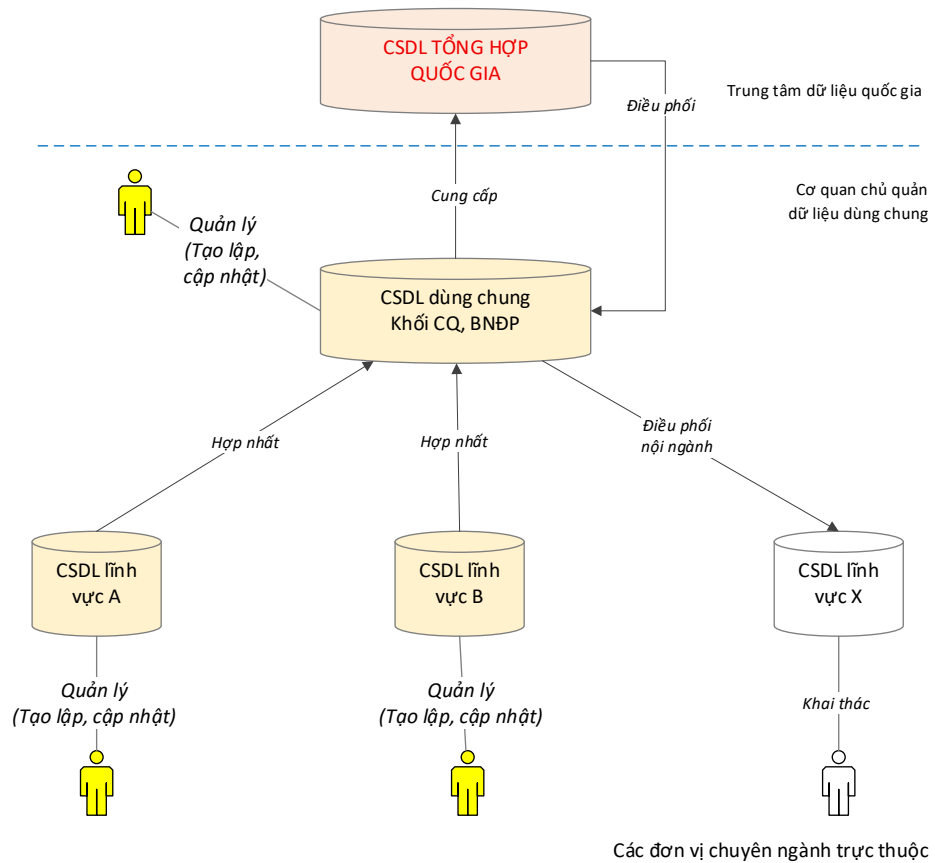
b) Mô hình 02: Quản lý phân quyền toàn bộ

**Hình 19: Mô hình phân quyền toàn bộ về quản lý dữ liệu dùng chung**

Trong mô hình quản lý phân quyền toàn bộ, mỗi nhóm dữ liệu dùng chung về một lĩnh vực được tạo lập, cập nhật, duy trì tại các cơ sở dữ liệu chuyên ngành/lĩnh vực và được hợp nhất (consolidate) vào cơ sở dữ liệu dùng chung của khối cơ quan, bộ ngành, địa phương. Sau đó, dữ liệu được cung cấp cho các hệ thống, các cơ sở dữ liệu của các lĩnh vực chuyên ngành sử dụng.

Các hệ thống và các cơ sở dữ liệu của các đơn vị chuyên ngành/lĩnh vực trực thuộc sử dụng dữ liệu dùng chung mà đơn vị không có từ cơ sở dữ liệu dùng chung.

c) Mô hình 03: Quản lý phân quyền một phần

**Hình 20: Mô hình phân quyền một phần về quản lý dữ liệu dùng chung**

Trong mô hình quản lý phân quyền một phần, một số nhóm dữ liệu dùng chung về một vài lĩnh vực được tạo lập, cập nhật, duy trì tại các hệ thống/cơ sở dữ liệu chuyên ngành/lĩnh vực và được hợp nhất (consolidate) vào cơ sở dữ liệu dùng chung của cả khối cơ quan, bộ ngành, địa phương.

Đồng thời, một số nhóm dữ liệu khác được quản lý tại hệ thống cơ sở dữ liệu dùng chung của khối cơ quan các bộ, ngành, địa phương. Sau đó, dữ liệu được cung cấp cho các hệ thống, các cơ sở dữ liệu của các lĩnh vực có nhu cầu sử dụng.

d) Khuyến nghị lựa chọn các mô hình triển khai

Bảng sau tóm tắt đặc trưng các mô hình:

Đặc trưng	Mô hình 01: Quản lý tập trung	Mô hình 02: Phân quyền quản lý toàn bộ	Mô hình 03: Phân quyền quản lý một phần
-----------	----------------------------------	---	--

Quản lý chất lượng dữ liệu	Một đơn vị chịu trách nhiệm.	Nhiều đơn vị chịu trách nhiệm: Đơn vị quản lý cơ sở dữ liệu dùng chung và các đơn vị trực thuộc ngành/lĩnh vực.	Nhiều đơn vị chịu trách nhiệm: Đơn vị quản lý cơ sở dữ liệu dùng chung và các đơn vị trực thuộc ngành/lĩnh vực.
Độ trễ dữ liệu	Thời gian thực	Có độ trễ nếu đồng bộ định kỳ.	Có độ trễ nếu đồng bộ định kỳ.

Bảng dưới đây nêu khuyến nghị về mô hình áp dụng cho từng nhóm dữ liệu dùng chung của mỗi lớp dữ liệu trong từng khối cơ quan, bộ ngành, địa phương:

Lớp dữ liệu	Mô hình 01: Quản lý tập trung	Mô hình 02: Phân quyền quản lý toàn bộ	Mô hình 03: Phân quyền quản lý một phần
Dữ liệu đặc tả (Siêu dữ liệu/từ điển dữ liệu)	Khuyến nghị	<i>Không khuyến nghị</i>	<i>Không khuyến nghị</i>
Dữ liệu danh mục dùng chung	Khuyến nghị	<i>Không khuyến nghị</i>	<i>Không khuyến nghị</i>
Dữ liệu chủ	Khuyến nghị: cho trường hợp làm chủ thiết kế về cơ sở dữ liệu	Khuyến nghị cho trường hợp không làm chủ được về chức năng và cơ sở dữ liệu chủ. Ví dụ: phần mềm thương mại	Khuyến nghị: cho trường hợp hỗn hợp mô hình 01 và mô hình 02. Ví dụ: Bao gồm cả phần mềm thương mại và phần mềm nội bộ.
Dữ liệu giao dịch/chuyên ngành	<i>Không khuyến nghị</i>	Khuyến nghị: Phần mềm nghiệp vụ chuyên ngành quản lý.	<i>Không khuyến nghị.</i>
Dữ liệu tổng hợp	<i>Không khuyến nghị</i>	Khuyến nghị: Phần mềm nghiệp vụ chuyên ngành quản lý.	<i>Không khuyến nghị</i>
Dữ liệu mở	<i>Không khuyến nghị</i>	Khuyến nghị: Các đơn vị trong lĩnh vực tạo lập	<i>Không khuyến nghị</i>

Các mô hình triển khai (Implementation Model) có thể áp dụng để quản lý dữ liệu dùng chung tại các khối cơ quan, bộ ngành, địa phương và đáp ứng chia sẻ, cung cấp dữ liệu cho cơ sở dữ liệu tổng hợp quốc gia. Các cơ quan, đơn vị lựa chọn mô hình phù hợp cho từng nhóm dữ liệu dùng chung và đặc thù của các hệ thống nghiệp vụ chuyên ngành, cơ sở dữ liệu chuyên ngành.

7. Mô hình tham chiếu về an toàn dữ liệu



Hình 21: Kiến trúc an toàn dữ liệu

a) Kiến trúc an toàn dữ liệu là hợp phần không thể tách rời của Khung Kiến trúc dữ liệu quốc gia, có vai trò thiết yếu trong việc bảo vệ chủ quyền dữ liệu, bảo đảm quyền riêng tư cá nhân, phòng chống rủi ro an ninh mạng và tạo nền tảng vững chắc cho chuyển đổi số quốc gia. Kiến trúc được thiết kế theo mô hình nhiều tầng (layered defense), tuân thủ nguyên tắc “phòng ngừa từ gốc - kiểm soát theo vai trò - giám sát liên tục - tuân thủ pháp lý - đánh giá định kỳ”, đồng thời bảo đảm tương thích với các tiêu chuẩn quốc tế như ISO/IEC 27001, NIST SP800, GDPR...

b) Kiến trúc an toàn dữ liệu bao gồm 8 lớp

- Lớp 1 - Dữ liệu gốc.
- Lớp 2 - Phân loại và gán nhãn dữ liệu.
- Lớp 3 - Quản lý truy cập và mã hóa dữ liệu.
- Lớp 4 - Che giấu dữ liệu, tính toán an toàn và truy vết.
- Lớp 5 - Giám sát và đánh giá theo khung “5 yếu tố an toàn”.
- Lớp 6 - Kiểm soát pháp lý và quản lý vòng đời dữ liệu.

- Lớp 7 - Công cụ kỹ thuật bắt buộc.

- Lớp 8 - Báo cáo và đánh giá an toàn dữ liệu định kỳ (bắt buộc áp dụng).

V. KHUNG THAM CHIẾU CHIẾN LƯỢC DỮ LIỆU, KIẾN TRÚC DỮ LIỆU

1. Xây dựng chiến lược dữ liệu

a) Các nội dung cơ bản của Chiến lược dữ liệu:

- Quan điểm

- Tầm nhìn

- Mục tiêu

+ Mục tiêu tổng quát

+ Mục tiêu cụ thể

- Nhiệm vụ và giải pháp

+ Hoàn thiện thể chế

+ Xây dựng kiến trúc dữ liệu

+ Xây dựng khung quản trị, quản lý dữ liệu

+ Xây dựng từ điển dữ liệu

+ Phát triển, thu thập, tạo lập dữ liệu

+ Phát triển ứng dụng phục vụ Chính phủ số, kinh tế số, xã hội số

+ Về phát triển hạ tầng dữ liệu

+ Bảo đảm an ninh dữ liệu, an toàn thông tin và bảo mật dữ liệu

+ Về nguồn lực (con người, kinh phí)

- Kinh phí thực hiện (nếu có)

- Tổ chức thực hiện.

b) Xây dựng danh mục dữ liệu của bộ, cơ quan trung ương, địa phương

Xây dựng danh mục dữ liệu mức khái niệm (các nhóm dữ liệu phản ánh các đối tượng/thực thể) của ngành/cơ quan theo các nhóm:

- Danh mục dữ liệu gốc của ngành/khối cơ quan:

+ Dữ liệu danh mục

+ Dữ liệu chủ

+ Dữ liệu giao dịch, dữ liệu nghiệp vụ chuyên ngành

- Danh mục dữ liệu khác cần thiết phục vụ các mục tiêu của bộ, cơ quan trung ương, địa phương:

- + Dữ liệu cần thu thập từ các đơn vị khác
- + Dữ liệu từ các nguồn mở
- + Dữ liệu tổng hợp
- Định hướng về dữ liệu mở, dữ liệu gia tăng
- Quy định về xây dựng từ điển dữ liệu cho từng nhóm dữ liệu
- c) Xác định các nhiệm vụ và giải pháp chiến lược
- Các nhiệm vụ xây dựng, cập nhật, duy trì dữ liệu
- Xác định các nhiệm vụ và giải pháp khai thác, sử dụng dữ liệu
- Xác định các nhiệm vụ và giải pháp chia sẻ dữ liệu.

2. Xây dựng kiến trúc dữ liệu

Khung kiến trúc dữ liệu cấp bộ, cơ quan trung ương, địa phương bao gồm các nội dung cơ bản sau:

- a) Tầm nhìn, mục tiêu và phạm vi kiến trúc
 - Tầm nhìn: Xác định tầm nhìn về kiến trúc dữ liệu trong lộ trình trung hạn.
 - Mục tiêu: xác định mục tiêu của kiến trúc dữ liệu gồm:
 - + Mục tiêu tổng quát
 - + Mục tiêu cụ thể
 - Phạm vi áp dụng: Phạm vi điều chỉnh của kiến trúc dữ liệu
- b) Nguyên tắc, yêu cầu
 - Nguyên tắc: Xác định các nguyên tắc chung của kiến trúc dữ liệu
 - Yêu cầu: Xác định các yêu cầu cụ thể về kiến trúc
- c) Xây dựng kiến trúc dữ liệu
 - Kiến trúc dữ liệu hiện trạng gồm:
 - + Mô tả hiện trạng kiến trúc dữ liệu của các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan khác trong hệ thống chính trị.
 - + Đánh giá khả năng đáp ứng các tầm nhìn và các mục tiêu
 - Kiến trúc dữ liệu mục tiêu: mô tả kiến trúc dữ liệu mục tiêu của ngành, khối cơ quan, bao gồm các nội dung cần có:
 - + Sơ đồ tổng quát kiến trúc dữ liệu: mô tả tổng quát các thành phần của kiến trúc dữ liệu đáp ứng lưu trữ dữ liệu; tích hợp, chia sẻ và điều phối dữ liệu; quản lý dữ liệu; vận hành dữ liệu; quản trị dữ liệu; quản lý từ điển dữ liệu; quản lý danh mục dữ liệu và quản trị dữ liệu.

+ Mô hình dữ liệu khái niệm: sơ đồ thiết kế sơ bộ nội dung của dữ liệu, trong đó thể hiện rõ các đối tượng được quản lý, các trường dữ liệu cơ bản đóng vai trò mô tả và định danh đối tượng quản lý, mối quan hệ giữa các đối tượng được quản lý trong cơ sở dữ liệu hoặc phạm vi dữ liệu của cơ quan, tổ chức.

+ Danh mục dữ liệu (Data catalog): mô tả danh mục các nhóm dữ liệu thuộc phạm vi quản lý. Danh mục chi tiết về dữ liệu/tài nguyên dữ liệu thuộc phạm vi quản lý được mô tả trong phần Phụ lục kèm theo.

+ Mô hình kiến trúc logic các cơ sở dữ liệu: tổ chức, sắp xếp rõ các cơ sở dữ liệu dùng chung và các cơ sở dữ liệu dùng riêng:

Các cơ sở dữ liệu dùng chung: từ điển dữ liệu quản lý siêu dữ liệu (dữ liệu đặc tả), dữ liệu chủ, dữ liệu giao dịch nghiệp vụ, dữ liệu thu thập từ các cơ quan khác, dữ liệu lớn thu thập từ nguồn mở, dữ liệu tổng hợp, dữ liệu phân tích, dữ liệu mở...

Các cơ sở dữ liệu dùng riêng phục vụ hoạt động nghiệp vụ chuyên ngành và quản trị nội bộ;

+ Mô hình logic luồng dữ liệu: mô hình tích hợp, chia sẻ và điều phối dữ liệu với các ngành/khoá cơ quan khác với Trung tâm dữ liệu quốc gia;

+ Mô hình các nền tảng dữ liệu, hạ tầng dữ liệu

+ Liệt kê và mô tả các danh mục dữ liệu chi tiết, cụ thể (trường thông tin) trong các cơ sở dữ liệu dùng chung.

- Kiến trúc an toàn dữ liệu 8 lớp theo Khung kiến trúc dữ liệu quốc gia.

d) Xác định các nhiệm vụ, giải pháp và tổ chức thực hiện

- Nhiệm vụ, giải pháp: xác định các nhiệm vụ, giải pháp để triển khai kiến trúc dữ liệu mục tiêu từ kiến trúc dữ liệu hiện trạng.

- Kinh phí thực hiện: xác định nguồn kinh phí và khái toán để triển khai thực hiện.

- Tổ chức thực hiện: đưa ra lộ trình và mốc triển khai các nhiệm vụ.

Chương III

KHUNG QUẢN TRỊ, QUẢN LÝ DỮ LIỆU QUỐC GIA

I. MỤC TIÊU, PHẠM VI

1. Mục tiêu

a) Thiết lập hệ thống quản trị dữ liệu thống nhất, minh bạch và hiệu quả trên toàn hệ thống cơ quan hành chính, bảo đảm quản lý dữ liệu xuyên suốt từ Trung ương đến địa phương và làm tài liệu tham chiếu cho các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan trung ương của các tổ chức chính trị, Viện kiểm sát nhân dân tối cao, Toà án nhân dân tối cao, các cơ quan của Quốc hội, các cơ quan Trung ương thuộc Mặt trận Tổ quốc Việt Nam, Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương, cơ quan, tổ chức, cá nhân trong hoạt động trực tiếp hoặc có liên quan đến các nội dung của Khung quản trị, quản lý dữ liệu quốc gia.

b) Hoàn thiện thể chế, hành lang pháp lý để dữ liệu trở thành tài sản có giá trị, được xác lập quyền sở hữu, quyền sử dụng, quyền khai thác và quyền giao dịch hợp pháp, từ đó thúc đẩy đổi mới sáng tạo, phát triển sản phẩm, dịch vụ dữ liệu và phát triển kinh tế số.

c) Bảo đảm chất lượng, tính toàn vẹn, an toàn, bảo mật và khả năng chia sẻ của dữ liệu số, tạo điều kiện thuận lợi cho việc khai thác, liên thông và tái sử dụng dữ liệu giữa các cơ quan nhà nước, tổ chức, doanh nghiệp và người dân trên cơ sở phân quyền, phân lớp để dữ liệu trở thành nền tảng cốt lõi phục vụ chuyển đổi số quốc gia.

2. Phạm vi

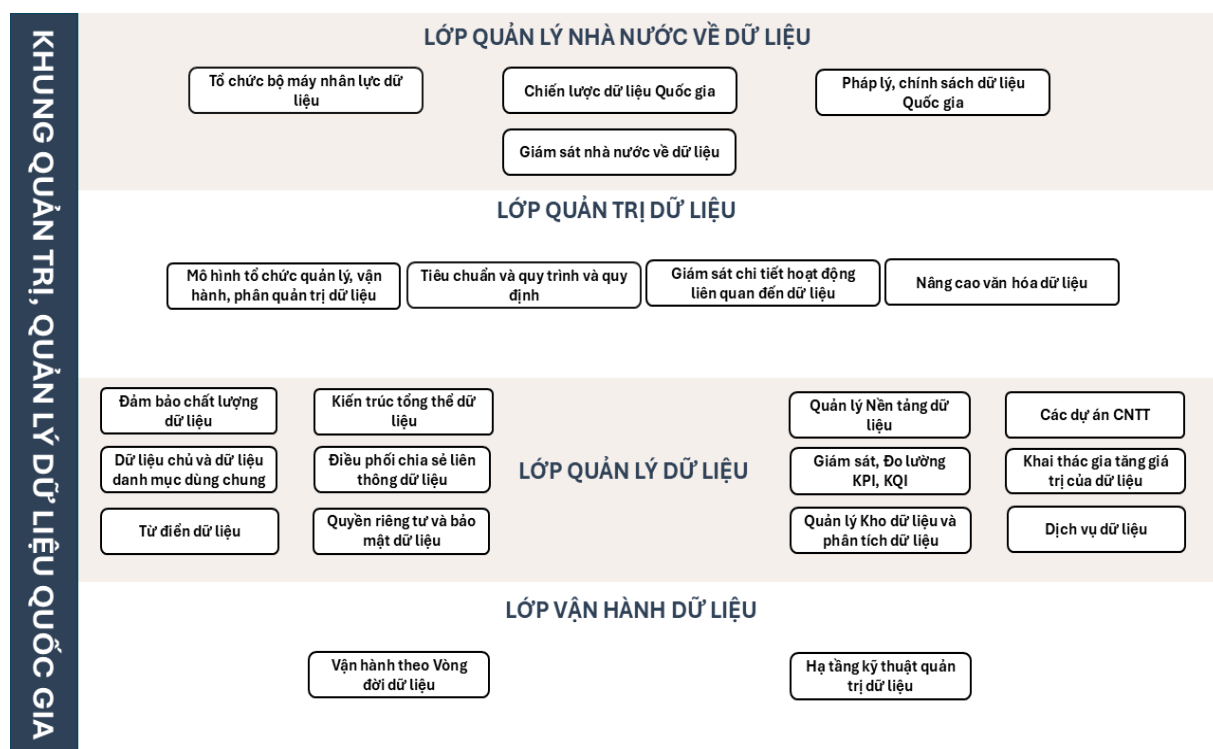
Áp dụng thống nhất tại các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan trung ương của các tổ chức chính trị, Viện kiểm sát nhân dân tối cao, Toà án nhân dân tối cao, các cơ quan của Quốc hội, các cơ quan Trung ương thuộc Mặt trận Tổ quốc Việt Nam, Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương, cơ quan, tổ chức, cá nhân trong hoạt động trực tiếp hoặc có liên quan đến các nội dung của Khung quản trị, quản lý dữ liệu quốc gia trên toàn bộ vòng đời dữ liệu số, từ thu thập, tạo lập, cập nhật, phân loại, dán nhãn, lưu trữ, bảo vệ, xử lý, chia sẻ, khai thác đến điều phối và hủy dữ liệu.

Quản lý toàn diện dữ liệu số trong các lĩnh vực phát triển kinh tế, xã hội, bảo đảm quốc phòng, an ninh, trật tự an toàn xã hội, phục vụ người dân và doanh nghiệp; làm nền tảng triển khai thống nhất các hoạt động xây dựng, kết nối, chia sẻ, tái sử dụng dữ liệu, hướng đến hình thành không gian dữ liệu tin cậy và hệ sinh thái dữ liệu quốc gia hiệu quả, an toàn, bền vững.

II. NGUYÊN TẮC

1. Thống nhất mô hình quản lý nhà nước về dữ liệu bảo đảm duy trì, phát triển dữ liệu và tuân thủ Khung kiến trúc dữ liệu quốc gia.
2. Bảo đảm xác định rõ vai trò và trách nhiệm của chủ sở hữu dữ liệu, người quản lý, người sử dụng dữ liệu.
3. Các quy trình quản lý dữ liệu phải công khai, rõ ràng, dễ hiểu và được ghi chép đầy đủ.
4. Bảo đảm hiệu quả quản lý, chất lượng dữ liệu: quản trị, quản lý tập trung, đúng đủ sạch sống; tuân thủ các quy định của pháp luật và các tiêu chuẩn, bảo đảm chất lượng dữ liệu, tính toàn vẹn và độ tin cậy dữ liệu; bảo đảm an ninh mạng và an toàn dữ liệu.

III. KHUNG QUẢN TRỊ, QUẢN LÝ DỮ LIỆU QUỐC GIA



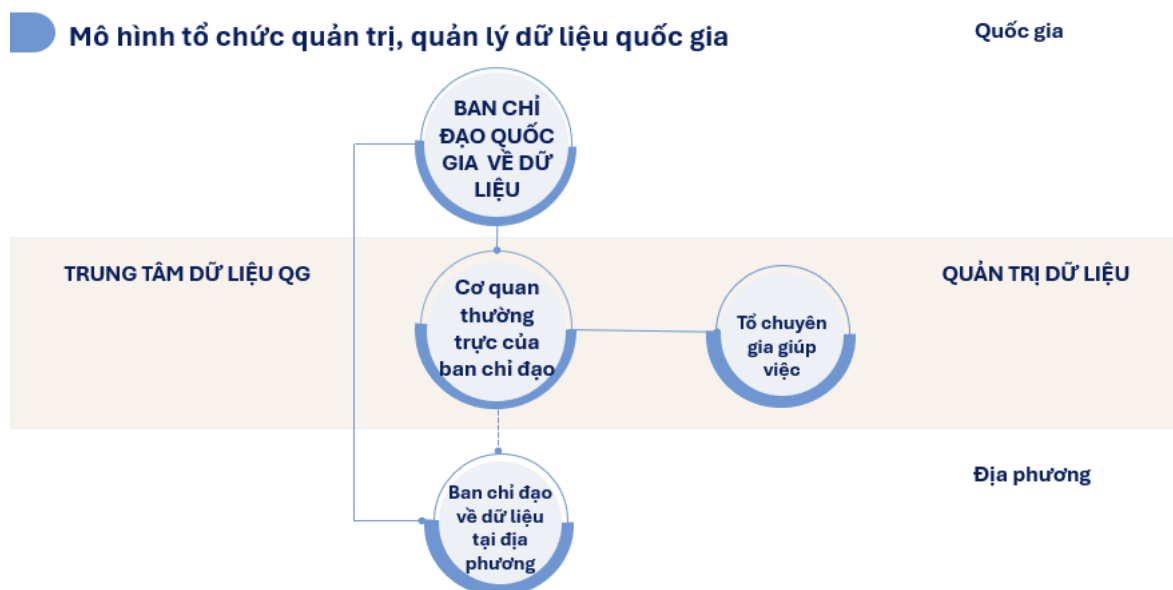
Hình 22: Khung Quản trị, quản lý dữ liệu quốc gia

Khung quản trị, quản lý dữ liệu quốc gia bao gồm 4 lớp sau:

A. Lớp quản lý nhà nước về dữ liệu

Lớp quản lý nhà nước về dữ liệu bao gồm các nội dung sau:

(1) Quản lý nhà nước về dữ liệu



Hình 23: Mô hình tổ chức quản trị, quản lý dữ liệu quốc gia

Ban Chỉ đạo quốc gia về dữ liệu đã được thành lập theo Quyết định số 2319/QĐ-TTg ngày 20/10/2025, chịu trách nhiệm giúp Chính phủ, Thủ tướng Chính phủ chỉ đạo, giám sát thực hiện các chủ trương, chiến lược, cơ chế chính sách về dữ liệu; xây dựng, phát triển Cơ sở dữ liệu quốc gia, Cơ sở dữ liệu chuyên ngành; tích hợp, đồng bộ, lưu trữ, chia sẻ, phân tích, khai thác, điều phối dữ liệu tại Cơ sở dữ liệu tổng hợp quốc gia; Giám sát xây dựng và thực hiện chiến lược quốc gia về dữ liệu, giám sát các bộ, ngành, địa phương thực hiện các chủ trương, chiến lược, cơ chế chính sách về dữ liệu, hoạt động xây dựng, khai thác, phát triển cơ sở dữ liệu chuyên ngành; tích hợp, đồng bộ, lưu trữ, chia sẻ, phân tích, khai thác, điều phối dữ liệu tại Cơ sở dữ liệu tổng hợp quốc gia.

Các bộ, ngành, địa phương thành lập ban chỉ đạo dữ liệu và chịu sự chỉ đạo trực tiếp của Ban chỉ đạo dữ liệu Quốc gia. Ban chỉ đạo dữ liệu sẽ chỉ định đơn vị, đầu mối phụ trách hoạt động quản trị và quản lý dữ liệu; tổ chức bộ máy và nhân lực phải được kiện toàn theo năng lực chuyên môn, gắn với các vị trí chức danh chuyên trách hoặc kiêm nhiệm. Chỉ đạo xây dựng và phát triển dữ liệu trong lĩnh vực phụ trách và theo mô hình quản trị quản lý dữ liệu thống nhất.

(2) Chiến lược dữ liệu quốc gia:

Chiến lược dữ liệu quốc gia là căn cứ để chỉ đạo các hoạt động xây dựng, phát triển, bảo vệ, quản trị, xử lý, sử dụng dữ liệu số. Chiến lược xác định rõ mục tiêu về xây dựng và phát triển dữ liệu trong phạm vi quốc gia phục vụ phát triển kinh tế - xã hội phục vụ chuyển đổi số quốc gia và phát triển kinh tế số gắn với bảo đảm quốc phòng, an ninh, đối ngoại, cơ yếu.

(3) Pháp lý, chính sách dữ liệu:

Để dữ liệu trở thành tài nguyên số phục vụ phát triển đất nước cần phải xây dựng và hoàn thiện khung pháp lý về phát triển dữ liệu và quản trị, quản lý dữ liệu để dữ liệu được quản trị tập trung, thống nhất tạo ra một nền tảng cốt lõi phục vụ quá trình chuyển đổi số. Các nội dung chính sách bao gồm:

- Quy chế về vai trò và trách nhiệm: quản trị dữ liệu, chủ sở hữu dữ liệu, quản lý dữ liệu, tạo lập dữ liệu, vận hành dữ liệu;

- Quy chế, quy trình về chia sẻ, điều phối dữ liệu

- Hợp tác quốc tế về dữ liệu: kết nối với các hệ thống dữ liệu quốc tế để hỗ trợ hội nhập kinh tế toàn cầu và thu hút đầu tư nước ngoài vào lĩnh vực công nghệ dữ liệu.

- Bảo đảm dữ liệu phục vụ chuyển đổi số, Chính phủ số và kinh tế số: khai thác dữ liệu phục vụ hoạt động chỉ đạo điều hành của chính phủ và bộ ngành địa phương; cung cấp các dịch vụ phân tích dữ liệu phục vụ phát triển kinh tế xã hội; khuyến khích đổi mới sáng tạo trên dữ liệu để phát triển kinh tế số.

(4) Giám sát nhà nước về dữ liệu:

- Giám sát nhà nước về dữ liệu là những hoạt động để bảo đảm quản trị, quản lý và sử dụng dữ liệu hiệu quả, an toàn và bảo mật. Nội dung giám sát bao gồm giám sát sự tuân thủ pháp luật, các quy trình, quy định trong quá trình hoạt động.

- Xây dựng bộ chỉ số hiệu suất với các yếu tố như tên chỉ số, người chịu trách nhiệm, công thức đo lường, nguồn dữ liệu, tần suất thu thập và mục tiêu định lượng. Ứng dụng các công cụ đo lường tự động để đánh giá định lượng.

- Báo cáo kết quả: theo dõi hiệu suất định kỳ, tổng hợp và báo cáo cho Ban chỉ đạo quốc gia về dữ liệu để bảo đảm các hoạt động quản trị dữ liệu được thực hiện hiệu quả và đạt mục tiêu đề ra.

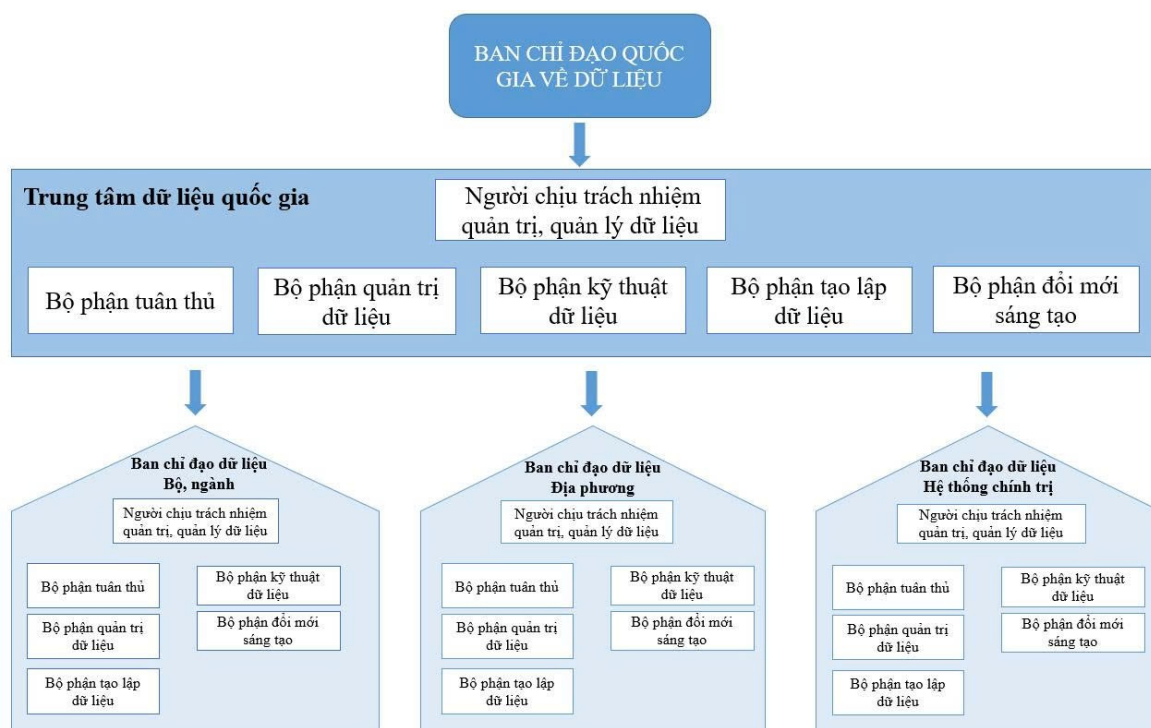
- Kế hoạch hành động khắc phục.

- Cơ chế xử lý vi phạm.

B. Lớp quản trị dữ liệu

(1) Mô hình tổ chức quản lý, vận hành, phân công trách nhiệm

- Mô hình tổ chức chung



Hình 24: Mô hình tổ chức quản lý, vận hành quản trị dữ liệu

Thành lập mô hình tổ chức quản lý, vận hành dữ liệu tại các cơ quan, tổ chức với vai trò và trách nhiệm như sau:

+ Bộ phận tuân thủ: giúp việc cho Người chịu trách nhiệm quản trị, quản lý dữ liệu thống nhất quản lý nhà nước về dữ liệu theo quy định của pháp luật. Thực hiện quản lý, giám sát, các hoạt động xây dựng, phát triển, bảo vệ, quản trị, xử lý, sử dụng, đánh giá chất lượng dữ liệu, bảo đảm an ninh, an toàn dữ liệu theo quy định của pháp luật; thanh tra, kiểm tra, giải quyết khiếu nại, tố cáo và xử lý vi phạm pháp luật về dữ liệu, tổ chức đánh giá, giám sát các hoạt động liên quan đến dữ liệu cốt lõi, dữ liệu quan trọng, dữ liệu cá nhân theo quy định của pháp luật.

+ Bộ phận quản trị dữ liệu: nghiên cứu, đề xuất các vấn đề liên quan đến hoàn thiện cơ sở pháp lý, chính sách, liên quan đến phát triển, chia sẻ, bảo vệ, quản trị, xử lý, khai thác dữ liệu; xây dựng chương trình, kế hoạch xây dựng, phát triển dữ liệu; xây dựng tiêu chuẩn kỹ thuật, quy chuẩn, quy chế, quy trình vận hành khai thác dữ liệu; thực hiện 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.

+ Bộ phận kỹ thuật dữ liệu: trực tiếp quản lý, sử dụng, tích hợp, đồng bộ, lưu trữ, chia sẻ, phân tích, khai thác, điều phối, bảo đảm hoạt động của các dịch vụ được cung cấp; xây dựng và tổ chức quản trị, vận hành hạ tầng vô trạm (hệ thống thiết bị công nghệ thông tin, hệ thống điện, hệ thống làm mát và các hệ thống, trang thiết bị khác) bảo đảm an toàn; tổ chức thực hiện công tác trực các hệ thống 24/7; khắc phục sự cố, bảo đảm an ninh vật lý.

+ Bộ phận tạo lập dữ liệu: giúp Người chịu trách nhiệm quản trị, quản lý dữ liệu tổ chức chỉ đạo, hướng dẫn, kiểm tra và tổ chức thực hiện tạo lập, thu thập, cập nhật, đồng bộ, khai thác, điều phối dữ liệu.

+ Bộ phận đổi mới sáng tạo: nghiên cứu, tổ chức triển khai các sản phẩm, dịch vụ về dữ liệu trong hoạt động trung gian dữ liệu, phân tích, tổng hợp dữ liệu, xác thực, xác nhận dữ liệu, sản dữ liệu thực hiện theo quy định của pháp luật; hoạt động thử nghiệm có kiểm soát, phát triển khoa học công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia theo quy định của pháp luật.

- Trách nhiệm phối hợp giữa Trung tâm dữ liệu quốc gia và bộ, ngành, địa phương, bao gồm:

Trách nhiệm phối hợp giữa Bộ phận Quản trị dữ liệu tại Trung tâm dữ liệu quốc gia và Bộ phận Quản trị dữ liệu của các bộ, ngành, địa phương.

+ Bộ phận Quản trị dữ liệu tại Trung tâm dữ liệu quốc gia: tổ chức các chương trình đào tạo, bồi dưỡng nghiệp vụ chung cho toàn hệ thống; hướng dẫn xây dựng, cập nhật và áp dụng các tiêu chuẩn, quy trình quản trị dữ liệu; tiếp nhận báo cáo định kỳ, đánh giá tình hình triển khai quản trị dữ liệu tại bộ ngành và phản hồi, hỗ trợ xử lý các vướng mắc phát sinh; đề xuất điều chỉnh hoặc hoàn thiện chính sách quản trị dữ liệu quốc gia dựa trên thực tiễn triển khai từ các Bộ ngành địa phương.

+ Bộ phận Quản trị dữ liệu tại các bộ, cơ quan trung ương, địa phương: thực hiện các nhiệm vụ quản trị dữ liệu theo hướng dẫn của Bộ phận Quản trị dữ liệu tại Trung tâm dữ liệu quốc gia; chủ động cập nhật, chuẩn hóa và chia sẻ các danh mục dữ liệu, thuật ngữ chuyên ngành, bảo đảm tích hợp vào khung quản trị dữ liệu quốc gia; phối hợp đánh giá, kiểm soát chất lượng dữ liệu và thực hiện các biện pháp cải tiến khi có sai lệch theo yêu cầu của Bộ phận Quản trị dữ liệu tại Trung tâm dữ liệu quốc gia; thường xuyên trao đổi, phản hồi về khó khăn, bất cập trong quá trình triển khai để Trung tâm hỗ trợ và điều chỉnh chính sách phù hợp.

Trách nhiệm phối hợp giữa Bộ phận Tạo lập dữ liệu tại Trung tâm dữ liệu quốc gia và các bộ, cơ quan trung ương, địa phương.

+ Bộ phận tạo lập dữ liệu tại Trung tâm dữ liệu quốc gia: hướng dẫn kỹ thuật và quy trình chuẩn trong việc thu thập, số hóa, chuẩn hóa và tích hợp dữ liệu từ các Bộ ngành địa phương vào hệ thống CSDL của Trung tâm dữ liệu quốc gia; đầu mối cung cấp công cụ, biểu mẫu, danh mục chuẩn và các nền tảng công nghệ hỗ trợ hoạt động tạo lập dữ liệu tại các bộ ngành địa phương; kiểm tra, đánh giá chất lượng dữ liệu đầu vào do các Bộ phận Tạo lập dữ liệu tại bộ ngành địa phương cung cấp, phản hồi và yêu cầu hiệu chỉnh nếu chưa đạt tiêu chuẩn; tổng hợp, tích hợp dữ liệu đã được chuẩn hóa từ các bộ ngành địa phương để làm cơ sở xây dựng các CSDL quốc gia.

+ Bộ phận Tạo lập dữ liệu tại các bộ, cơ quan trung ương, địa phương: Thực hiện thu thập, rà soát, chuẩn hóa và cập nhật dữ liệu theo đúng hướng dẫn của Bộ phận Tạo lập dữ liệu tại Trung tâm dữ liệu quốc gia; bảo đảm dữ liệu được tạo lập có tính đầy đủ, chính xác, đúng định dạng và gắn với các danh mục chuẩn, từ vựng thống nhất; phối hợp kiểm thử, đánh giá chất lượng dữ liệu với Trung tâm dữ liệu quốc gia; chịu trách nhiệm khắc phục các sai sót, thiếu hụt hoặc sai lệch phát hiện trong quá trình kiểm tra liên ngành; định kỳ báo cáo kết quả triển khai, các khó khăn và đề xuất hiệu chỉnh quy trình, tiêu chuẩn bảo đảm thực tiễn triển khai hiệu quả; tham gia các chương trình đào tạo, tập huấn do Trung tâm dữ liệu quốc gia tổ chức để nâng cao năng lực cho đội ngũ làm công tác tạo lập dữ liệu.

(2) Xây dựng chi tiết các quy định, quy trình và tiêu chuẩn

Xây dựng các quy định, quy trình và tiêu chuẩn phù hợp với Khung quản trị dữ liệu quốc gia: Quy định/quy trình/tiêu chuẩn về quản trị dữ liệu, Quy định/quy trình/tiêu chuẩn an toàn, bảo vệ dữ liệu; Quy định/quy trình/tiêu chuẩn bảo đảm chất lượng dữ liệu; Quy định/quy trình/tiêu chuẩn quản lý dữ liệu chủ và dữ liệu danh mục dùng chung; Quy định/quy trình/tiêu chuẩn kiến trúc dữ liệu; Quy trình làm sạch dữ liệu, xử lý dữ liệu lỗi; Quy định/quy trình/tiêu chuẩn Tù điển dữ liệu và Các quy định/quy trình/tiêu chuẩn phát sinh trong quá trình vận hành.

Các quy định/quy trình/tiêu chuẩn cần mô tả rõ ràng các bước thực hiện, điều kiện đầu vào và đầu ra, vai trò và trách nhiệm, các chỉ số hiệu suất (KPI) và cơ chế truyền thông, cập nhật.

Các quy định, quy trình, tiêu chuẩn phải được phê duyệt, công bố nội bộ và ghi nhận các sửa đổi chính thức.

(3) Giám sát chi tiết các hoạt động liên quan đến dữ liệu

- Việc giám sát các hoạt động liên quan đến dữ liệu phải được thực hiện theo nguyên tắc liên tục, khách quan, minh bạch, có thể kiểm tra, truy vết và đánh giá được, bảo đảm việc thu thập, lưu trữ, xử lý, chia sẻ và sử dụng dữ liệu tuân thủ đúng quy định pháp luật và Khung quản trị dữ liệu quốc gia.

- Hành vi sử dụng dữ liệu phải để lại dấu vết, có cơ chế cảnh báo, kiểm soát rủi ro và được lưu trữ phục vụ công tác thanh tra, kiểm tra, điều tra khi cần thiết.

- Việc giám sát bao gồm nhưng không giới hạn với các hoạt động sau:

- + Giám sát việc bảo đảm chất lượng dữ liệu.
- + Giám sát hoạt động quản lý dữ liệu chủ và dữ liệu danh mục dùng chung.
- + Giám sát hoạt động quản lý từ điển dữ liệu.
- + Giám sát kiến trúc tổng thể dữ liệu.

- + Giám sát hoạt động điều phối chia sẻ liên thông dữ liệu.
- + Giám sát hoạt động quản lý quyền riêng tư và bảo mật dữ liệu.
- + Giám sát hoạt động quản lý nền tảng dữ liệu.
- + Giám sát đo lường KPI, KQI.
- + Giám sát việc quản lý kho dữ liệu và phân tích dữ liệu.
- + Giám sát các dự án công nghệ thông tin.
- + Giám sát hoạt động khai thác gia tăng giá trị của dữ liệu.
- + Giám sát hoạt động dịch vụ dữ liệu.

- Cơ chế giám sát gồm:

+ Thiết lập hệ thống giám sát tập trung cấp quốc gia đối với các hoạt động dữ liệu trọng yếu, bảo đảm khả năng thu thập log truy cập, ghi nhận lịch sử xử lý và phát hiện các hành vi bất thường.

+ Áp dụng các công nghệ giám sát tự động, phân tích hành vi, phân tích rủi ro để chủ động phát hiện vi phạm và ngăn ngừa sự cố dữ liệu.

+ Các cơ quan, tổ chức quản lý, khai thác dữ liệu có trách nhiệm xây dựng và vận hành hệ thống giám sát nội bộ, đồng bộ dữ liệu giám sát về hệ thống quốc gia theo quy định.

- Trách nhiệm giám sát, gồm:

+ Trung tâm dữ liệu quốc gia là đầu mối tổ chức, giám sát, kiểm tra định kỳ và đột xuất các hoạt động dữ liệu trên toàn quốc.

+ Cơ quan chủ quản dữ liệu có trách nhiệm tổ chức giám sát tại đơn vị, lưu trữ log truy cập, bảo đảm tính toàn vẹn và sẵn sàng phục vụ kiểm tra.

+ Tổ chức, cá nhân sử dụng dữ liệu phải chịu trách nhiệm về các hành vi sử dụng dữ liệu của mình và chấp hành quy định về giám sát, truy vết, báo cáo.

(4) Xây dựng văn hóa dữ liệu

Văn hóa dữ liệu được hình thành khi dữ liệu được coi là tài nguyên và được sử dụng làm cơ sở để hỗ trợ ra quyết định. Văn hóa dữ liệu được phát triển trên cơ sở các nguyên tắc: tăng cường quản lý và chia sẻ dữ liệu; ra quyết định dựa trên dữ liệu; bảo đảm tính minh bạch về dữ liệu và trách nhiệm giải trình.

Giải pháp chính để xây dựng văn hóa dữ liệu là thực hiện quản trị, quản lý dữ liệu theo khung quản trị dữ liệu; thực hiện đào tạo kỹ năng phân tích, ra quyết định, chia sẻ sáng kiến đổi mới sáng tạo dựa trên dữ liệu.

C. Lớp quản lý dữ liệu

(1) Quản lý chất lượng dữ liệu

- Tiêu chuẩn chất lượng dữ liệu

Xây dựng Bộ tiêu chuẩn chất lượng dữ liệu quốc gia cơ quan thuộc hệ thống chính trị chủ trì xây dựng và ban hành Bộ tiêu chuẩn chất lượng dữ liệu quốc gia để áp dụng thống nhất trên toàn hệ thống chính trị, làm căn cứ triển khai các sáng kiến quản lý chất lượng dữ liệu chuyên biệt tại các bộ, ngành, địa phương và tổ chức liên quan. Bộ tiêu chuẩn này tối thiểu bao gồm:

- Tiêu chí chất lượng dữ liệu gồm: Tính đầy đủ (Completeness): dữ liệu phải hiện diện đầy đủ các trường cần thiết để phục vụ nghiệp vụ; tính nhất quán (Consistency): dữ liệu phải đồng nhất và không mâu thuẫn giữa các hệ thống; tính chính xác (Accuracy): dữ liệu phản ánh đúng thực tế khách quan, không bị sai lệch; tính kịp thời (Timeliness): dữ liệu được cập nhật đúng thời điểm cần thiết; tính duy nhất (Uniqueness): không có bản ghi trùng lặp đối với dữ liệu định danh hoặc dữ liệu chủ.

- Phương pháp đo lường và đánh giá chất lượng dữ liệu, bao gồm việc: thiết lập ngưỡng phân trăm chấp nhận được cho từng chiều chất lượng; xây dựng công thức tính điểm tổng hợp để đánh giá mức độ tuân thủ; xây dựng và công bố Chỉ số chất lượng dữ liệu (Data Quality Index - DQI) được cập nhật định kỳ.

Ban hành Quy trình giám sát và xác định vấn đề chất lượng dữ liệu. Quy trình này cần thống nhất áp dụng trên toàn quốc, bảo đảm khả năng phát hiện sớm và xử lý chủ động các vấn đề liên quan đến chất lượng dữ liệu.

Xây dựng Quy trình khắc phục vấn đề chất lượng dữ liệu, bao gồm các bước: phân tích nguyên nhân gốc rễ (root-cause analysis); xây dựng phương án khắc phục và phòng ngừa tái diễn; thiết lập kế hoạch triển khai chi tiết, kèm theo xác nhận kết quả thực hiện.

Xây dựng và phê duyệt cam kết mức dịch vụ (Service Level Agreement – SLA), SLA phải xác định: thời gian xử lý và khắc phục sự cố dữ liệu; các tiêu chí đo lường hiệu quả cải thiện chất lượng; trách nhiệm và chế tài khi vi phạm cam kết.

- Vận hành hệ thống bảo đảm chất lượng dữ liệu

Tài liệu hóa quy tắc chất lượng dữ liệu: các bộ, ngành, địa phương cần xây dựng và công bố các quy tắc chất lượng đối với các trường dữ liệu nghiệp vụ quan trọng cấp quốc gia.

Thực hiện lập hồ sơ dữ liệu (data profiling): định kỳ thực hiện profiling để đánh giá thực trạng dữ liệu đang lưu trữ để phát hiện bất thường, lỗi sai hoặc mâu thuẫn giữa các hệ thống.

Đánh giá và điều chỉnh ngưỡng chất lượng: căn cứ kết quả profiling, tiến hành đánh giá tính phù hợp và cập nhật quy tắc hoặc ngưỡng cho các chỉ số

chất lượng đã xác lập.

Phân tích nguyên nhân và xây dựng báo cáo khắc phục, báo cáo phải bao gồm: mô tả cụ thể vấn đề phát sinh; phân tích ảnh hưởng đến hoạt động nghiệp vụ; kết quả truy vết và xác định nguyên nhân; đề xuất biện pháp cải thiện và phòng ngừa.

Xây dựng và phê duyệt kế hoạch khắc phục chất lượng dữ liệu, kế hoạch cần xác định: thời gian thực hiện; các mốc triển khai chính; đơn vị phụ trách và nguồn lực đi kèm.

Theo dõi SLA và đánh giá việc thực thi: thiết lập hệ thống giám sát việc thực hiện các SLA đã cam kết giữa các đơn vị liên quan.

Duy trì nhật ký vấn đề chất lượng dữ liệu (Data Quality Issue Log). Nhật ký này phải bao gồm: mô tả vấn đề phát sinh; biện pháp khắc phục hoặc phòng ngừa; mức cải thiện đạt được; tỷ lệ vấn đề đã xử lý trên tổng số; số lượng trường hợp vi phạm SLA.

Công cụ tự động hóa quản lý chất lượng dữ liệu. Triển khai các công cụ công nghệ hỗ trợ tự động hóa, bao gồm: phát hiện tự động các vấn đề chất lượng theo quy tắc thiết lập; phân luồng và giao việc xử lý dữ liệu cho đúng đơn vị phụ trách; cảnh báo sớm và hiển thị bảng điều khiển (dashboard) theo thời gian thực; báo cáo xu hướng và chỉ số chất lượng để hỗ trợ ra quyết định kịp thời.

(2) Quản lý dữ liệu chủ và dữ liệu danh mục dùng chung quốc gia

- Quản lý dữ liệu chủ

Chính phủ phải xác định và lập danh sách các dữ liệu chủ được sử dụng quốc gia, dựa trên danh mục dữ liệu và mô hình dữ liệu hiện có.

Rà soát hệ thống hiện tại và xác định các hệ thống tạo, đọc, cập nhật hoặc xóa dữ liệu chủ.

Xây dựng quy trình tạo mới, cập nhật và xóa/lưu trữ dữ liệu chủ và bảo đảm quy trình được phê duyệt trước khi triển khai.

Phân công vai trò sở hữu và giám sát dữ liệu chủ tương ứng cho các bên liên quan.

Dữ liệu chủ phải được định nghĩa dưới dạng thuộc tính siêu dữ liệu trong Từ điển nghiệp vụ.

Cơ quan được giao nhiệm vụ cần định kỳ rà soát và cập nhật danh sách dữ liệu chủ theo tình trạng sử dụng thực tế.

Thực hiện quản lý phiên bản để bảo đảm khả năng truy vết các thay đổi đối với dữ liệu chủ.

- Quản lý dữ liệu danh mục dùng chung quốc gia

Chính phủ phải xác định và lập danh sách các dữ liệu danh mục dùng chung được sử dụng quốc gia, dựa trên danh mục dữ liệu và mô hình dữ liệu hiện có.

Phải xác định và lập danh sách các đối tượng dữ liệu danh mục dùng chung được sử dụng trong quy trình nghiệp vụ của mình, dựa trên phân tích danh mục dữ liệu và mô hình dữ liệu hiện có.

Rà soát lại toàn bộ hệ thống thông tin hiện có và xác định những hệ thống nào đang đọc dữ liệu danh mục dùng chung.

Phát triển quy trình tạo mới, cập nhật và xóa/lưu trữ dữ liệu danh mục dùng chung, đồng thời bảo đảm quy trình này được phê duyệt trước khi áp dụng.

Thực hiện phân công vai trò sở hữu và giám sát (ownership & stewardship) cho các bên liên quan phù hợp để quản lý dữ liệu tham chiếu.

Dữ liệu danh mục dùng chung phải được bổ sung dưới dạng siêu dữ liệu (metadata) trong Từ điển nghiệp vụ (Business Glossary).

Cần định kỳ rà soát và cập nhật danh sách dữ liệu danh mục dùng chung căn cứ vào thực tế sử dụng trong các quy trình nghiệp vụ.

Phải thực hiện quản lý phiên bản để bảo đảm khả năng truy vết các cập nhật đối với danh sách dữ liệu danh mục dùng chung.

- Công cụ tự động hóa dữ liệu danh mục dùng chung và dữ liệu chủ

Xác định các yêu cầu cần thiết để tạo lập và cung cấp dữ liệu có thẩm quyền cho các hệ thống đích, bao gồm: nguồn dữ liệu cần tích hợp; các hệ thống đích tiêu thụ dữ liệu; quy tắc bảo đảm chất lượng dữ liệu; mô hình tích hợp và điều phối cần thiết.

Căn cứ vào yêu cầu, cơ quan cần lựa chọn mô hình kiến trúc tham chiếu phù hợp (ví dụ: hợp nhất - consolidation, đăng ký - registry, đồng tồn tại - coexistence) để thiết kế trung tâm dữ liệu danh mục dùng chung và dữ liệu chủ (RMD hub).

Thiết kế kiến trúc giải pháp bao gồm: ma trận CRUD (Create, Read, Update, Delete) xác định nguồn dữ liệu chủ; luồng dữ liệu giữa nguồn và RMD hub; bản ghi có thẩm quyền giữa Hub và hệ thống tiêu thụ; quan hệ giữa dữ liệu tham chiếu và ứng dụng sử dụng; mô hình dữ liệu khái niệm, logic, vật lý và mối quan hệ giữa các bảng; từ điển dữ liệu mô tả chi tiết dữ liệu và logic chuyển đổi từ nguồn đến đích.

Cần lựa chọn và triển khai một công cụ tự động hóa quản lý dữ liệu tham chiếu và dữ liệu chủ, đáp ứng tối thiểu các tiêu chí: thực thi quy tắc chất lượng dữ liệu để bảo đảm độ chính xác và tính cập nhật; kiểm soát kỹ thuật và an ninh

trong việc truy cập dữ liệu có thẩm quyền.

Công cụ phải có tính năng quản lý phiên bản để lưu vết lịch sử các thay đổi đối với dữ liệu tham chiếu và dữ liệu chủ.

(3) Từ điển dữ liệu dùng chung (nội dung cụ thể tại Chương IV).

(4) Kiến trúc dữ liệu

- Mô tả tài liệu kiến trúc dữ liệu

Kiến trúc dữ liệu cần mô tả đầy đủ các thành phần tối thiểu như sau: danh mục các nguồn dữ liệu đầu vào; các quy trình nghiệp vụ hiện hành có sử dụng dữ liệu; các hệ thống lưu trữ, xử lý, phân tích và cung cấp dữ liệu; mẫu kiến trúc dữ liệu theo các quy trình thu nhận, xử lý và phân phối dữ liệu hiện tại.

- Phê duyệt kiến trúc dữ liệu: phê duyệt kiến trúc dữ liệu trình thu nhận, xử lý và phân phối dữ liệu hiện tại.

- Giám sát và cập nhật kiến trúc dữ liệu đối với các dữ liệu: có thay đổi trong cấu trúc dữ liệu; có thay đổi về phương thức tích hợp hoặc chia sẻ dữ liệu; có thay đổi về nguồn dữ liệu hoặc hệ thống lưu trữ.

- Kiến trúc dữ liệu phải được lưu trữ tập trung tại kho lưu trữ dữ liệu do đơn vị đầu mối quản trị dữ liệu quản lý, đồng thời thiết lập phân quyền truy cập phù hợp cho các nhóm đối tượng liên quan theo quy định về an toàn, bảo mật thông tin.

- Tất cả các thay đổi đối với kiến trúc dữ liệu phải được theo dõi, kiểm soát và ghi nhận theo cơ chế quản lý phiên bản, bảo đảm khả năng truy vết, phục hồi và kiểm toán khi cần thiết.

(5) Quản lý quyền riêng tư và an toàn bảo mật dữ liệu

- Quản lý quyền riêng tư dữ liệu

Nghĩa vụ của Bên kiểm soát dữ liệu: đơn vị kiểm soát dữ liệu cam kết bảo vệ tất cả dữ liệu cá nhân có trong hệ thống, bao gồm cả dữ liệu nhận từ đơn vị khác và dữ liệu đã chia sẻ ra bên ngoài.

Khi xử lý dữ liệu cá nhân, đơn vị phải bảo đảm: thu thập dữ liệu phù hợp giới hạn trong phạm vi, mục tiêu pháp lý hoặc nghiệp vụ; xử lý dữ liệu đúng với mục đích đã đăng ký, tuyên bố; dữ liệu được cập nhật bổ sung phù hợp với mục đích xử lý; không lưu trữ dữ liệu dưới dạng có khả năng nhận diện khi mục đích thu thập hoặc xử lý đã hoàn tất.

Chỉ yêu cầu cung cấp dữ liệu tối thiểu từ chủ thể dữ liệu nếu không thể thu thập từ hệ thống điện tử sẵn có.

Triển khai biện pháp tổ chức và kỹ thuật để bảo vệ dữ liệu khỏi mất mát, phá hoại, thay đổi trái phép, rò rỉ, tấn công.

Bảo đảm an toàn cho tất cả hệ thống và thiết bị lưu trữ liên quan đến xử lý dữ liệu để ngăn chặn truy cập trái phép.

Nếu dữ liệu được xử lý bởi đơn vị thứ ba cần bảo đảm đơn vị xử lý có biện pháp kỹ thuật và tổ chức đầy đủ, có hợp đồng rõ ràng với điều khoản về thời gian lưu trữ, xóa dữ liệu, tuân thủ quy định và chỉ chia sẻ dữ liệu cho đơn vị thứ ba khi có mục đích rõ ràng, hợp pháp và được xác định trước.

Công bố tuyên bố quyền riêng tư trên trang thông tin điện tử của đơn vị, bao gồm: mục đích thu thập dữ liệu; dữ liệu nào là bắt buộc/tùy chọn; cách thức thu thập, xử lý, lưu trữ và hủy dữ liệu; danh sách các đơn vị nhận dữ liệu và liệu dữ liệu có được chuyển ra nước ngoài; hệ quả nếu không cung cấp dữ liệu.

Thông báo về quyền của chủ thể dữ liệu, bao gồm: quyền được thông tin; quyền truy cập dữ liệu của mình; quyền yêu cầu chỉnh sửa, cập nhật dữ liệu.

Xây dựng cơ chế hủy dữ liệu an toàn, ngăn chặn truy cập trái phép đồng thời thông báo khi có sự cố rò rỉ, mất mát, tấn công dữ liệu.

Nghĩa vụ của Đơn vị xử lý dữ liệu cam kết: bảo vệ tất cả dữ liệu nhận được; chỉ xử lý theo chỉ đạo từ đơn vị kiểm soát dữ liệu.

- Quản lý An toàn bảo mật dữ liệu

Nguyên tắc an toàn bảo mật dữ liệu tuân thủ nguyên tắc “ phòng ngừa từ gốc - kiểm soát theo vai trò - giám sát liên tục - tuân thủ pháp lý - đánh giá định kỳ”, đồng thời bảo đảm tương thích với Kiến trúc an toàn bảo mật dữ liệu quốc gia.

Tất cả các cơ sở dữ liệu chuyên ngành, hệ thống có chứa dữ liệu chủ, cũng như các hệ thống thông tin dùng chung, kết nối liên thông hoặc phục vụ phân tích, điều hành đều phải tuân thủ đầy đủ các lớp cấu trúc, nguyên tắc thiết kế và yêu cầu kỹ thuật theo các nguyên tắc an toàn bảo mật dữ liệu. Việc tuân thủ được áp dụng theo hướng linh hoạt về mức độ, tùy theo cấp độ nhạy cảm của dữ liệu, quy mô hệ thống và vai trò trong Khung kiến trúc dữ liệu quốc gia, nhưng vẫn bảo đảm nguyên tắc thống nhất về kiểm soát rủi ro và bảo vệ an toàn thông tin. Những nguyên tắc này bao gồm:

- Khung an toàn dữ liệu 8 lớp:

+ *Lớp 1 - Dữ liệu gốc*

Là tầng nền tảng của toàn bộ hệ thống thông tin. Mỗi tập dữ liệu cần được gán mã định danh duy nhất và cấu trúc metadata chuẩn hóa. Gắn thông tin về cơ quan sở hữu, đơn vị quản lý kỹ thuật và lịch sử cập nhật.

+ *Lớp 2 - Phân loại và gán nhãn dữ liệu*

Việc phân loại dữ liệu là bắt buộc để xác định mức độ rủi ro, áp dụng các biện pháp bảo vệ phù hợp và lập kế hoạch đầu tư hiệu quả. Các tập dữ liệu đều phải có nhãn dữ liệu được quản lý qua metadata và phục vụ tự động hóa kiểm

soát quyền truy cập, lập nhật ký sử dụng và đánh giá rủi ro.

+ *Lớp 3 - Quản lý truy cập và mã hóa dữ liệu*

Việc quản lý truy cập và mã hoá dữ liệu cần áp dụng các biện pháp kỹ thuật sau:

Phân quyền truy cập dữ liệu phải được thực hiện thông qua các cơ chế phân quyền theo vai trò hoặc phân quyền theo thuộc tính: cơ chế phân quyền cần được xây dựng dựa trên các vai trò chức năng cụ thể của người dùng trong tổ chức (như quản trị hệ thống, nghiệp vụ, kiểm tra...), kết hợp với đơn vị công tác hoặc phạm vi chức năng tương ứng. Việc áp dụng cơ chế phân quyền theo vai trò bảo đảm chỉ những cá nhân được ủy quyền hợp lệ mới được truy cập tới các lớp dữ liệu phù hợp với nhiệm vụ được giao.

Xác thực đa yếu tố (MFA): áp dụng với tất cả tài khoản người dùng có quyền quản trị hệ thống hoặc quản trị dữ liệu; người dùng truy cập vào dữ liệu thuộc cấp độ nhạy cảm từ cấp 4 trở lên, bao gồm: dữ liệu tài chính, dữ liệu dân cư, hồ sơ y tế, thông tin điều hành cấp chiến lược hoặc dữ liệu có ảnh hưởng trực tiếp đến an toàn xã hội và vận hành quốc gia.

Mã hóa dữ liệu: thực hiện mã hóa dữ liệu khi truyền, khi lưu trữ, dữ liệu sao lưu, bắt buộc mã hóa tại nguồn trước khi truyền sang môi trường lưu trữ khác.

Quản lý khóa mã hóa: triển khai cơ chế phân vùng logic cho quản trị khóa (Key Management Service) theo từng bộ/ngành/địa phương.

+ *Lớp 4 - Che giấu dữ liệu, tính toán an toàn và truy vết*

Che giấu dữ liệu (Masking): áp dụng theo vai trò người dùng, bảo đảm nguyên tắc "tiết lộ tối thiểu cần thiết". Chỉ hiển thị phần thông tin phù hợp với quyền hạn và mục đích công việc của người truy cập. Việc che giấu dữ liệu là bắt buộc đối với dữ liệu cá nhân ở cấp độ 3 trở lên và dữ liệu quan trọng, cốt lõi ở cấp độ 4 và 5.

Băm và mã hóa ngữ cảnh (Hashing/Tokenization): áp dụng đối với các thông tin định danh cá nhân và mã hóa đặc thù trong quá trình xử lý dữ liệu lớn, học máy (AI) hoặc chia sẻ liên ngành. Việc token hóa bảo đảm tách rời khóa định danh khỏi nội dung gốc, giúp giảm thiểu rủi ro lộ lọt dữ liệu và hỗ trợ tuân thủ các quy định của pháp luật về bảo vệ dữ liệu cá nhân.

Xử lý trong vùng an toàn (Confidential Computing): yêu cầu đối với dữ liệu nhạy cảm cấp độ 4 trở lên, sử dụng môi trường Trusted Execution Environment (TEE).

Truy vết dữ liệu (Data Lineage): là yêu cầu bắt buộc đối với tất cả hệ thống có chức năng chia sẻ, tích hợp hoặc phân tích dữ liệu. Dữ liệu truy vết phải bảo đảm lưu lại đầy đủ các thông tin sau: nguồn gốc dữ liệu; thời điểm và điều kiện thu thập; các hành động xử lý đã thực hiện; tổ chức hoặc đơn vị thực hiện hành

động đó. Truy vết phải được tổ chức thành hệ thống tập trung hoặc tích hợp với giải pháp SIEM/DAM để phục vụ kiểm toán và cảnh báo sự cố.

+ Lớp 5 - Giám sát và đánh giá theo “5 yếu tố an toàn”

Các hệ thống dữ liệu phải được đánh giá tuân thủ 5 yếu tố sau:

Người dùng an toàn (Safe People): cá nhân truy cập vào hệ thống phải được xác thực định danh, phân quyền rõ ràng theo vai trò. Hệ thống phải ghi nhận đầy đủ lịch sử truy cập và có khả năng kiểm tra, truy vết (audit) theo thời gian thực hoặc hậu kiểm.

Mục đích sử dụng an toàn (Safe Projects): việc thu thập, xử lý, chia sẻ và lưu trữ dữ liệu chỉ được thực hiện khi đã có văn bản phê duyệt chính thức, hợp đồng sử dụng dữ liệu, hoặc căn cứ pháp lý cụ thể từ cơ quan có thẩm quyền.

Dữ liệu phù hợp (Safe Data): chỉ được khai thác dữ liệu đúng theo mục tiêu, phạm vi và phân loại đã xác định. Các thay đổi mục đích sử dụng phải được đánh giá lại mức độ rủi ro và cập nhật chính sách bảo vệ phù hợp.

Môi trường an toàn (Safe Settings): hệ thống phải được triển khai trong môi trường hạ tầng có kiểm soát vật lý và mạng, bao gồm phân vùng bảo mật, cách ly môi trường thử nghiệm khỏi môi trường sản xuất và bảo vệ hệ thống đầu cuối khỏi truy cập trái phép.

Kết quả đầu ra an toàn (Safe Outputs): thông tin đầu ra phải được kiểm duyệt bảo đảm không tiết lộ dữ liệu cá nhân, dữ liệu nhạy cảm hoặc thông tin nội bộ. Không được chia sẻ dữ liệu thô nếu không có biện pháp bảo vệ tương ứng như ẩn danh, che giấu dữ liệu hoặc mã hóa.

+ Lớp 6 - Kiểm soát pháp lý và quản lý vòng đời dữ liệu

Mỗi hệ thống thông tin phải xây dựng và ban hành quy định quản lý vòng đời dữ liệu, bảo đảm dữ liệu được kiểm soát chặt chẽ từ khâu tạo lập đến khi bị xóa bỏ, tuân thủ đầy đủ các quy định pháp luật về bảo vệ dữ liệu cá nhân và an toàn thông tin.

Nội dung quản lý vòng đời dữ liệu bao gồm các nội dung tối thiểu sau: nội dung thu thập dữ liệu theo nguyên tắc mục đích cụ thể, hợp pháp, minh bạch; quy trình lưu trữ, cập nhật, khai thác và chia sẻ dữ liệu theo từng cấp độ nhạy cảm; cơ chế hủy/xóa dữ liệu định kỳ hoặc khi hết giá trị sử dụng, có xác nhận lưu vết và audit kiểm tra.

Rà soát và xóa dữ liệu định kỳ: hệ thống phải thiết lập quy trình rà soát dữ liệu không còn giá trị nghiệp vụ, dữ liệu tạm thời hoặc lỗi thời; việc xóa dữ liệu phải được thực hiện theo quy trình đã được phê duyệt, có ghi nhận log hành động và xác nhận trách nhiệm của đơn vị quản trị.

Chia sẻ dữ liệu giữa các cơ quan, đơn vị: việc chia sẻ dữ liệu phải được

thiết lập thông qua hợp đồng dữ liệu, biên bản ghi nhớ (MOU) hoặc văn bản thỏa thuận có giá trị pháp lý tương đương. Văn bản chia sẻ dữ liệu phải quy định rõ: chủ thể chịu trách nhiệm pháp lý chính về dữ liệu được chia sẻ; quyền sở hữu dữ liệu, bao gồm việc sử dụng, chỉnh sửa, lưu trữ lại hoặc phân phối lại; phạm vi và thời gian khai thác hợp lệ; cam kết bảo mật, biện pháp kỹ thuật bảo vệ dữ liệu và trách nhiệm xử lý sự cố khi có vi phạm hoặc rò rỉ thông tin.

+ Lớp 7 - Công cụ kỹ thuật bắt buộc

Để bảo đảm năng lực giám sát, kiểm soát và bảo vệ dữ liệu toàn trình, các hệ thống thuộc phạm vi Khung Kiến trúc dữ liệu quốc gia phải áp dụng các công cụ kỹ thuật chuyên biệt, có khả năng tích hợp và vận hành tập trung. Việc triển khai các công cụ này được áp dụng đối với: các hệ thống kho dữ liệu, nền tảng tích hợp chia sẻ, trung tâm dữ liệu và hệ thống phân tích dữ liệu; các hệ thống có xử lý dữ liệu cá nhân, dữ liệu quan trọng, dữ liệu cốt lõi.

Các công cụ kỹ thuật bao gồm:

Hệ thống quản lý siêu dữ liệu và truy vết dữ liệu (Metadata & Lineage Tools): Có khả năng quản lý metadata theo tiêu chuẩn ISO/IEC 11179 hoặc tương đương; Ghi nhận đầy đủ lịch sử dữ liệu (lineage), bao gồm: nguồn gốc, thời điểm tạo lập, đơn vị xử lý, quá trình biến đổi và đối tượng truy cập.

Hệ thống giám sát sự kiện bảo mật (SIEM - Security Information and Event Management): bắt buộc triển khai đối với các hệ thống có xử lý dữ liệu từ cấp độ nhạy cảm 3 trở lên; cho phép ghi nhận, cảnh báo và phân tích các sự kiện liên quan đến bảo mật dữ liệu theo thời gian thực; có khả năng tích hợp với trung tâm điều hành an toàn thông tin (SOC) cấp ngành hoặc cấp tỉnh/thành phố.

Hệ thống ngăn ngừa rò rỉ dữ liệu (DLP - Data Loss Prevention): phát hiện và ngăn chặn hành vi rò rỉ dữ liệu qua email, USB, sao chép thủ công, in ấn trái phép; bắt buộc áp dụng đối với dữ liệu chứa thông tin cá nhân, dữ liệu nhạy cảm, dữ liệu quốc phòng - an ninh hoặc thuộc danh mục bí mật nhà nước.

Hệ thống giám sát cơ sở dữ liệu (DAM - Database Activity Monitoring): theo dõi toàn bộ truy vấn, cập nhật, sao chép, xóa dữ liệu trên các hệ quản trị cơ sở dữ liệu quan trọng; bắt buộc đối với hệ thống có lưu trữ dữ liệu tập trung (data warehouse, data mart, CSDL quốc gia/chuyên ngành); hỗ trợ truy vết và kiểm toán đầy đủ.

Hệ thống token hóa và làm mờ dữ liệu (Tokenization, Masking, Hashing): Token hóa bắt buộc đối với dữ liệu định danh cá nhân trong quá trình huấn luyện AI hoặc khai thác dữ liệu mở; áp dụng cơ chế masking (che dấu thông tin) theo vai trò và ngữ cảnh sử dụng.

Bảng điều khiển tuân thủ (Compliance Dashboard): hiển thị trạng thái tuân

thủ các tiêu chuẩn kỹ thuật, pháp lý về an toàn dữ liệu; báo cáo được trực quan hóa theo đơn vị chủ quản, loại dữ liệu và trạng thái cảnh báo; có khả năng gửi cảnh báo tự động tới cơ quan giám sát cấp trên khi có vi phạm nghiêm trọng.

Yêu cầu tích hợp và vận hành tập trung: các công cụ nêu trên bắt buộc có khả năng tích hợp với Nền tảng chia sẻ, điều phối dữ liệu (NDOP), nền tảng phân tích tập trung và các hệ thống giám sát dữ liệu theo quy định.

+ Lớp 8 - Báo cáo và đánh giá an toàn dữ liệu định kỳ

Để bảo đảm công tác kiểm tra, giám sát, minh bạch hóa tình trạng an toàn dữ liệu và tuân thủ pháp luật, các cơ quan, đơn vị triển khai hệ thống dữ liệu thực hiện chế độ báo cáo và đánh giá an toàn dữ liệu theo các yêu cầu sau:

Thực hiện báo cáo ngay khi có sự cố an toàn thông tin mức độ nghiêm trọng (cấp độ III trở lên theo quy định) và theo yêu cầu của cơ quan quản lý nhà nước có thẩm quyền với một số thông tin báo cáo cơ bản sau:

STT	Nhóm chỉ số	Nội dung đánh giá
1	Mức độ mã hóa dữ liệu	Tỷ lệ dữ liệu có áp dụng mã hóa trong lưu trữ và truyền tải.
2	Truy cập trái phép bị ngăn chặn	Số lượng và tỷ lệ truy cập không hợp lệ bị phát hiện/ngăn chặn.
3	Tốc độ phản ứng sự cố	Thời gian trung bình từ khi phát hiện đến khi khắc phục.
4	Nhật ký truy cập	Tỷ lệ hành vi truy cập có log đầy đủ và không thể bị sửa đổi.
5	Tỷ lệ kiểm tra định kỳ hoàn thành	Kết quả tự đánh giá, rà soát an toàn định kỳ.
6	Kiểm toán bên thứ ba (nếu có)	Tên đơn vị kiểm toán, thời gian và mức độ tuân thủ.

Tổ chức thực hiện và chịu trách nhiệm: đơn vị vận hành dữ liệu có trách nhiệm tổng hợp, chuẩn bị và gửi báo cáo theo đúng yêu cầu; đơn vị sở hữu dữ liệu chịu trách nhiệm pháp lý về tính chính xác, đầy đủ của báo cáo; đơn vị điều phối dữ liệu thực hiện tổng hợp, giám sát và đề xuất các biện pháp cải thiện tại cấp bộ/ngành/địa phương.

Kiểm tra, giám sát và xử lý vi phạm: việc kiểm tra, giám sát và xử lý vi phạm thực hiện theo quy định của pháp luật.

(6) Điều phối, chia sẻ liên thông dữ liệu

- Cơ chế điều phối và kết nối liên thông

Xây dựng và vận hành hệ thống điều phối liên thông dữ liệu quốc gia,

bảo đảm kết nối kỹ thuật, phân luồng truy cập, kiểm soát an ninh bảo mật và giám sát sử dụng dữ liệu.

Thiết lập cơ chế điều phối tập trung thông qua nền tảng tích hợp, chia sẻ dữ liệu quốc gia, bảo đảm khả năng giám sát truy vết, đo lường hiệu quả và xử lý sự cố.

Áp dụng các giao thức, tiêu chuẩn kỹ thuật mở, thống nhất định dạng và mô hình dữ liệu, bảo đảm khả năng kết nối giữa các hệ thống dữ liệu của cơ quan, tổ chức.

- Quy định về quyền truy cập và chia sẻ

Việc truy cập, sử dụng dữ liệu liên thông phải tuân thủ phân quyền, phân cấp, có mục đích rõ ràng, được kiểm soát và ghi nhận truy vết.

Việc chia sẻ dữ liệu cá nhân, dữ liệu nhạy cảm phải có đồng thuận của chủ thể dữ liệu, trừ các trường hợp pháp luật có quy định khác.

- Giám sát và xử lý vi phạm

Các hoạt động chia sẻ, liên thông dữ liệu phải được giám sát tập trung, tự động, ghi nhận đầy đủ lịch sử truy cập, sử dụng và chia sẻ.

Các hành vi lợi dụng việc chia sẻ, liên thông để truy cập trái phép, làm sai lệch, rò rỉ, mua bán, sử dụng sai mục đích dữ liệu đều bị xử lý theo quy định của pháp luật.

(7) Quản lý nền tảng dữ liệu

- Quản lý nền tảng dữ liệu bảo đảm tính sẵn sàng, khả năng mở rộng, hiệu năng và chất lượng dữ liệu phục vụ các chức năng nghiệp vụ, điều hành, phân tích và chia sẻ.

- Luồng tích hợp dữ liệu

Dữ liệu được tích hợp theo các luồng chức năng chính sau: tích hợp dữ liệu theo lô (batch) từ các hệ thống nghiệp vụ hiện có; tích hợp dữ liệu theo thời gian thực (real-time) hoặc gần thời gian thực (near real-time) thông qua cơ chế đồng bộ; tích hợp dữ liệu từ các nguồn bên ngoài thông qua API, cổng chia sẻ dữ liệu hoặc các giao thức truyền nhận an toàn.

Việc tích hợp dữ liệu phải tuân thủ quy trình kiểm soát đầu vào (data ingestion), bao gồm xác thực nguồn, kiểm tra định dạng, kiểm tra dữ liệu trùng và gán siêu dữ liệu (metadata) nhận dạng.

Luồng dữ liệu sau tích hợp cần được phân loại và lưu trữ theo mô hình tầng dữ liệu: thô (raw), xử lý (refined), chuẩn hóa (curated), phục vụ phân tích (analytic-ready).

- Nguyên tắc tổ chức lưu trữ

Dữ liệu cần được tổ chức lưu trữ theo tầng truy cập để tối ưu hiệu suất, chi phí và mục tiêu sử dụng: dữ liệu có tần suất truy cập cao, yêu cầu phản hồi tức thì (ví dụ: dữ liệu giao dịch gần thời gian thực) được tổ chức thành vùng Nóng (Hot tier/zone/data); dữ liệu có tần suất truy cập trung bình, phục vụ báo cáo định kỳ, đối chiếu và phân tích nghiệp vụ được tổ chức thành vùng Ấm (Warm tier/zone/data); dữ liệu lịch sử, lưu trữ lâu dài, phục vụ kiểm toán, tra soát hoặc phân tích xu hướng dài hạn được tổ chức thành vùng Lạnh (Cold tier/zone/data).

Việc tổ chức lưu trữ phải bảo đảm khả năng mở rộng, bảo mật, truy xuất thuận tiện và tuân thủ chính sách vòng đời dữ liệu (data lifecycle). Các quy định về di chuyển dữ liệu giữa các tầng (tiering policy) cần được thiết lập rõ ràng, có cơ chế tự động hóa và nhật ký theo dõi (audit log).

- Chuẩn hoá dữ liệu

Dữ liệu phải được chuẩn hóa theo mô hình dữ liệu thống nhất, bao gồm định nghĩa, kiểu dữ liệu, đơn vị đo lường, mã hóa, định danh chủ thể và các siêu dữ liệu.

Các bảng dữ liệu sau chuẩn hóa cần bảo đảm không dư thừa, không mâu thuẫn và có khả năng tích hợp xuyên hệ thống.

Việc chuẩn hóa phải tuân thủ danh mục chuẩn quốc gia (nếu có), hoặc được điều phối và phê duyệt bởi cơ quan có thẩm quyền quản lý dữ liệu.

- Tự động hóa bảo đảm chất lượng dữ liệu trên nền tảng dữ liệu

Các công cụ và quy trình kiểm soát chất lượng dữ liệu phải được thiết lập tự động. Cơ chế tự động phát hiện và cảnh báo lỗi dữ liệu cần được tích hợp vào pipeline xử lý, cho phép khôi phục (rollback) hoặc cách ly (quarantine) dữ liệu sai lệch.

Hệ thống tự động phải hỗ trợ phân tích nguyên nhân sai lệch (root cause analysis) và đề xuất phương án cải thiện chất lượng. Kết quả kiểm tra chất lượng phải được ghi nhận, theo dõi và báo cáo định kỳ đến các đơn vị quản lý dữ liệu và cơ quan giám sát.

(8) Quản lý giám sát và đo lường dữ liệu (KPI, KQI)

- Bộ chỉ số (KPI, KQI) để theo dõi, đánh giá và cải tiến hoạt động quản lý dữ liệu tại các cơ quan nhà nước và xác định mức độ trưởng thành của tổ chức về các mặt quản lý dữ liệu, bao gồm chính sách, quy trình, vai trò, chất lượng dữ liệu, công nghệ hỗ trợ và khả năng vận hành. Quá trình đánh giá giúp nhận diện điểm mạnh, điểm yếu, từ đó làm cơ sở xây dựng kế hoạch cải tiến, hoàn thiện năng lực quản trị dữ liệu phù hợp với yêu cầu phát triển Chính phủ số, kinh tế số và chuyển đổi số.

- Các chỉ số Hiệu suất chính (KPI - Key Performance Indicator)

KPI là các chỉ số tập trung đánh giá chất lượng vận hành thực tế của hệ thống quản trị dữ liệu, như hiệu quả xử lý sự cố, độ tin cậy của dữ liệu, tính minh bạch và tự động hóa trong quy trình vận hành. KPI giúp đánh giá và đo lường chất lượng kết quả quản trị dữ liệu thực tế, bổ trợ cho KQI (Chỉ số Chất lượng chính) để có cái nhìn toàn diện hơn về năng lực tổ chức.

- Chỉ số Chất lượng chính (KQI - Key Quality Indicator)

KQI là các chỉ số tập trung đánh giá chất lượng vận hành thực tế của hệ thống quản trị dữ liệu, như hiệu quả xử lý sự cố, độ tin cậy của dữ liệu, tính minh bạch và tự động hóa trong quy trình vận hành. KQI giúp đánh giá “chúng ta làm tốt đến mức nào” và đo lường chất lượng kết quả quản trị dữ liệu thực tế, bổ trợ cho KPI để có cái nhìn toàn diện hơn về năng lực tổ chức.

- Bộ chỉ số đánh giá

Nhóm năng lực thiết lập nền tảng quản trị dữ liệu

Bao gồm các KPI và KQI liên quan đến khung quản trị, vai trò trách nhiệm, chính sách và tiêu chuẩn về dữ liệu

Loại chỉ số	Chỉ số (KPI/KQI)	Mô tả	Tiêu chí chấm điểm (0-5)
KPI	Tỷ lệ áp dụng chính thức khung quản trị dữ liệu (%)	Phần trăm các bộ phận/ đơn vị áp dụng khung quản trị đã được ban hành và công nhận	0: 0% không có khung
			1: <10% có khung nhưng chưa phê duyệt hoặc chưa áp dụng
			2: 10-29% áp dụng, chưa phổ biến rộng
			3: 30-69% áp dụng, có cập nhật hàng năm
			4: 70-99% áp dụng, đánh giá hiệu quả định kỳ
			5: 100% áp dụng, tự động cập nhật, báo cáo liên tục
KPI	Tỷ lệ dữ liệu/tập dữ liệu có vai trò rõ ràng (%)	Phần trăm dữ liệu được chỉ định các vai trò trong mô hình tổ chức, quản lý, vận hành, phân công trách nhiệm rõ ràng	0: 0% không xác định vai trò
			1: <10% rời rạc hoặc chưa chính thức
			2: 10-29%, có nhưng chưa đồng bộ
			3: 30-69%, đa số dữ liệu lớn đã xác định, có đào tạo
			4: 70-99%, có ma trận trách nhiệm và cập nhật định kỳ
			5: 100%, tích hợp hệ thống, kiểm tra hiệu quả định kỳ
KPI	Tỷ lệ chính sách và tiêu chuẩn dữ	Phần trăm chính sách, tiêu chuẩn quản trị dữ	0: Không có chính sách nào
			1: <10% có chính sách/thủ tục chưa bắt buộc
			2: 10-29% chính sách nhưng hiệu quả thấp, chưa được tuân thủ

	liệu được cập nhật (%)	liệu được ban hành, duy trì và tuân thủ	3: Hơn 30-69%, có cập nhật định kỳ, phổ biến
			4: 70-99%, giám sát tuân thủ, báo cáo
			5: 100%, tự động cập nhật, tuân thủ toàn tổ chức
KQI	Tỷ lệ cập nhật, đánh giá hiệu quả định kỳ (%)	Tỷ lệ các chính sách được rà soát, đánh giá và cập nhật hiệu quả theo định kỳ (quý, năm)	0: <10% không theo dõi
			1: 10-30% thủ công, chưa chuẩn
			2: 30-50% một phần tự động
			3: 50-70% có dashboard, báo cáo định kỳ
			4: 70-99% đánh giá thường xuyên
			5: 100% tự động, cảnh báo và cải tiến liên tục
KQI	Tỷ lệ báo cáo kiểm soát quyền truy cập audit đúng hạn (%)	Phần trăm báo cáo truy cập và kiểm soát quyền được phê duyệt và hoàn thành đầy đủ, đúng hạn	0: <10% báo cáo đúng hạn
			1: 10-30% báo cáo cơ bản
			2: 30-50% có báo cáo
			3: 50-70% báo cáo đều, có dashboard
			4: 70-99% báo cáo tự động
			5: 100% liên tục, cảnh báo kịp thời
KQI	Tỷ lệ phản hồi, xử lý sự cố liên quan đến quản trị dữ liệu (%)	Phần trăm sự cố, phản hồi liên quan đến quản trị dữ liệu được xử lý kịp thời (trong vòng 30 ngày)	0: <10% xử lý kịp
			1: 10-30% xử lý
			2: 30-50% xử lý
			3: 50-70% xử lý kịp
			4: 70-99% xử lý nhanh
			5: 100% xử lý triệt để, báo cáo minh bạch

Nhóm năng lực Kiểm soát chất lượng và tuân thủ

Bao gồm các chỉ số liên quan đến kiểm soát chất lượng dữ liệu, metadata và truy vết dữ liệu (data lineage).

Loại chỉ số	Chỉ số (KPI/KQI)	Mô tả	Tiêu chí chấm điểm (0-5)
KPI	Tỷ lệ dữ liệu có kiểm soát chất lượng tự động (%)	Phần trăm dữ liệu được kiểm soát chất lượng qua các giải pháp tự động hóa	0: Không kiểm soát tự động
			1: <10% kiểm soát thủ công ít hiệu quả
			2: 10-29% có rule kiểm tra cơ bản

			3: 30-69% kiểm soát đồng bộ, dashboard review
			4: 70-99% theo dõi tự động xử lý lỗi
			5: 100%, quy trình tự cải tiến, báo cáo realtime
KPI	Tỷ lệ dữ liệu có lineage rõ ràng, hỗ trợ audit (%)	Phần trăm dữ liệu có mô tả lineage chính xác, có công cụ hỗ trợ audit	0: Không có lineage
			1: <10%, chỉ thủ công, thiếu công cụ
			2: 10-29%, mô tả tay, không dùng công cụ visual
			3: 30-69%, công cụ hỗ trợ, audit ít nhất 1 lần/năm
			4: 70-99% có lineage, audit tự phục vụ
			5: 100%, lineage đầy đủ, dashboard trực quan, hỗ trợ audit liên tục
KPI	Tỷ lệ dữ liệu có metadata đầy đủ, cập nhật định kỳ (%)	Phần trăm dữ liệu có metadata đầy đủ, được cập nhật định kỳ và chính xác	0: Không có metadata
			1: <10% metadata thủ công
			2: 10-29% metadata sơ khai, chưa chuẩn
			3: 30-69% quản lý tập trung, cập nhật 1-2 lần/năm
			4: 70-99% metadata tự động cập nhật, kết nối pipeline
			5: 100% metadata đồng bộ tự động, cảnh báo lỗi tự động
KQI	Tỷ lệ yêu cầu truy vết lineage đáp ứng SLA (%)	Phần trăm yêu cầu truy vết lineage được hoàn thành kịp thời, theo đúng thỏa thuận dịch vụ (SLA)	0: <10% yêu cầu
			1: 10-30% đáp ứng
			2: 30-50% đáp ứng
			3: 50-70% đáp ứng
			4: 70-99% đáp ứng
			5: 100% đáp ứng SLA, tự động hóa báo cáo
KQI	Tỷ lệ cảnh báo lỗi dữ liệu được xử lý đúng hạn (%)	Phần trăm cảnh báo lỗi dữ liệu được phát hiện tự động và giải quyết kịp thời	0: <10% xử lý
			1: 10-30% xử lý
			2: 30-50% xử lý
			3: 50-70% xử lý
			4: 70-99% xử lý
			5: 100% xử lý ngay, có báo cáo rõ ràng
KQI	Tỷ lệ metadata dữ liệu được duy trì và	Phần trăm bản ghi metadata được duy trì, cập nhật đầy	0: <10% metadata
			1: 10-30% cập nhật
			2: 30-50% cập nhật
			3: 50-70% cập nhật

	cập nhật định kỳ (%)	đủ, đúng kỳ	4: 70-99% cập nhật tự động
			5: 100% đồng bộ tự động, cảnh báo lỗi tự động

Nhóm năng lực nền tảng công nghệ hỗ trợ và tự động hóa

Bao gồm các chỉ số liên quan đến nền tảng công nghệ phục vụ vận hành, tự động hóa và giám sát quản trị dữ liệu.

Loại chỉ số	Chỉ số (KPI/KQI)	Mô tả	Tiêu chí chấm điểm (0-5)
KPI	Có nền tảng công nghệ hỗ trợ quản trị dữ liệu	Mức độ sử dụng nền tảng công nghệ quản trị metadata, chất lượng dữ liệu, lineage	0: Không có nền tảng
			1: Có nhưng ít được dùng, hỗ trợ yếu
			2: Công cụ phân tán, sử dụng hạn chế
			3: Nền tảng đa năng, hỗ trợ tích hợp nhiều domain
			4: Tối ưu hóa, tự động alert, dashboard
			5: Nền tảng AI hỗ trợ tự động đo lường, dự báo, phân tích
KPI	Tỷ lệ nghiệp vụ quản trị dữ liệu vận hành tự động (%)	Phần trăm nghiệp vụ quản trị dữ liệu được thực hiện tự động, giám phụ thuộc thủ công	0: 0% thủ công toàn bộ
			1: <10% tự động
			2: 10-29% tự động
			3: 30-69% tự động
			4: 70-99% tự động
			5: 100% tự động hóa đầy đủ, tích hợp sâu
KQI	Số lần xảy ra sự cố hệ thống quản trị gây mất truy cập (> 1h)	Số lần mất truy cập hoặc gián đoạn thao tác nghiêm trọng ảnh hưởng đến công tác quản trị	0: >5 lần
			1: 3-5 lần
			2: 2 lần
			3: 1 lần
			4: 0 lần nhưng có cảnh báo
			5: 0 lần, tự động phục hồi, không ảnh hưởng nghiệp vụ
KQI	Tỷ lệ quy trình quản trị dữ liệu	Phần trăm các quy trình quản trị, vận hành,	0: 0% tự động
			1: 10-30% tự động
			2: 30-50% tự động

	được tự động hóa (%)	báo cáo được thực hiện tự động, tự giám sát	3: 50-70% tự động
			4: 70-99% tự động
			5: 100% tự động và tối ưu liên tục

- Phương pháp tính điểm và phân loại mức độ trưởng thành về năng lực quản trị dữ liệu

Đánh giá từng chỉ số KPI và KQI cho mỗi tổ chức, điền điểm 0 - 5 dựa vào số liệu thực tế hoặc câu trả lời khảo sát. Tính điểm trung bình nhóm năng lực bằng trung bình cộng các điểm KPI và KQI trong nhóm đó. Tính điểm tổng thể bằng trung bình cộng điểm của 3 nhóm năng lực.

Áp dụng thang điểm xếp loại mức độ trưởng thành về năng lực quản trị dữ liệu:

Điểm trung bình	Mức xếp loại
0.0 – 0.99	Không có năng lực
1.0 – 1.99	Không ổn định
2.0 – 2.99	Đang phát triển
3.0 – 3.99	Đã xác định, có quy trình
4.0 – 4.49	Quản lý hiệu quả
4.5 – 5.0	Tối ưu, tự động hóa

- Kết quả đo lường là căn cứ để xếp hạng mức độ trưởng thành về năng lực quản trị dữ liệu của cơ quan, đơn vị.

(9) Quản lý kho dữ liệu và phân tích dữ liệu

Kho dữ liệu là nền tảng lưu trữ dữ liệu chuyên biệt, được thiết kế để tích hợp dữ liệu từ nhiều nguồn khác nhau theo mô hình phân tích, phục vụ việc tổng hợp, báo cáo và ra quyết định.

Phân tích nghiệp vụ tập hợp các công cụ, kỹ thuật và dịch vụ để khai thác dữ liệu trong kho dữ liệu để tạo ra thông tin giá trị, hỗ trợ lãnh đạo, quản lý và vận hành tổ chức.

- Vai trò của kho dữ liệu

Kho dữ liệu đóng vai trò là trung tâm tích hợp, chuẩn hóa và lưu trữ dữ liệu phục vụ cho các công cụ phân tích nghiệp vụ.

Hệ thống khai thác dữ liệu từ kho dữ liệu hoặc các kho dữ liệu chuyên biệt (data mart) để tạo ra báo cáo, bảng điều khiển (dashboard), mô hình phân tích xu hướng, dự báo và ra quyết định.

Các chỉ tiêu, cấu trúc báo cáo và logic nghiệp vụ được thiết lập trong hệ thống khai thác dữ liệu phải gắn kết chặt chẽ với mô hình dữ liệu kho bảo đảm tính nhất quán, toàn vẹn và khả năng truy vết nguồn dữ liệu.

- Cơ chế tổng hợp chỉ tiêu trong kho dữ liệu

Cơ chế tổng hợp chỉ tiêu là quy trình xác định, tính toán và đối soát các chỉ tiêu phân tích, được thực hiện trên cơ sở dữ liệu tập trung, tuân thủ mô hình logic thống nhất.

Việc tổng hợp chỉ tiêu phải dựa trên logic nghiệp vụ chuẩn, sử dụng công cụ ETL (Extract - Transform - Load) hoặc ELT, với luồng xử lý dữ liệu minh bạch và có khả năng kiểm soát.

Các chỉ tiêu phân tích sau tổng hợp phải được lưu trữ theo thời gian, theo đơn vị tổ chức, lĩnh vực chuyên ngành và hỗ trợ phân tích đa chiều (OLAP).

- Tổ chức các chỉ tiêu phân tích trong kho dữ liệu

Chỉ tiêu phân tích phải được tổ chức trong danh mục chuẩn, có mã định danh, định nghĩa nghiệp vụ, đơn vị tính, tần suất cập nhật và quy tắc tính toán. Hệ thống chỉ tiêu được phân loại theo các nhóm mục tiêu: điều hành, giám sát, đánh giá hiệu quả, dự báo hoặc cảnh báo sớm. Các thay đổi về cấu trúc hoặc logic của chỉ tiêu phải được kiểm soát thông qua quy trình quản lý thay đổi (change management), bảo đảm tính thống nhất trong toàn bộ hệ thống.

- Tổ chức kho dữ liệu chuyên biệt (Data Mart)

Data mart là kho dữ liệu chuyên biệt, được thiết kế phục vụ nhu cầu phân tích của một đơn vị, lĩnh vực hoặc chức năng cụ thể. Việc tổ chức data mart phải bảo đảm tuân thủ mô hình dữ liệu logic và vật lý thống nhất, đồng thời có khả năng tích hợp linh hoạt. Các data mart có thể được tổ chức theo hướng, theo miền (domain-based), theo cơ cấu tổ chức (department-based) hoặc theo nghiệp vụ (subject-oriented), bảo đảm phục vụ tối ưu các nhóm người dùng chuyên biệt.

- Yêu cầu kỹ thuật và vận hành

Kho dữ liệu và data mart phải bảo đảm tính mở rộng, hiệu suất cao và khả năng xử lý khối lượng lớn dữ liệu. Phải có cơ chế kiểm soát chất lượng dữ liệu đầu vào, dữ liệu xử lý trung gian và dữ liệu kết quả sau tổng hợp chỉ tiêu. Các hệ thống phân tích phải hỗ trợ truy cập theo phân quyền, ghi nhận lịch sử truy cập và có công cụ kiểm soát truy vết dữ liệu từ báo cáo đến nguồn gốc ban đầu (data lineage).

(10) Quản lý các dự án công nghệ thông tin

- Khái niệm và nguyên tắc: quản lý các dự án công nghệ thông tin là quá trình lập kế hoạch, triển khai, giám sát và đánh giá các chương trình, đề án, dự án có liên quan đến dữ liệu do các cơ quan nhà nước chủ trì hoặc phối hợp thực hiện.

- Phạm vi áp dụng với các dự án liên quan đến các hoạt động: thu thập và tích hợp dữ liệu; làm sạch, chuẩn hóa và phân loại dữ liệu; phân tích, dự báo và mô hình hóa dữ liệu; xây dựng cơ sở dữ liệu; triển khai hệ thống dữ liệu mở, chia sẻ liên thông dữ liệu.

- Quản trị rủi ro và thay đổi

Thiết lập quy trình kiểm soát thay đổi trong quá trình thực hiện dự án; Xác định rủi ro về dữ liệu (rò rỉ, sai lệch, lỗi tích hợp) và xây dựng phương án xử lý.

(11) Khai thác giá trị của dữ liệu

- Dữ liệu mở

Nguyên tắc công bố dữ liệu mở

Dữ liệu mở được công bố theo các nguyên tắc: công khai theo danh mục dữ liệu mở được phê duyệt; bảo đảm không vi phạm bảo mật, quyền riêng tư; tuân thủ định dạng mở (CSV, JSON, XML...), có metadata rõ ràng; cập nhật dữ liệu định kỳ và có API truy cập nếu cần.

Kênh và hình thức khai thác

- Cổng dữ liệu quốc gia phục vụ người dân, doanh nghiệp; các API dữ liệu mở cho phép truy cập dữ liệu theo thời gian thực, có giới hạn; chia sẻ có điều kiện, có thu phí hoặc hợp tác công - tư; dịch vụ phân tích, dashboard, truy vấn dữ liệu theo yêu cầu.

- Thị trường dữ liệu

Tạo nguồn thu từ dữ liệu

Thực hiện đánh giá khai thác dữ liệu để xác định và lập danh mục đầy đủ các sản phẩm dữ liệu có tiềm năng tạo ra nguồn thu nhưng phải bảo đảm tuân thủ quy định của pháp luật trong việc điều tiết thị trường dữ liệu, cạnh tranh lành mạnh và an ninh quốc gia.

(12) Quản lý dịch vụ dữ liệu

- Khái niệm: là tập hợp các năng lực, công cụ và nền tảng kỹ thuật cho phép cung cấp, khai thác, sử dụng dữ liệu một cách hiệu quả, bảo đảm khả năng truy cập, tiêu thụ, phân tích và khai thác dữ liệu theo nhiều hình thức khác nhau, phục vụ yêu cầu quản trị, vận hành và ra quyết định.

- Thành phần dịch vụ dữ liệu gồm: lớp tiêu thụ dữ liệu (Consumption Layer - Semantic Model); dịch vụ phân tích nghiệp vụ (Business Intelligence);

kênh tiêu thụ dữ liệu (Consumption Channels); giao diện lập trình dữ liệu (Data API); dịch vụ tìm kiếm dữ liệu (Data Search); kênh truyền dữ liệu thời gian thực (Data Streaming Channel).

- Yêu cầu quản lý và vận hành dịch vụ dữ liệu: dịch vụ dữ liệu được quản lý theo quy trình, có kiểm soát truy cập, giám sát chất lượng dữ liệu và bảo vệ thông tin. Các giao diện cung cấp dịch vụ dữ liệu (API, dashboard, search...) có tính sẵn sàng, hiệu suất, toàn vẹn và bảo mật. Hệ thống cung cấp dịch vụ dữ liệu hỗ trợ phân quyền truy cập theo vai trò, ghi nhật ký hoạt động và tuân thủ quy định pháp luật liên quan.

- Trách nhiệm vận hành dịch vụ dữ liệu: bảo đảm duy trì hạ tầng kỹ thuật ổn định, an toàn và mở rộng linh hoạt theo nhu cầu khai thác dữ liệu. Cập nhật định kỳ mô hình, danh mục dữ liệu bảo đảm tính nhất quán và khả năng liên thông. Hỗ trợ người dùng truy cập, sử dụng và khai thác hiệu quả dịch vụ dữ liệu; tiếp nhận và xử lý sự cố, phản hồi yêu cầu cải tiến dịch vụ.

D. Lớp vận hành dữ liệu

(1) Vận hành vòng đời dữ liệu

- Quản lý tập trung, xuyên suốt: vòng đời dữ liệu được quản lý từ khi hình thành đến khi xóa, hủy theo quy trình chuẩn hóa, đồng bộ dữ liệu.

- An toàn, bảo mật và tuân thủ pháp luật: toàn bộ quá trình bảo đảm bảo mật dữ liệu, tuân thủ quy định pháp luật về an ninh mạng, bảo vệ dữ liệu cá nhân.

- Tối ưu hóa giá trị dữ liệu: dữ liệu được bảo đảm chuẩn hóa, làm sạch, phân tích, chia sẻ để phục vụ quản lý, điều hành và phát triển các dịch vụ số.

- Trách nhiệm giải trình: các cơ quan, đơn vị tham gia quản lý, khai thác dữ liệu chịu trách nhiệm về tính chính xác, đầy đủ, an toàn dữ liệu trong suốt vòng đời dữ liệu.

- Xây dựng cơ chế giám sát, phân quyền, ghi lại nhật ký và ban hành quy chế quản lý vòng đời dữ liệu của các cơ quan, đơn vị.

(2) Hạ tầng kỹ thuật quản trị dữ liệu

Bảo đảm vận hành hệ thống hạ tầng kỹ thuật, công nghệ, cấp phát tài nguyên, an toàn, bảo mật, tăng cường sử dụng các công nghệ và công nghệ hỗ trợ quản trị dữ liệu.

Chương IV

TỪ ĐIỂN DỮ LIỆU DÙNG CHUNG

I. MỤC TIÊU, PHẠM VI

1. Mục tiêu

Xây dựng và vận hành Từ điển dữ liệu dùng chung làm công cụ chuẩn hóa dữ liệu thống nhất, bảo đảm tính đồng bộ, tương thích, liên thông giữa các cơ sở dữ liệu, hệ thống thông tin phục vụ hiệu quả quản lý nhà nước, chuyển đổi số quốc gia và hội nhập quốc tế.

2. Phạm vi

Từ điển dữ liệu dùng chung áp dụng thống nhất tại các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan trung ương của các tổ chức chính trị, Viện kiểm sát nhân dân tối cao, Toà án nhân dân tối cao, các cơ quan của Quốc hội, các cơ quan Trung ương thuộc Mặt trận Tổ quốc Việt Nam, Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương, cơ quan, tổ chức, cá nhân trong hoạt động trực tiếp hoặc có liên quan đến các nội dung của Từ điển dữ liệu dùng chung.

II. NGUYÊN TẮC

Việc xây dựng và áp dụng Từ điển dữ liệu dùng chung tuân thủ các nguyên tắc sau:

1. Chuẩn hóa: Sử dụng thuật ngữ, định nghĩa và cấu trúc thống nhất, phù hợp với pháp luật và tiêu chuẩn quốc gia, hạn chế tối đa việc tạo thuật ngữ trùng lặp.

2. Nhất quán, xuyên suốt: áp dụng nhất quán, xuyên suốt trong tất cả các hoạt động kết nối, chia sẻ giữa các cơ sở dữ liệu, hệ thống thông tin phục vụ hiệu quả quản lý nhà nước, chuyển đổi số quốc gia và hội nhập quốc tế.

3. Tương thích và liên thông: hỗ trợ trao đổi, liên thông và tích hợp dữ liệu giữa các hệ thống, nền tảng dữ liệu trong nước và quốc tế.

4. Khả năng mở rộng: thiết kế linh hoạt, có khả năng mở rộng, cập nhật theo thay đổi chính sách, nghiệp vụ hoặc công nghệ; hỗ trợ mô tả và chú giải đa ngôn ngữ.

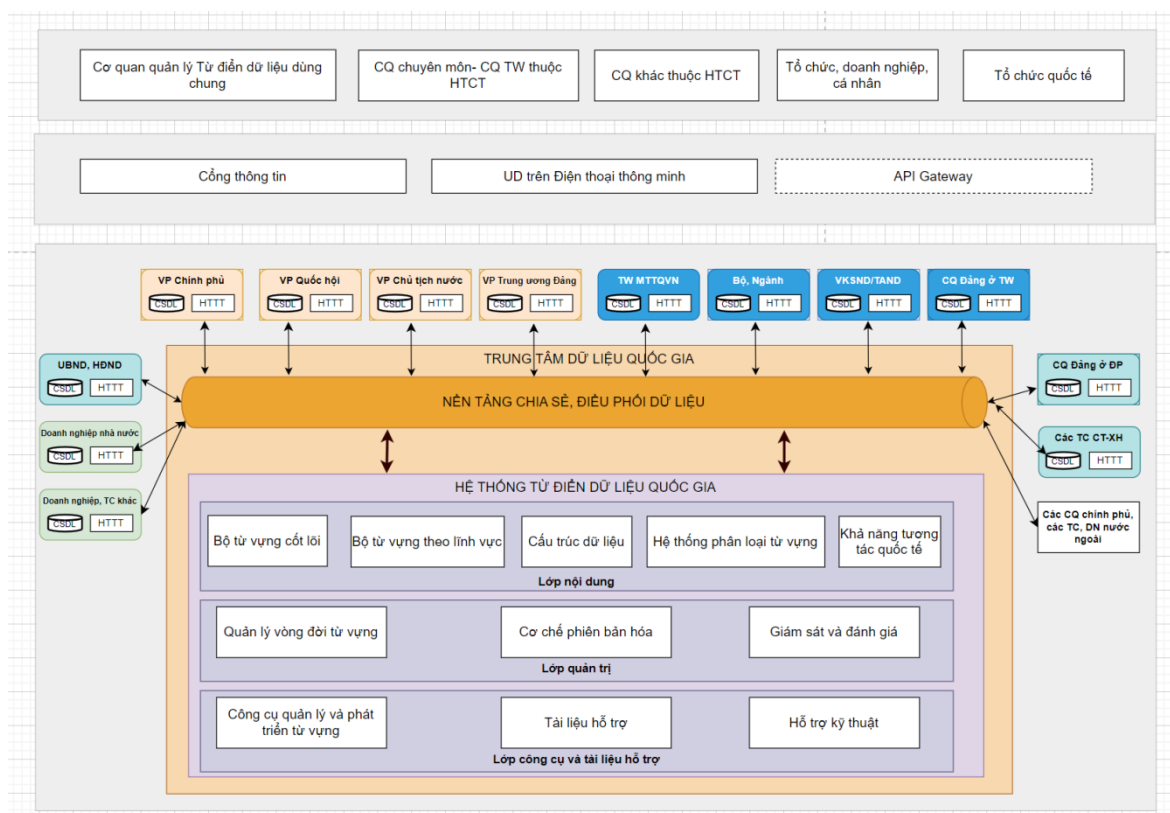
5. Quản lý tập trung, phối hợp liên ngành: xác định rõ vai trò, trách nhiệm của Cơ quan quản lý hệ thống Từ điển dữ liệu dùng chung, cơ quan quản lý ngành, đơn vị chủ sở hữu dữ liệu và nhóm chuyên gia; thực hiện phối hợp liên ngành để tránh trùng lặp và tăng hiệu quả sử dụng dữ liệu.

6. Bảo đảm an toàn thông tin, an ninh mạng: kiểm soát truy cập theo phân loại dữ liệu; bảo vệ dữ liệu cá nhân, an toàn thông tin theo Luật Dữ liệu, Luật bảo vệ dữ liệu cá nhân và các quy định liên quan.

III. HỆ THỐNG TỪ ĐIỂN DỮ LIỆU DÙNG CHUNG

1. Sơ đồ tổng quát hệ thống

a) Sơ đồ



Hình 25. Mô hình tổng quan của Hệ thống Từ điển dữ liệu dùng chung

b) Mô tả

Sơ đồ tổng quát Hệ thống từ điển dữ liệu dùng chung bao gồm các thành phần sau:

- Các cơ quan, tổ chức, cá nhân cung cấp nội dung, quản lý, sử dụng hệ thống Từ điển bao gồm: cơ quan quản lý hệ thống từ điển dữ liệu dùng chung; cơ quan chuyên môn bao gồm các bộ, cơ quan ngang bộ, cơ quan thuộc Chính phủ, cơ quan trung ương của các tổ chức chính trị, Viện kiểm sát nhân dân tối cao, Tòa án nhân dân tối cao, các cơ quan của Quốc hội, các cơ quan Trung ương thuộc Mặt trận Tổ quốc Việt Nam, Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương và các tổ chức quốc tế;

- Các kênh giao tiếp bao gồm các kênh số như ứng dụng trên điện thoại thông minh, ứng dụng trên máy tính để bàn (Desktop), ứng dụng web, API gateway;

- Nền tảng Chia sẻ, điều phối dữ liệu (NDOP) - tất cả hoạt động kết nối, chia sẻ dữ liệu bắt buộc đều phải thực hiện thông qua Nền tảng chia sẻ, điều phối dữ liệu để thực hiện việc giám sát, truy vết và đánh giá hiệu quả.

- Hệ thống của các cơ quan được kết nối, chia sẻ dữ liệu thông qua NDOP để trao đổi dữ liệu từ điển.

2. Lớp nội dung

a) Bộ từ vựng cốt lõi: tập hợp thuật ngữ cơ bản dùng chung và xuyên suốt trong nhiều ngành, hệ thống và cấp quản lý khác nhau (dữ liệu chủ quốc gia, dữ liệu danh mục dùng chung và các dữ liệu dùng chung liên ngành khác), bảo đảm tính phổ cập (tính chung), khả năng tái sử dụng.

b) Từ vựng theo lĩnh vực: thuật ngữ chuyên biệt cho từng ngành (y tế, giáo dục, tài chính,...), tương thích với bộ từ vựng cốt lõi để bảo đảm tính nhất quán.

c) Cấu trúc dữ liệu: tổ chức theo thực thể (entity) và mối quan hệ (relationship) giữa các thực thể, hỗ trợ mô tả rõ ràng kèm định dạng kỹ thuật để dễ liên kết.

d) Hệ thống phân loại từ vựng (taxonomy): xây dựng danh mục phân loại theo ngành, chủ đề, loại thực thể, mức độ nhạy cảm,... để hỗ trợ tìm kiếm, khám phá dữ liệu (data discovery) và quản lý dữ liệu.

đ) Khả năng tương tác quốc tế: bảo đảm trao đổi và liên thông dữ liệu xuyên biên giới thông qua cơ chế ánh xạ (data mapping) tới các chuẩn quốc tế phù hợp.

3. Lớp quản trị

Thiết lập quy trình quản lý vòng đời từ vựng theo các bước sau:

a) Khởi tạo/Đề xuất: các bộ/ngành, địa phương, các cơ quan, tổ chức có thể đề xuất từ vựng mới hoặc thay đổi thuật ngữ đã có.

b) Xây dựng/Rà soát sơ bộ: cơ quan quản lý hệ thống từ điển dữ liệu dùng chung thẩm định hồ sơ đề xuất, rà soát trùng lặp, ảnh hưởng, có thể yêu cầu đơn vị chuyên môn hỗ trợ mô hình hóa và đánh giá.

c) Tham vấn/Lấy ý kiến: đối với từ vựng có phạm vi sử dụng liên ngành hoặc tác động lớn, thực hiện tham vấn với các bên liên quan để đánh giá ý nghĩa, tính liên ngành.

d) Phê duyệt/Đăng ký: từ vựng được phê duyệt chính thức và được đăng ký công khai trên Hệ thống Từ điển dữ liệu dùng chung.

Cơ chế phiên bản hóa: quản lý các thay đổi, bảo đảm tính nhất quán và tránh gây gián đoạn cho các hệ thống hiện hành khi cập nhật:

- Phiên bản Chính (Major Release): được phát hành theo chu kỳ 3 năm, bao gồm cập nhật toàn diện cả Bộ Từ vựng cốt lõi và kiến trúc kỹ thuật của Hệ thống từ điển;

- Phiên bản Phụ (Minor Release): có thể được phát hành 12 tháng (không cùng năm với phiên bản chính), các phiên bản này chỉ cập nhật nội dung liên quan đến các lĩnh vực chuyên ngành (Domains) và các bảng mã;

- Quản lý thay đổi: các thay đổi được ghi lại chi tiết trong nhật ký phiên bản, các phiên bản cũ được duy trì trong một khoảng thời gian xác định để hỗ trợ các hệ thống chưa kịp cập nhật.

Giám sát và đánh giá:

- Giám sát liên tục: cơ quan quản lý hệ thống từ điển dữ liệu dùng chung giám sát việc sử dụng Từ điển dữ liệu dùng chung thông qua phản hồi từ các cơ quan, tổ chức và người dùng. Các vấn đề (như thuật ngữ không phù hợp, lỗi dữ liệu) được ghi nhận và xử lý theo quy trình phê duyệt.

- Đánh giá định kỳ: cơ quan quản lý hệ thống từ điển dữ liệu dùng chung thực hiện đánh giá hàng năm hoặc khi có nhu cầu để xem xét tính phù hợp của các thuật ngữ và định nghĩa hiện có, đánh giá nhu cầu thêm thuật ngữ mới hoặc sửa đổi thuật ngữ hiện có.

4. Công cụ và tài liệu hỗ trợ

- a) Các công cụ số hỗ trợ quản lý và phát triển từ vựng: phân hệ quản lý từ vựng; trang thông tin điện tử Từ điển dữ liệu dùng chung thuộc Cổng dữ liệu quốc gia; công cụ kiểm tra tính tương thích dữ liệu; trình dựng mô hình dữ liệu; công cụ tìm kiếm.

- b) Tài liệu, hỗ trợ kỹ thuật: tài liệu hướng dẫn sử dụng, triển khai đào tạo và hỗ trợ kỹ thuật.

5. Các thành phần mô tả cho từng thuật ngữ

Mỗi thuật ngữ trong Từ điển dữ liệu dùng chung được mô tả theo hai nhóm chính: định nghĩa nghiệp vụ (để hiểu và dùng đúng) và định nghĩa kỹ thuật (để hệ thống tương tác). Các thành phần chính bao gồm:

Định nghĩa nghiệp vụ (Business)

(1) Mã định danh duy nhất (ID)

- Mô tả: một định danh duy nhất trong hệ thống Từ điển dữ liệu dùng chung để nhận diện mỗi từ vựng, thường là một chuỗi mã đảm bảo nhận diện duy nhất trong các hệ thống và hỗ trợ tích hợp với các nền tảng quốc tế

- Yêu cầu: phải duy nhất, không thay đổi, tuân thủ ISO/IEC 11179 hoặc W3C URI, dễ dàng tích hợp vào các hệ thống quản lý dữ liệu

- Ví dụ: ndg:ID-001

(2) Tên nghiệp vụ

- Mô tả: tên hiển thị rõ ràng, dễ hiểu, phản ánh ý nghĩa nghiệp vụ đảm bảo người dùng hiểu đúng thuật ngữ, hỗ trợ tái sử dụng dữ liệu.

- Yêu cầu: ngắn gọn, dễ nhận diện, hỗ trợ đa ngôn ngữ (tiếng Việt và tiếng Anh); tuân thủ quy ước đặt tên được quy định.

- Ví dụ: HoVaTen (tiếng Việt), FullName (tiếng Anh).

(3) Định nghĩa nghiệp vụ

- Mô tả: mô tả chi tiết khái niệm và ý nghĩa nghiệp vụ, đảm bảo hiểu đúng ý nghĩa, hỗ trợ người dùng tái sử dụng dữ liệu.

- Yêu cầu: cụ thể, không mơ hồ, có thể đa ngôn ngữ để tích hợp quốc tế

- Ví dụ: HoVaTen: tên đầy đủ của cá nhân, bao gồm họ, tên đệm, và tên chính, dùng để xác định danh tính.

(4) Mô tả ngữ cảnh sử dụng

- Mô tả: chỉ rõ nơi và hoạt động mà từ vựng được sử dụng trong các quy trình nghiệp vụ, hỗ trợ sử dụng đúng ngữ cảnh, giúp các bên hiểu mục đích sử dụng.

- Yêu cầu: liên kết với quy trình nghiệp vụ cụ thể, Cung cấp thông tin về bối cảnh sử dụng (ví dụ: dịch vụ công, quản lý hành chính).

- Ví dụ: HoVaTen được dùng trong đăng ký dịch vụ công, xác minh danh tính, và quản lý hồ sơ dân cư.

(5) Nhóm dữ liệu / Phân loại

- Mô tả: phân loại theo nhóm dữ liệu hoặc lĩnh vực (dân cư, giáo dục, doanh nghiệp) hỗ trợ tìm kiếm và phân loại dữ liệu.

- Yêu cầu: tuân thủ quy tắc phân loại, hỗ trợ tìm kiếm và khám phá dữ liệu.

- Ví dụ: nhóm dữ liệu: Dân cư; Phân loại: Thông tin cá nhân.

(6) Mối quan hệ với thực thể

- Mô tả: xác định từ vựng là phần tử dữ liệu có thể là thuộc tính của đối tượng nào (Cá nhân, Hộ khẩu, Tổ chức) và cách liên kết với các từ vựng khác (ví dụ: mối quan hệ phân cấp như "là một phần của", "là loại của"), đảm bảo liên kết với các thực thể khác, hỗ trợ xây dựng mối quan hệ ngữ nghĩa.

- Yêu cầu: rõ ràng, hỗ trợ xây dựng ngữ nghĩa dữ liệu (ontology); Tuân thủ các tiêu chuẩn như W3C RDF/OWL.

- Ví dụ: HoVaTen là thuộc tính của thực thể CaNhan; liên kết với DiaChiThuongTru qua mối quan hệ hasAddress.

(7) Đơn vị tạo ra dữ liệu

- Mô tả: cơ quan chủ quản tạo ra dữ liệu gốc, đảm bảo minh bạch nguồn gốc, xác định trách nhiệm quản lý dữ liệu.

- Yêu cầu: ghi rõ tên cơ quan để xác định trách nhiệm tạo lập và quản lý.

- Ví dụ: Bộ Công an (tạo dữ liệu HoVaTen từ Cơ sở dữ liệu quốc gia về Dân cư).

(8) Đơn vị sở hữu dữ liệu

- Mô tả: cơ quan là chủ sở hữu dữ liệu.

- Yêu cầu: xác định rõ chủ sở hữu dữ liệu tuân thủ các quy định pháp luật.

(9) Ngôn ngữ

- Mô tả: ngôn ngữ sử dụng (tiếng Việt, tiếng Anh, song ngữ), hỗ trợ trao đổi dữ liệu toàn cầu, tăng khả năng tiếp cận cho người dùng quốc tế.

- Yêu cầu: hỗ trợ đa ngôn ngữ, tuân thủ ISO 639-1, Đảm bảo khả năng tích hợp quốc tế.

- Ví dụ: Tiếng Việt (HoVaTen), tiếng Anh (FullName).

(10) Phiên bản / Trạng thái

- Mô tả: phiên bản hiện tại, trạng thái (đang sử dụng, lỗi thời, đang xem xét), và lịch sử thay đổi, đảm bảo tính minh bạch và khả năng truy xuất, hỗ trợ việc theo dõi sự phát triển của bộ từ vựng.

- Yêu cầu: cập nhật định kỳ, tuân thủ quản lý vòng đời dữ liệu, ghi chú về các thay đổi đối với định nghĩa hoặc cách sử dụng thuật ngữ theo thời gian.

- Ví dụ: Phiên bản 1.0, trạng thái: Đang sử dụng, cập nhật: 2025-07-21.

(11) Tài liệu liên quan

- Mô tả: trích dẫn quy định pháp luật, tiêu chuẩn, hoặc thông tư liên quan, đảm bảo cung cấp nguồn kiểm chứng.

- Yêu cầu: liên kết đến tài liệu chính thức, như Luật Dữ liệu (2024) hoặc ISO/IEC 11179.

- Ví dụ: Luật Căn cước 2023.

(12) Từ đồng nghĩa / gần nghĩa

- Mô tả: các từ đồng nghĩa hoặc gần nghĩa để hỗ trợ tìm kiếm thông minh và xử lý ngữ nghĩa.

- Yêu cầu: xác định rõ, hỗ trợ các công cụ tìm kiếm.

- Ví dụ: HoVaTen → TenDayDu, TenCaNhan.

(13) Thuật ngữ liên quan (Related Terms)

- Mô tả: các thuật ngữ khác có mối liên hệ gần gũi hoặc thuộc cùng nhóm khái niệm, hỗ trợ xây dựng mối quan hệ ngữ nghĩa, khám phá dữ liệu.

- Yêu cầu: liên kết với ontology, hỗ trợ tìm kiếm và khám phá dữ liệu, Tuân thủ W3C RDF/OWL.

- Ví dụ: HoVaTen liên quan đến CCCD, NgaySinh, GioiTinh.

(14) Thuộc tính (Attributes)

- Mô tả: danh sách các thuộc tính hoặc tính chất định nghĩa từ vựng từ góc độ nghiệp vụ, đảm bảo hiểu rõ cấu trúc nghiệp vụ của thực thể, hỗ trợ tái sử dụng thông tin nghiệp vụ.

- Yêu cầu: liệt kê các thuộc tính mà thực thể có, giúp hiểu rõ cấu trúc nghiệp vụ; Tuân thủ ISO/IEC 11179 để định nghĩa thuộc tính.

- Ví dụ: đối với thực thể CaNhan, các thuộc tính nghiệp vụ bao gồm HoVaTen, NgaySinh, GioiTinh, DiaChiThuongTru, QuocTich,...

Định nghĩa kỹ thuật (Technical)

(15) Tên trường kỹ thuật (Technical Name)

- Mô tả: tên kỹ thuật dùng trong hệ thống, thường theo quy ước đặt tên, hỗ trợ tích hợp vào hệ thống kỹ thuật, đảm bảo tương thích với các hệ thống quốc tế.

- Yêu cầu: tuân thủ quy ước đặt tên kỹ thuật, dễ tích hợp vào cơ sở dữ liệu.

- Ví dụ: CaNhan.full_name.

(16) Loại dữ liệu

- Mô tả: kiểu dữ liệu cơ bản (String, Integer, Date, Boolean), đảm bảo xử lý dữ liệu đồng nhất, hỗ trợ tích hợp với các hệ thống quốc tế.

- Yêu cầu: tuân thủ ISO/IEC 11404 (General Purpose Datatypes)

- Ví dụ: String (cho HoVaTen).

(17) Độ dài / Định dạng

- Mô tả: số ký tự tối đa hoặc định dạng cụ thể của dữ liệu đảm bảo biểu diễn dữ liệu thống nhất, hỗ trợ tích hợp vào các hệ thống khác.

- Yêu cầu: tuân thủ theo quy tắc (như các tiêu chuẩn như ISO 8601 (ngày tháng) hoặc ISO 4217 (tiền tệ)).

- Ví dụ: HoVaTen: Độ dài tối đa 50 ký tự; NgaySinh: Định dạng YYYY-MM-DD (ví dụ: 1990-01-01).

(18) Kiểu dữ liệu mô hình (RDF/OWL)

- Mô tả: kiểu dữ liệu trong mô hình RDF/OWL (RDFS:Literal, xsd:string, xsd:date, owl:Class) hỗ trợ dữ liệu liên kết (linked data) và truy vấn ngữ nghĩa.

- Yêu cầu: tuân thủ theo quy tắc (W3C RDF/OWL) để hỗ trợ dữ liệu ngữ nghĩa.

- Ví dụ: xsd:string (cho HoVaTen), xsd:date (cho NgaySinh).

(19) Quan hệ với trường khác

- Mô tả: mối quan hệ với các trường khác (hasAddress, isPartOf, linkedTo), đảm bảo liên kết ngữ nghĩa.

- Yêu cầu: hỗ trợ xây dựng ontology, tuân thủ W3C RDF/OWL.

- Ví dụ: HoVaTen thuộc CaNhan, DiaChiThuongTru có quan hệ hasAddress với CaNhan.

(20) Thuộc tính (Attributes)

- Mô tả: đối với các thuật ngữ là thực thể hoặc đối tượng, danh sách các thuộc tính kỹ thuật định nghĩa chúng, bao gồm loại dữ liệu, độ dài, định dạng, v.v. đảm bảo xử lý dữ liệu đồng nhất.

- Yêu cầu: cung cấp chi tiết kỹ thuật cho mỗi thuộc tính để hỗ trợ xử lý dữ liệu.

- Ví dụ: Đối với CaNhan, các thuộc tính kỹ thuật bao gồm:

- + HoVaTen: Loại dữ liệu String, độ dài tối đa 50 ký tự.

- + NgaySinh: Loại dữ liệu Date, định dạng YYYY-MM-DD.

(21) Danh sách giá trị

- Mô tả: danh sách các giá trị cho phép (nếu có) đảm bảo dữ liệu hợp lệ.

- Yêu cầu: xác định rõ ràng, phù hợp với kiểu dữ liệu.

- Ví dụ: GioiTinh: [Nam, Nữ, Khác].

(22) Quy tắc chuẩn hóa

- Mô tả: quy tắc định dạng dữ liệu (in hoa, không dấu, số hóa) đảm bảo biểu diễn thống nhất, hỗ trợ tích hợp dữ liệu.

- Yêu cầu: đảm bảo tính nhất quán trong biểu diễn dữ liệu.

- Ví dụ: HoVaTen: In hoa chữ cái đầu (Nguyen Van A).

(23) API mô tả truy cập

- Mô tả: Endpoint API và định dạng truyền tải dữ liệu hỗ trợ tích hợp vào ứng dụng.

- Yêu cầu: tuân thủ các chuẩn như RESTful API, hỗ trợ các định dạng như JSON, XML.

- Ví dụ: GET /api/CaNhan/{id}, định dạng JSON.

(24) Tương thích chuẩn quốc tế (Mapping quốc tế)

- Mô tả: ánh xạ với các tiêu chuẩn quốc tế (như Schema.org, DCAT, NIEM, FOAF, SKOS,...), hỗ trợ trao đổi dữ liệu toàn cầu, tăng khả năng tái sử dụng dữ liệu.

- Yêu cầu: bảo đảm khả năng tương thích với các hệ thống toàn cầu.

- Ví dụ:

+ HoVaTen → schema:name

+ CCCD → nc:PersonIdentification

+ NgaySinh → schema:birthDate

+ GioiTinh → nc:PersonSexCode

+ DiaChiThuongTru → nc:LocationAddress

+ QuocTich → schema:nationality

(25) Miền giá trị hợp lệ (Validation Rules)

- Mô tả: quy tắc kiểm tra tính hợp lệ của dữ liệu (ràng buộc bắt buộc), đảm bảo dữ liệu hợp lệ.

- Yêu cầu: bảo đảm tính toàn vẹn, hợp lệ của dữ liệu.

- Ví dụ: CCCD: 12 ký tự số, bắt buộc.

IV. XÂY DỰNG TỪ ĐIỂN DỮ LIỆU

Bước	Nội dung thực hiện
1. Phân tích: Xác định hiện trạng dữ liệu và nghiệp vụ của bộ, ngành để xây dựng từ vựng phù hợp	- Đánh giá nhu cầu dữ liệu dùng chung - Rà soát tài liệu, biểu mẫu, quy trình nghiệp vụ và hệ thống công nghệ thông tin hiện có. - Thu thập danh sách thuật ngữ, trường dữ liệu từ các cơ sở dữ liệu chuyên ngành. - Gom nhóm các thuật ngữ tương tự nhưng khác tên, xác định từ vựng cần chuẩn hóa. Kết quả đầu ra: danh sách từ vựng thô (bao gồm thuật ngữ, mô tả, ngữ cảnh sử dụng).
2. Chuẩn hóa	- Chuẩn hóa tên, định nghĩa, loại bỏ từ đồng nghĩa hoặc gây nhầm lẫn. - Phân loại từ vựng: từ vựng cốt lõi (dùng chung, ví dụ: thông tin cá nhân), từ vựng chuyên ngành (y tế, giáo dục, tài chính...).

	- Gắn thuật ngữ với ngữ cảnh nghiệp vụ và cơ quan chịu trách nhiệm. - Ánh xạ với các chuẩn quốc tế chuyên ngành (nếu có) Kết quả đầu ra: danh mục từ vựng chuẩn hóa, có định nghĩa và phân loại rõ ràng.
3. Mô hình hóa: Thiết kế cấu trúc kỹ thuật cho từ vựng, phục vụ cả người dùng và hệ thống công nghệ thông tin	- Thiết kế cấu trúc kỹ thuật: tên, mô tả, kiểu dữ liệu, định dạng, mã hóa,... - Thiết lập quan hệ phân cấp - Mô hình hóa bằng chuẩn kỹ thuật Kết quả đầu ra: mô hình từ vựng có cấu trúc thống nhất.
4. Xác thực: Bảo đảm độ chính xác và khả năng áp dụng của từ vựng	- Thẩm định kỹ thuật: kiểm tra logic, định dạng, kiểu dữ liệu. - Thẩm định ngữ nghĩa: bảo đảm phù hợp pháp lý và hành chính. - Tham vấn chuyên ngành: gửi từ vựng đến các đơn vị liên quan để xác nhận. Kết quả đầu ra: danh mục từ vựng đã xác thực
5. Công bố	Đăng tải trên cổng Từ điển dữ liệu dùng chung để sử dụng chung, bảo đảm khả năng truy cập mở và tái sử dụng
6. Quản trị	Thiết lập quy trình quản lý vòng đời, cơ chế phiên bản hóa, cập nhật định kỳ, gắn mã định danh duy nhất cho từng thuật ngữ; giám sát áp dụng.