

BỘ GIÁO DỤC VÀ ĐÀO TẠO CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

Số: 1866 /QĐ-BGDĐT

Hà Nội, ngày 29 tháng 06 năm 2026

QUYẾT ĐỊNH

**Ban hành Khung kiến trúc dữ liệu, Khung quản trị, quản lý dữ liệu
Bộ Giáo dục và Đào tạo (Phiên bản 1.0)**

BỘ TRƯỞNG BỘ GIÁO DỤC VÀ ĐÀO TẠO

Căn cứ Luật Dữ liệu số 60/2024/QH15;

Căn cứ Nghị định số 37/2025/NĐ-CP ngày 26 tháng 02 năm 2025 của Chính phủ quy định chức năng, nhiệm vụ, quyền hạn và cơ cấu tổ chức của Bộ Giáo dục và Đào tạo;

Căn cứ Nghị định số 165/2025/NĐ-CP ngày 30 tháng 6 năm 2025 của Chính phủ quy định chi tiết và biện pháp thi hành Luật Dữ liệu;

Căn cứ Nghị định số 88/2026/NĐ-CP ngày 28 tháng 3 năm 2026 của Chính phủ quy định quản lý dữ liệu giáo dục và đào tạo;

Căn cứ Quyết định số 2439/QĐ-TTg ngày 04 tháng 11 năm 2025 của Thủ tướng Chính phủ về việc ban hành 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);

Căn cứ Quyết định số 2629/QĐ-TTg ngày 01 tháng 12 năm 2025 của Thủ tướng Chính phủ về việc phê duyệt Chương trình phát triển Chính phủ số;

Căn cứ Quyết định số 11/QĐ-BCĐCP ngày 27 tháng 02 năm 2026 ban hành kế hoạch hoạt động năm 2026 của Ban Chỉ đạo của Chính phủ về phát triển khoa học, công nghệ, đổi mới sáng tạo, chuyển đổi số và Đề án 06;

Căn cứ Quyết định số 3090/QĐ-BKHCN ngày 08 tháng 10 năm 2025 của Bộ trưởng Bộ Khoa học và Công nghệ về ban hành Khung kiến trúc tổng thể quốc gia số;

Căn cứ Quyết định số 3009/QĐ-BGDĐT ngày 31 tháng 10 năm 2025 của Bộ trưởng Bộ Giáo dục và Đào tạo về Phê duyệt Chiến lược dữ liệu đến năm 2030 của Bộ Giáo dục và Đào tạo;

Căn cứ Quyết định số 2920/QĐ-BGDĐT ngày 22 tháng 10 năm 2025 của Bộ trưởng Bộ Giáo dục và Đào tạo về ban hành Chiến lược chuyển đổi số của Bộ Giáo dục và Đào tạo giai đoạn 2025 – 2030, định hướng đến năm 2035;

Theo đề nghị của Cục trưởng Cục Khoa học, Công nghệ và Thông tin.

QUYẾT ĐỊNH:

Điều 1. Ban hành kèm theo Quyết định này Khung kiến trúc dữ liệu, Khung quản trị, quản lý dữ liệu Bộ Giáo dục và Đào tạo (Phiên bản 1.0).

Điều 2. Quyết định này có hiệu lực kể từ ngày ký.

Điều 3. Chánh Văn phòng Bộ, Cục trưởng Cục Khoa học, Công nghệ và Thông tin, Thủ trưởng các đơn vị thuộc Bộ Giáo dục và Đào tạo và các tổ chức, cá nhân liên quan chịu trách nhiệm thi hành Quyết định này. /.

Nơi nhận:

- Như Điều 3;
- Thủ tướng Chính phủ (để báo cáo);
- Các Phó Thủ tướng Chính phủ (để báo cáo);
- BCĐ của Chính phủ về phát triển KHCN, ĐMST, CDS và Đề án 06 (để báo cáo);
- BCĐ quốc gia về dữ liệu (để báo cáo);
- Tổ công tác triển khai Đề án 06; cải cách hành chính, CDS gắn với Đề án 06;
- Văn phòng Chính phủ;
- Các Bộ: Công an, KHCN (để phối hợp);
- Các Thứ trưởng (để chỉ đạo thực hiện);
- Các đơn vị thuộc, trực thuộc Bộ (để thực hiện);
- Các Sở Giáo dục và Đào tạo (để thực hiện);
- Lưu: VT, KHCNTT.



BỘ TRƯỞNG

Hoàng Minh Sơn
Hoàng Minh Sơn

BỘ GIÁO DỤC VÀ ĐÀO TẠO

KHUNG KIẾN TRÚC DỮ LIỆU BỘ GIÁO DỤC VÀ ĐÀO TẠO (PHIÊN BẢN 1.0)

*(Kèm theo Quyết định số 1866/QĐ-BGDĐT ngày 29 tháng 6 năm 2026
của Bộ trưởng Bộ Giáo dục và Đào tạo)*

NĂM 2026

MỤC LỤC

I. TẦM NHÌN, MỤC TIÊU VÀ PHẠM VI KIẾN TRÚC	6
1. Tầm nhìn	6
2. Mục tiêu	7
2.1. Mục tiêu tổng quát	7
2.2. Mục tiêu cụ thể.....	7
3. Phạm vi áp dụng.....	8
3.1. Phạm vi dữ liệu	8
3.2. Phạm vi tổ chức.....	9
3.3. Phạm vi hệ thống.....	9
II. XÂY DỰNG KIẾN TRÚC DỮ LIỆU	9
1. Nguyên tắc và yêu cầu kiến trúc	9
2. Kiến trúc dữ liệu hiện tại.....	12
3. Kiến trúc dữ liệu mục tiêu.....	13
3.1. Sơ đồ kiến trúc dữ liệu Bộ Giáo dục và Đào tạo	14
3.2. Mô hình dữ liệu	16
3.3. Mô hình logic luồng dữ liệu.....	17
3.4. Mô hình nền tảng dữ liệu	22
4. Kiến trúc an toàn dữ liệu 8 lớp.....	30
4.1. Lớp Dữ liệu	30
4.2. Lớp Phân loại và gán nhãn dữ liệu.....	31
4.3. Lớp Quản lý truy cập và mã hóa dữ liệu.....	31
4.4. Lớp Che giấu dữ liệu và truy vết	32
4.5. Lớp Giám sát và đánh giá an toàn dữ liệu	32
4.6. Lớp Kiểm soát pháp lý và vòng đời dữ liệu.....	33
4.7. Lớp Công cụ kỹ thuật bảo mật.....	33
4.8. Lớp Báo cáo và đánh giá định kỳ.....	34
III. NHIỆM VỤ VÀ GIẢI PHÁP	34
1. Giai đoạn 1 (2026-2027)	34
2. Giai đoạn 2 (2027-2030)	34

IV. KINH PHÍ THỰC HIỆN	35
V. TỔ CHỨC THỰC HIỆN.....	35

THUẬT NGỮ VÀ CÁC TỪ NGỮ VIẾT TẮT

STT	Thuật ngữ/Chữ viết tắt	Mô tả
1	ATTT	An toàn thông tin
2	API	Application Programming Interface: Giao diện lập trình ứng dụng: Bộ quy tắc cho phép các phần mềm trao đổi dữ liệu với nhau một cách chuẩn hóa.
3	GD&ĐT	Giáo dục và Đào tạo
4	CĐS	Chuyển đổi số
5	CNTT	Công nghệ thông tin
6	CSDL	Cơ sở dữ liệu
7	DLP	Data Loss Prevention: Phòng chống thất thoát dữ liệu: Giải pháp giám sát và ngăn dữ liệu nhạy cảm bị rò rỉ ra ngoài (qua email, USB, mạng...), bằng cách phát hiện và chặn các luồng truyền dữ liệu vi phạm chính sách.
8	IAM	Identity and Access Management: Quản lý định danh và truy cập: quản lý "ai là ai" và "ai được làm gì" – cấp phát, xác thực, phân quyền và thu hồi quyền truy cập của người dùng/ứng dụng vào tài nguyên dữ liệu.
9	ISO	International Organization for Standardization: Tổ chức tiêu chuẩn quốc tế
10	LGSP/LDOP	Nền tảng Tích hợp, chia sẻ dữ liệu cấp Bộ/Tỉnh
11	NDOP	Nền tảng Chia sẻ, điều phối dữ liệu Quốc gia
12	MFA	Multi-Factor Authentication: Xác thực đa yếu tố
13	SIEM	Security Information and Event Management: thu thập, tổng hợp và phân tích log/sự kiện an ninh từ nhiều nguồn để phát hiện sớm, cảnh báo và điều tra các mối đe dọa.
14	SSO	Single Sign-On: Đăng nhập một lần: Cho phép người dùng đăng nhập một lần duy nhất để truy cập nhiều hệ thống/ứng dụng khác nhau, không phải nhập lại thông tin đăng nhập cho từng hệ thống.

15	AGN	Agent Node : Cổng điều phối dữ liệu tại chỗ
16	AI	Artificial Intelligence: Trí tuệ nhân tạo
17	BI	Business Intelligence: Hệ thống phân tích, báo cáo thông minh
18	CDC	Change Data Capture :Thu thập dữ liệu thay đổi
19	ETL/ELT	Extract-Transform-Load/Extract-Load-Transform: Trích xuất - Chuyển đổi - Nạp / Trích xuất - Nạp - Chuyển đổi
20	IDS/IPS	Intrusion Detection System/Intrusion Prevention System: Hệ thống phát hiện và ngăn chặn xâm nhập: IDS phát hiện các hành vi xâm nhập bất thường và cảnh báo; IPS đi xa hơn, chủ động ngăn chặn lưu lượng tấn công theo thời gian thực.
21	KMS	Key Management System: Hệ thống quản lý khóa mã hóa: quản lý vòng đời các khóa mã hóa (tạo, lưu trữ, luân chuyển, thu hồi), đảm bảo dữ liệu được mã hóa an toàn và khóa không bị lộ.
22	ML	Machine Learning: Học máy
23	NIST	National Institute of Standards and Technology: Viện Tiêu chuẩn và Công nghệ Quốc gia (Hoa Kỳ)
24	ODS	Operational Data Store: Cơ sở dữ liệu vận hành
25	OLAP	Online Analytical Processing: Xử lý phân tích trực tuyến
26	OLTP	Online Transaction Processing: Xử lý giao dịch trực tuyến
27	RDBMS	Relational Database Management Systems: Hệ quản trị cơ sở dữ liệu quan hệ
28	SSOT	Single Source of Truth: Nguồn dữ liệu tin cậy duy nhất
29	TTHC	Thủ tục hành chính

I. TẦM NHÌN, MỤC TIÊU VÀ PHẠM VI KIẾN TRÚC

1. Tầm nhìn

Đến năm 2030, Bộ Giáo dục và Đào tạo hướng tới vận hành toàn bộ các hoạt động quản lý nhà nước, chỉ đạo điều hành và cung cấp dịch vụ công trong lĩnh vực giáo dục trên nền tảng dữ liệu số thống nhất, an toàn, thông minh, có khả năng phân tích, dự báo và hỗ trợ ra quyết định ở mức cao, phục vụ hiệu quả công tác quản lý giáo dục quốc gia.

Kiến trúc dữ liệu Bộ Giáo dục và Đào tạo được xây dựng phù hợp, đồng bộ với Khung kiến trúc dữ liệu quốc gia, Khung quản trị dữ liệu quốc gia và Kiến trúc Chính phủ số trong lĩnh vực giáo dục, hướng tới hình thành không gian dữ liệu giáo dục quốc gia thống nhất, kết nối liền mạch với hệ thống dữ liệu của Chính phủ, các bộ, ngành, địa phương và toàn bộ hệ thống cơ sở giáo dục.

Dữ liệu trở thành nguồn lực cốt lõi cho công tác quản lý, hoạch định chính sách và nâng cao chất lượng giáo dục: Dữ liệu giáo dục của Bộ Giáo dục và Đào tạo được quy hoạch, chuẩn hóa, tích hợp và quản trị thống nhất, sẵn sàng phục vụ các ứng dụng phân tích dữ liệu lớn và trí tuệ nhân tạo (AI). Dữ liệu không chỉ phục vụ hoạt động tác nghiệp thường xuyên của các đơn vị mà còn đóng vai trò là đầu vào thiết yếu cho:

- Các hệ thống phân tích, thống kê giáo dục quốc gia theo từng cấp học, lĩnh vực và địa phương;
- Các mô hình dự báo quy mô người học, đội ngũ nhà giáo, nhu cầu cơ sở vật chất và nguồn nhân lực;
- Hệ thống hỗ trợ ra quyết định trong các lĩnh vực then chốt như quy hoạch mạng lưới trường lớp, tuyển sinh, đào tạo, kiểm định chất lượng giáo dục và phát triển đội ngũ;
- Các trợ lý số và mô hình AI phục vụ công tác tham mưu, cảnh báo sớm rủi ro, đánh giá tác động chính sách giáo dục.

Việc khai thác hiệu quả dữ liệu giúp quá trình xây dựng, điều chỉnh chính sách giáo dục trở nên chính xác hơn, kịp thời hơn và dựa trên bằng chứng dữ liệu, đáp ứng yêu cầu đổi mới căn bản, toàn diện giáo dục và đào tạo. Đồng thời, dữ liệu số được khai thác để từng bước thay thế thành phần hồ sơ giấy, tái cấu trúc và đơn giản hóa thủ tục hành chính, thực hiện nguyên tắc “cung cấp thông tin một lần”, giảm thiểu yêu cầu cung cấp lại thông tin, nâng cao chất lượng dịch vụ công và trải nghiệm của người dân, doanh nghiệp.

Hệ sinh thái dữ liệu giáo dục kết nối liên thông, liền mạch ở quy mô quốc gia: Hệ sinh thái dữ liệu của Bộ Giáo dục và Đào tạo được kết nối chặt chẽ với hệ sinh thái dữ liệu Chính phủ và các bộ, ngành, địa phương có liên quan. Thông qua các nền tảng chia sẻ dữ liệu quốc gia (NDOP), địa phương (LDOP) và nền tảng tích hợp dữ liệu của Bộ, dữ liệu giáo dục được liên thông:

- Giữa các đơn vị thuộc và trực thuộc Bộ Giáo dục và Đào tạo;
- Giữa Bộ Giáo dục và Đào tạo với các bộ, ngành liên quan;

- Giữa Bộ Giáo dục và Đào tạo với các cơ sở giáo dục, tổ chức xã hội, tổ chức quốc tế và cộng đồng nghiên cứu theo cơ chế phù hợp;

- Giữa Bộ Giáo dục và Đào tạo với các địa phương trong công tác quản lý, điều hành, giám sát và phát triển giáo dục.

Qua đó hình thành Hệ sinh thái dữ liệu giáo dục đồng bộ, thống nhất, phục vụ mục tiêu phát triển nguồn nhân lực và nâng cao chất lượng giáo dục quốc gia. Cán bộ, công chức, viên chức và các hệ thống hỗ trợ được trao quyền tiếp cận và khai thác dữ liệu toàn diện. Trong môi trường làm việc số hiện đại, cán bộ, công chức, viên chức của Bộ Giáo dục và Đào tạo từ trung ương đến đơn vị trực thuộc có khả năng:

- Truy cập, khai thác và sử dụng dữ liệu giáo dục mọi lúc, mọi nơi theo đúng phân quyền;

- Vận hành các ứng dụng dữ liệu trên nhiều thiết bị, bảo đảm an toàn, an ninh thông tin;

- Sử dụng các công cụ phân tích dữ liệu, bảng điều khiển (dashboard) và hệ thống báo cáo trực tuyến;

- Ứng dụng các trợ lý số và mô hình AI nhằm nâng cao năng suất làm việc và hiệu quả chuyên môn.

Điều này góp phần đổi mới phương thức quản lý và điều hành giáo dục, nâng cao hiệu quả hoạt động của bộ máy, đồng thời thúc đẩy văn hóa làm việc dựa trên dữ liệu trong toàn ngành giáo dục.

2. Mục tiêu

2.1. Mục tiêu tổng quát

Hình thành kiến trúc dữ liệu thống nhất trong đó lấy Cơ sở dữ liệu quốc gia về Giáo dục và Đào tạo làm gốc cùng các nguồn dữ liệu từ các Cơ sở dữ liệu chuyên ngành, các hệ thống thông tin của Bộ GD&ĐT để làm nền tảng cho quản trị, điều hành và phát triển toàn diện giáo dục. Đây là nền tảng quan trọng để triển khai các hệ thống thông tin phục vụ công tác quản lý nhà nước về giáo dục, chỉ đạo điều hành của Lãnh đạo Bộ và hoạt động quản trị nội bộ của ngành giáo dục; đồng thời là căn cứ bắt buộc để thẩm định thiết kế, bảo đảm sự phù hợp với kiến trúc dữ liệu và Kiến trúc Chính phủ số trong các dự án đầu tư ứng dụng công nghệ thông tin, chuyển đổi số của Bộ Giáo dục và Đào tạo.

2.2. Mục tiêu cụ thể

Các mục tiêu cụ thể về dữ liệu số của Bộ Giáo dục và Đào tạo trong giai đoạn 2026–2030 được xác định nhằm hình thành hệ sinh thái dữ liệu giáo dục thống nhất, phục vụ hiệu quả công tác quản lý nhà nước, cung cấp dịch vụ công trong lĩnh vực giáo dục và hỗ trợ ra quyết định. Các mục tiêu cụ thể bao gồm:

- Hoàn thành xây dựng nền tảng dữ liệu giáo dục tập trung của Bộ Giáo dục và Đào tạo, có khả năng lưu trữ, quản lý và tích hợp toàn bộ dữ liệu của các đơn vị thuộc và trực

thuộc Bộ; cho phép truy cập, khai thác và phân tích dữ liệu theo đúng phân quyền, phục vụ công tác quản lý, điều hành và giám sát hoạt động giáo dục trên phạm vi toàn quốc.

- 100% dữ liệu về người học, nhà giáo, cán bộ quản lý giáo dục, cơ sở giáo dục, học bạ số, văn bằng chứng chỉ số được số hóa, chuẩn hóa và tích hợp vào Cơ sở dữ liệu quốc gia về Giáo dục và Đào tạo. Các dữ liệu này phải bảo đảm các tiêu chí “Đúng – Đủ – Sạch – Sống – Thống nhất – Dùng chung”, trở thành nguồn dữ liệu chủ dùng chung cho toàn ngành giáo dục.

- 100% cơ sở dữ liệu chuyên ngành giáo dục và đào tạo được quản lý tập trung, kết nối đồng bộ, thông suốt với Nền tảng tích hợp, chia sẻ dữ liệu (LGSP) và Cổng dữ liệu của Bộ Giáo dục và Đào tạo.

- Người dân, doanh nghiệp chỉ phải cung cấp thông tin một lần khi thực hiện thủ tục hành chính trong các lĩnh vực do Bộ Giáo dục và Đào tạo quản lý; 100% hồ sơ, kết quả giải quyết thủ tục hành chính, giấy tờ do Bộ Giáo dục và Đào tạo ban hành được lưu trữ trong kho dữ liệu số của Bộ; 80% dữ liệu được chia sẻ và sử dụng lại trong các thủ tục hành chính, hoạt động hành chính và sử dụng thay thế hoàn toàn bản giấy;

- 100% công tác tổng hợp, báo cáo định kỳ và đột xuất (trừ dữ liệu thuộc danh mục bí mật nhà nước theo quy định) của các đơn vị thuộc và trực thuộc Bộ Giáo dục và Đào tạo được thực hiện hoàn toàn trên môi trường số; từng bước tự động hóa tối đa các quy trình thu thập, tổng hợp, đối soát số liệu và trình ký báo cáo, bảo đảm tính chính xác, kịp thời, minh bạch và giảm thiểu trùng lặp.

- 70% yêu cầu khai thác dữ liệu, thông tin phục vụ công tác xây dựng chính sách giáo dục, quy hoạch mạng lưới cơ sở giáo dục, quản lý tuyển sinh, đánh giá chất lượng, thẩm định chương trình đào tạo, dự báo nhu cầu nguồn nhân lực và chỉ đạo điều hành được thực hiện, ra quyết định dựa trên dữ liệu (data-driven decision making).

3. Phạm vi áp dụng

3.1. Phạm vi dữ liệu

Kiến trúc dữ liệu áp dụng đối với toàn bộ dữ liệu phát sinh trong hoạt động quản lý nhà nước của Bộ Giáo dục và Đào tạo, bao gồm các lĩnh vực: giáo dục mầm non; giáo dục phổ thông; giáo dục thường xuyên; giáo dục đại học; giáo dục nghề nghiệp; khảo thí và kiểm định chất lượng giáo dục; phát triển đội ngũ nhà giáo và cán bộ quản lý giáo dục; chuyển đổi số giáo dục; và các hoạt động quản trị nội bộ của Bộ, không bao gồm dữ liệu thuộc danh mục bí mật nhà nước theo quy định của pháp luật.

Phạm vi áp dụng cũng bao gồm dữ liệu được tích hợp, khai thác từ các hệ thống thông tin và cơ sở dữ liệu bên ngoài phục vụ công tác quản lý nhà nước và chỉ đạo, điều hành trong lĩnh vực giáo dục như: dữ liệu dân cư; dữ liệu hộ tịch, dữ liệu đất đai, dữ liệu y tế; dữ liệu lao động – việc làm; dữ liệu thống kê; dữ liệu đăng ký doanh nghiệp (đối với các cơ sở giáo dục ngoài công lập); dữ liệu khoa học – công nghệ; dữ liệu địa lý – hành chính; và các nguồn dữ liệu khác phục vụ công tác phân tích, dự báo và hoạch định chính sách giáo dục. Việc khai

thác, kết nối, chia sẻ và sử dụng dữ liệu từ các cơ sở dữ liệu quốc gia, cơ sở dữ liệu chuyên ngành và các hệ thống thông tin không thuộc phạm vi quản lý của Bộ Giáo dục và Đào tạo phải tuân thủ đúng mục đích, phạm vi và thẩm quyền được giao theo quy định của pháp luật. Không thực hiện sao chép hoặc lưu trữ tập trung dữ liệu từ các hệ thống này khi chưa có căn cứ pháp lý hoặc yêu cầu quản lý nhà nước theo quy định. Hoạt động đối soát, xác thực và khai thác dữ liệu được ưu tiên thực hiện thông qua các phương thức kết nối, chia sẻ dữ liệu dùng chung, giao diện lập trình ứng dụng (API), định danh điện tử, mã hóa, ghi vết truy cập và các thỏa thuận chia sẻ dữ liệu giữa các cơ quan, tổ chức có liên quan.

Việc kết nối, chia sẻ và khai thác dữ liệu từ các cơ sở dữ liệu quốc gia, cơ sở dữ liệu chuyên ngành phải tuân thủ quy định tại Nghị định số 278/2025/NĐ-CP ngày 22 tháng 10 năm 2025 của Chính phủ về kết nối, chia sẻ dữ liệu bắt buộc giữa các cơ quan thuộc hệ thống chính trị và các văn bản quy phạm pháp luật khác có liên quan.

3.2. Phạm vi tổ chức

Kiến trúc dữ liệu được áp dụng đối với các đơn vị thuộc và trực thuộc Bộ Giáo dục và Đào tạo; các cơ quan, tổ chức được giao thực hiện nhiệm vụ quản lý nhà nước về giáo dục; các cơ sở giáo dục trong phạm vi quản lý của Bộ và các tổ chức, đơn vị có liên quan trong quá trình tạo lập, quản lý, khai thác và chia sẻ dữ liệu giáo dục.

Các cơ quan, tổ chức, đơn vị khi xây dựng mới, nâng cấp, mở rộng các hệ thống thông tin, cơ sở dữ liệu hoặc triển khai các hoạt động có liên quan đến dữ liệu giáo dục có trách nhiệm tuân thủ các yêu cầu, nguyên tắc và tiêu chuẩn được quy định trong Kiến trúc dữ liệu của Bộ Giáo dục và Đào tạo.

3.3. Phạm vi hệ thống

Kiến trúc dữ liệu áp dụng đối với toàn bộ các hệ thống thông tin, cơ sở dữ liệu, nền tảng số đang vận hành, được nâng cấp, mở rộng hoặc đầu tư xây dựng mới thuộc phạm vi quản lý của Bộ Giáo dục và Đào tạo.

Kiến trúc dữ liệu là căn cứ để:

- Lập kế hoạch đầu tư, thiết kế kỹ thuật, lựa chọn giải pháp công nghệ;
- Thẩm định, phê duyệt các dự án, nhiệm vụ ứng dụng công nghệ thông tin, chuyển đổi số;
- Tổ chức kết nối, tích hợp, chia sẻ và khai thác dữ liệu giữa các hệ thống thông tin trong và ngoài Bộ.

II. XÂY DỰNG KIẾN TRÚC DỮ LIỆU

1. Nguyên tắc và yêu cầu kiến trúc

Các nguyên tắc thiết kế dưới đây nhằm bảo đảm Kiến trúc dữ liệu Bộ Giáo dục và Đào tạo duy trì được tính nhất quán, linh hoạt, khả mở rộng và tuân thủ chuẩn, đồng thời phù hợp với định hướng phát triển Chính phủ số, Chuyển đổi số ngành giáo dục và Hệ sinh thái dữ liệu quốc gia:

- Tuân thủ Khung kiến trúc Chính phủ số và Khung kiến trúc dữ liệu quốc gia

Kiến trúc dữ liệu Bộ Giáo dục và Đào tạo phải được xây dựng phù hợp, thống nhất với Khung kiến trúc Chính phủ số Việt Nam, Khung kiến trúc dữ liệu quốc gia và các quy định, tiêu chuẩn về mô hình dữ liệu, tích hợp, chia sẻ dữ liệu và an toàn thông tin.

Các tiêu chuẩn kỹ thuật, danh mục dữ liệu dùng chung, định danh dữ liệu, quy tắc kết nối và chia sẻ dữ liệu phải bảo đảm tương thích với các tiêu chuẩn quốc gia và các nền tảng dùng chung của Chính phủ, tạo điều kiện thuận lợi cho việc kết nối, liên thông dữ liệu trong phạm vi toàn quốc.

- Phù hợp với Kiến trúc tổng thể của Bộ Giáo dục và Đào tạo

Kiến trúc dữ liệu là một cấu phần quan trọng trong Kiến trúc Chính phủ số Bộ Giáo dục và Đào tạo, do đó phải thống nhất, đồng bộ với kiến trúc nghiệp vụ, kiến trúc ứng dụng, kiến trúc công nghệ và kiến trúc an toàn thông tin đã được ban hành. Mọi điều chỉnh, thay đổi liên quan đến dữ liệu đều phải được đánh giá tác động đối với các kiến trúc thành phần khác (và ngược lại), bảo đảm tính đồng bộ, nhất quán trong toàn ngành giáo dục.

- Tập trung, chuẩn hóa và dùng chung

Dữ liệu được quản lý theo nguyên tắc “tập trung quản trị, phân tán tạo lập, liên thông khai thác”, ưu tiên xây dựng và quản lý thống nhất dữ liệu chủ, dữ liệu danh mục dùng chung phục vụ toàn ngành giáo dục. Các tập dữ liệu chủ và dữ liệu dùng chung quan trọng của cơ quan quản lý, tổ chức, cơ sở giáo dục, cán bộ, nhà giáo, người học và hồ sơ học tập suốt đời phải được quản lý tập trung, chuẩn hóa và sử dụng bởi nhiều hệ thống thông tin khác nhau, nhằm giảm trùng lặp, bảo đảm tính thống nhất và nâng cao hiệu quả khai thác dữ liệu.

- Phân quyền, phân cấp và phân lớp dữ liệu

Kiến trúc dữ liệu được thiết kế nhằm hỗ trợ mô hình phân quyền, phân cấp hợp lý trong việc tạo lập, cập nhật và quản lý dữ liệu trên toàn ngành giáo dục. Theo đó, dữ liệu chủ (master data) được phép tạo lập và cập nhật trên các hệ thống cơ sở dữ liệu chuyên ngành do Bộ Giáo dục và Đào tạo quản lý. Sau đó toàn bộ dữ liệu chủ được hợp nhất trên CSDL quốc gia về Giáo dục và Đào tạo nhằm hình thành bản ghi dữ liệu duy nhất toàn ngành rồi chia sẻ về các hệ thống chuyên ngành để thực hiện khai thác.

Dữ liệu danh mục dùng chung (Reference Data): Hệ thống mã dùng chung, bảng phân loại, chuẩn định danh và các danh mục quản lý thống nhất, được quản lý tập trung tại CSDL quốc gia về Giáo dục và Đào tạo và được tham chiếu trong nội ngành và quốc gia. Các hệ thống thông tin khác chỉ khai thác và sử dụng các dữ liệu này thông qua cơ chế kết nối, chia sẻ dữ liệu.

Đối với các loại dữ liệu phát sinh trong quá trình vận hành, việc tạo lập và cập nhật dữ liệu được thực hiện phân tán tại các đơn vị chuyên môn, cơ sở giáo dục và địa phương theo thẩm quyền và chức năng quản lý. Trong đó, chỉ các dữ liệu thuộc phạm vi quản lý thống nhất của Nền tảng quản trị số về giáo dục và đào tạo mới được đồng bộ, hợp nhất, chuẩn hóa và

quản lý tập trung tại nền tảng dữ liệu của Bộ Giáo dục và Đào tạo nhằm bảo đảm tính toàn vẹn, nhất quán và phục vụ công tác quản lý, điều hành trên phạm vi toàn ngành.

Đối với các dữ liệu phục vụ nghiệp vụ đặc thù hoặc nhu cầu quản lý riêng của địa phương, các đơn vị, địa phương được chủ động tổ chức quản lý, khai thác và cập nhật trong phạm vi hệ thống của mình; đồng thời phải tuân thủ các nguyên tắc, tiêu chuẩn và kiến trúc dữ liệu do Bộ Giáo dục và Đào tạo ban hành để bảo đảm khả năng kết nối, tích hợp và chia sẻ dữ liệu khi cần thiết.

Bên cạnh đó, dữ liệu được phân lớp rõ ràng thành các nhóm như: dữ liệu chủ, dữ liệu danh mục dùng chung, dữ liệu gốc, dữ liệu dẫn xuất, thứ cấp, dữ liệu mở, dữ liệu nhạy cảm và dữ liệu mật. Việc phân lớp này là cơ sở để xác định rõ trách nhiệm quản lý, phạm vi khai thác, chia sẻ dữ liệu cũng như các yêu cầu về bảo mật, an toàn thông tin tương ứng, phù hợp với quy định của pháp luật và các quy định chuyên ngành của Bộ Giáo dục và Đào tạo.

- Kiến trúc hướng dịch vụ và ưu tiên API (API-First & Service-Oriented)

Các ứng dụng và hệ thống thông tin thuộc phạm vi quản lý của Bộ Giáo dục và Đào tạo không truy cập trực tiếp vào cơ sở dữ liệu; mọi hoạt động trao đổi, chia sẻ dữ liệu phải được thực hiện thông qua API chuẩn hóa, có kiểm soát, do Bộ quản lý tập trung. Việc áp dụng nguyên tắc API-First giúp:

- + Nâng cao mức độ an toàn và khả năng kiểm soát truy cập dữ liệu;
- + Giảm sự phụ thuộc chặt chẽ giữa các hệ thống thông tin;
- + Tạo điều kiện thuận lợi cho việc mở rộng, nâng cấp và thay thế hệ thống trong tương lai.

- An toàn thông tin và tuân thủ pháp lý

Dữ liệu giáo dục phải được bảo vệ theo mô hình nhiều lớp, tuân thủ đầy đủ các quy định của pháp luật về an toàn thông tin mạng, bảo vệ bí mật nhà nước, bảo vệ dữ liệu cá nhân và các tiêu chuẩn, quy chuẩn quốc gia có liên quan. Mọi hoạt động truy cập, khai thác, sử dụng và chia sẻ dữ liệu phải được giám sát, kiểm soát và ghi vết đầy đủ, sẵn sàng phục vụ công tác thanh tra, kiểm tra khi cần thiết. Ưu tiên áp dụng nguyên tắc “an toàn và bảo vệ dữ liệu ngay từ khâu thiết kế” (security & privacy by design) trong toàn bộ vòng đời các hệ thống xử lý dữ liệu giáo dục.

- Dữ liệu là tài sản chiến lược của ngành giáo dục

Dữ liệu được xác định là tài sản chiến lược, đóng vai trò cốt lõi trong công tác quản lý nhà nước về giáo dục, giám sát chất lượng, phát triển đội ngũ, nâng cao hiệu quả đào tạo và xây dựng chính sách giáo dục. Vì vậy, dữ liệu phải được đầu tư, bảo vệ, quản lý và khai thác tương xứng với giá trị, đồng thời thúc đẩy hình thành văn hóa làm việc dựa trên dữ liệu trong toàn ngành giáo dục.

- Bảo đảm chất lượng dữ liệu là trách nhiệm chung

Chất lượng dữ liệu (đúng, đủ, sạch, sống, thống nhất, dùng chung) là nền tảng cho việc khai thác và sử dụng hiệu quả dữ liệu giáo dục. Việc bảo đảm chất lượng dữ liệu không chỉ

là trách nhiệm của một đơn vị hay bộ phận chuyên trách mà là trách nhiệm chung của tất cả các tổ chức, cá nhân tham gia vào quá trình tạo lập, cập nhật, quản lý và khai thác dữ liệu trong suốt vòng đời dữ liệu.

2. Kiến trúc dữ liệu hiện tại

a) Phân tích hiện trạng

Tình trạng phân tán, cát cứ dữ liệu vẫn còn phổ biến trong các đơn vị thuộc và trực thuộc Bộ Giáo dục và Đào tạo. Trong thời gian qua, nhiều hệ thống thông tin được xây dựng để phục vụ từng lĩnh vực nghiệp vụ riêng lẻ như giáo dục mầm non, giáo dục phổ thông, giáo dục thường xuyên, giáo dục đại học, khảo thí, kiểm định chất lượng giáo dục, đào tạo giáo viên, quản lý nhà giáo – cán bộ quản lý giáo dục, chuyển đổi số giáo dục, cũng như các hệ thống thông tin phục vụ công tác quản trị nội bộ của Bộ.

Mỗi lĩnh vực, cấp học hoặc chương trình thường được triển khai theo đặc thù riêng, thiếu sự phối hợp và chuẩn hóa thống nhất ở cấp Bộ, dẫn đến một số hạn chế điển hình sau:

- Dữ liệu bị phân mảnh, trùng lặp:

Cùng một đối tượng thông tin như người học, nhà giáo, cán bộ quản lý giáo dục, cơ sở giáo dục, chương trình đào tạo, ngành đào tạo, kết quả học tập, kết quả kiểm định thường được thu thập, quản lý và lưu trữ lặp lại ở nhiều hệ thống khác nhau. Mỗi hệ thống sử dụng mã định danh, cấu trúc dữ liệu và cách định nghĩa khác nhau, dẫn đến dữ liệu thiếu nhất quán, không đồng bộ và khó bảo đảm tính chính xác trên phạm vi toàn ngành. Thực trạng này gây nhiều khó khăn cho việc tổng hợp, phân tích dữ liệu xuyên cấp học, xuyên địa phương, đặc biệt trong các hoạt động cần dữ liệu tích hợp như thống kê giáo dục quốc gia, theo dõi chất lượng giáo dục, quy hoạch mạng lưới cơ sở giáo dục, đánh giá chính sách tuyển sinh hoặc dự báo nhu cầu nguồn nhân lực.

- Chất lượng dữ liệu chưa cao:

Dữ liệu trong nhiều hệ thống thông tin giáo dục chưa được quản lý và bảo đảm theo đầy đủ các tiêu chí “đúng, đủ, sạch, sống, thống nhất, dùng chung”. Việc thiếu các quy tắc và tiêu chuẩn chuẩn hóa dữ liệu thống nhất ở cấp Bộ khiến dữ liệu được nhập và cập nhật theo nhiều cách khác nhau, dẫn đến sai lệch, thiếu hụt hoặc khó đối soát. Các quy trình làm sạch dữ liệu, phân loại, cập nhật định kỳ và giám sát chất lượng dữ liệu còn chưa có quy định rõ ràng chỉ dựa trên đánh giá chủ quan, chưa được thiết lập đầy đủ, đồng bộ hoặc chưa ứng dụng các hệ thống công nghệ thông tin để quản lý dữ liệu. Một số dữ liệu không rõ nguồn gốc, thiếu lịch sử thay đổi, hoặc không được cập nhật kịp thời, làm hạn chế khả năng sử dụng dữ liệu cho phân tích, dự báo, tham mưu chính sách và chỉ đạo điều hành. Điều này đặc biệt ảnh hưởng đến các lĩnh vực có quy mô lớn và biến động nhanh như tuyển sinh, đào tạo đại học, đánh giá chất lượng giáo dục và phát triển đội ngũ nhà giáo.

- Thiếu nền tảng quản lý dữ liệu thống nhất và dùng chung:

Bộ Giáo dục và Đào tạo hiện chưa hình thành đầy đủ hệ thống quản lý dữ liệu chủ (Master Data) và dữ liệu danh mục dùng chung (Reference Data) theo hướng quản lý tập trung ở cấp Bộ, chưa phân loại dữ liệu để quản trị, quản lý dữ liệu hiệu quả và an toàn dữ liệu. Việc mỗi hệ thống tự xây dựng danh mục riêng dẫn đến xung đột mã hóa, không tương thích khi

tích hợp dữ liệu, là nguyên nhân gốc rễ gây nên tình trạng dữ liệu thiếu đồng bộ, hạn chế khả năng chia sẻ, liên thông và hình thành bức tranh tổng thể, xuyên suốt về toàn ngành giáo dục.

b) Đánh giá chung

Kiến trúc dữ liệu hiện tại của Bộ Giáo dục và Đào tạo về cơ bản vẫn là tập hợp các cơ sở dữ liệu và hệ thống thông tin chuyên ngành rời rạc, được xây dựng theo nhu cầu quản lý của từng đơn vị, từng lĩnh vực hoặc từng giai đoạn, chưa tuân thủ một mô hình kiến trúc dữ liệu tổng thể thống nhất cho toàn Bộ.

Hệ thống dữ liệu hiện nay còn bộc lộ nhiều bất cập, thể hiện ở các điểm sau:

- Chưa hình thành được bức tranh tổng quan về dữ liệu giáo dục của Bộ, bao gồm: danh mục dữ liệu chủ và dữ liệu dùng chung (như dữ liệu người học, nhà giáo, cơ sở giáo dục, kết quả học tập, chương trình đào tạo...); dữ liệu tích hợp từ các cơ sở dữ liệu dùng chung quốc gia; dữ liệu tổng hợp và phân tích phục vụ chỉ đạo điều hành; và dữ liệu mở phục vụ xã hội. Các lớp dữ liệu quan trọng chưa được phân loại, chuẩn hóa và mô hình hóa thống nhất.

- Chưa xác định rõ mô hình điều phối dữ liệu giữa dữ liệu dùng chung của Bộ Giáo dục và Đào tạo với dữ liệu dùng chung quốc gia; giữa dữ liệu dùng chung cấp Bộ với dữ liệu dùng riêng tại từng cấp học, từng lĩnh vực chuyên ngành (như mầm non, phổ thông, đại học, kiểm định, khảo thí). Việc thiếu cơ chế định danh, chuẩn hóa và đồng bộ dữ liệu xuyên suốt gây khó khăn trong quản lý, chia sẻ và khai thác dữ liệu.

Những hạn chế nêu trên cho thấy các điểm yếu mang tính hệ thống, bao gồm: thiếu tính thống nhất và toàn diện trong thiết kế; chưa bảo đảm khả năng liên thông, tích hợp giữa các hệ thống; và chưa có cơ chế điều phối dữ liệu một cách bài bản, khoa học. Điều này cản trở việc hình thành bức tranh dữ liệu giáo dục toàn cảnh, chính xác và kịp thời, phục vụ công tác quản lý nhà nước, hoạch định chính sách giáo dục, giám sát chất lượng và chỉ đạo điều hành của Lãnh đạo Bộ.

Vì vậy, việc xây dựng một kiến trúc dữ liệu mục tiêu mang tính chiến lược, lấy dữ liệu làm trung tâm, hướng tới sự đồng bộ, thống nhất và liên thông trong toàn ngành giáo dục là yêu cầu cấp thiết. Kiến trúc dữ liệu này sẽ tạo nền tảng vững chắc cho chuyển đổi số toàn diện trong lĩnh vực giáo dục và đào tạo, góp phần phát triển nguồn nhân lực và nâng cao chất lượng giáo dục quốc gia.

3. Kiến trúc dữ liệu mục tiêu

Trong đó, LGSP là nền tảng tích hợp, chia sẻ dữ liệu ngành Giáo dục đang được Bộ Giáo dục và Đào tạo triển khai và vận hành theo các quy định hiện hành. LDOP là định hướng nâng cấp, hoàn thiện LGSP theo Khung kiến trúc dữ liệu quốc gia, nhằm mở rộng và chuẩn hóa các năng lực điều phối, tích hợp, chia sẻ, quản trị và khai thác dữ liệu ở mức độ cao hơn.

Trong giai đoạn hiện nay, LDOP được xem là mô hình mục tiêu, còn LGSP là nền tảng hiện hữu, từng bước được nâng cấp, bổ sung chức năng để đáp ứng đầy đủ vai trò của LDOP, bảo đảm tính kế thừa, không gián đoạn vận hành và không phát sinh nền tảng song song. LGSP/LDOP hỗ trợ hai cơ chế chia sẻ dữ liệu: Chia sẻ theo yêu cầu (pull) và Chia sẻ chủ động (push).

Đồng thời, nền tảng cho phép thu thập, tích hợp dữ liệu từ các hệ thống thông tin, cơ sở dữ liệu chuyên ngành giáo dục của Bộ, các Sở Giáo dục và Đào tạo, cơ sở giáo dục, cũng như từ các cơ sở dữ liệu quốc gia. Trên cơ sở đó, hình thành các cơ sở dữ liệu dùng chung, cơ sở dữ liệu tổng hợp và kho dữ liệu phân tích, phục vụ hiệu quả công tác quản lý, điều hành giáo dục.

+ Công điều phối dữ liệu tại chỗ (AGN – Agent Node)

Công điều phối dữ liệu tại chỗ (AGN – Agent Node) là thành phần kỹ thuật then chốt, nhằm bảo đảm an toàn thông tin và tách biệt vùng dữ liệu nội bộ của Bộ Giáo dục và Đào tạo với nền tảng điều phối, chia sẻ dữ liệu quốc gia (NDOP).

AGN thực hiện vai trò là đầu mối kết nối kỹ thuật duy nhất giữa Bộ Giáo dục và Đào tạo với Trung tâm Dữ liệu Quốc gia, phục vụ việc điều phối, chia sẻ và đồng bộ dữ liệu theo đúng quy định của Nhà nước. Thành phần này bảo đảm việc trao đổi dữ liệu giữa Bộ với các bộ, ngành, địa phương và các cơ quan bên ngoài được thực hiện: đúng chuẩn kỹ thuật, đúng thẩm quyền, đúng phân quyền và có giám sát, kiểm soát, ghi vết đầy đủ.

AGN không thực hiện điều phối nghiệp vụ dữ liệu mà đóng vai trò là đầu mối kết nối bảo mật, hỗ trợ LDOP của Bộ trong việc kết nối an toàn với các nền tảng dữ liệu bên ngoài.

+ Nền tảng chia sẻ, điều phối dữ liệu (tên tiếng Anh là National Data Orchestration Platform, viết tắt là NDOP)

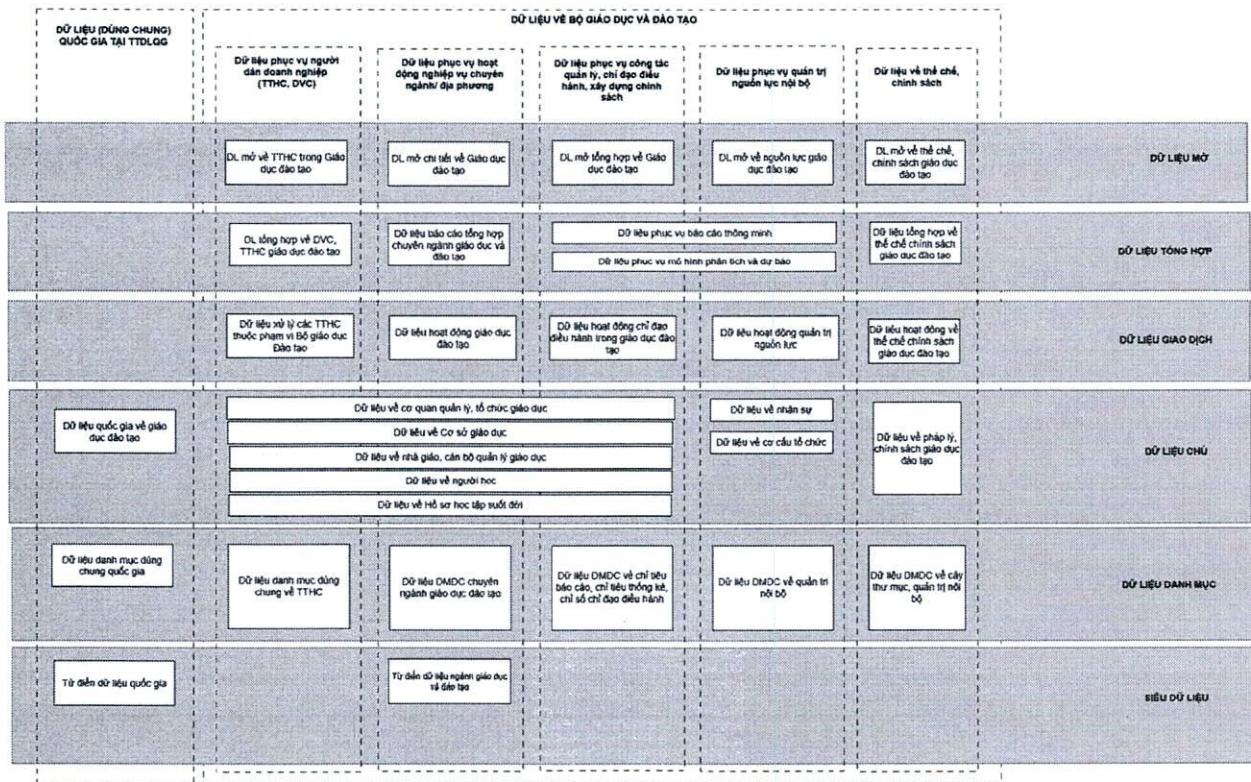
Nền tảng chia sẻ, điều phối dữ liệu quốc gia (NDOP) đóng vai trò là lớp trung gian điều phối dữ liệu ở cấp quốc gia. Trong lĩnh vực giáo dục, NDOP là kênh giao tiếp thực hiện việc trao đổi dữ liệu giữa Cơ sở dữ liệu quốc gia của Bộ Giáo dục và Đào tạo với Trung tâm dữ liệu quốc gia và các bộ, ngành khác khai thác, đồng thời hỗ trợ phân phối dữ liệu ngược trở lại cho Cơ sở dữ liệu quốc gia theo cơ chế kiểm soát tập trung.

- Lớp lưu trữ và xử lý: Lớp lưu trữ và xử lý dữ liệu là lớp lõi có vai trò tiếp nhận yêu cầu, tạo lập, lưu trữ, tích hợp và xử lý dữ liệu phục vụ khai thác, phân tích và điều hành trong toàn ngành giáo dục. Lớp này bao gồm hai thành phần chính: (i) **Nguồn dữ liệu** là các nền tảng, hệ thống thông tin và cơ sở dữ liệu tạo ra dữ liệu tại các cơ quan quản lý giáo dục, các tổ chức và cơ sở giáo dục, đây là thành phần tạo lập dữ liệu cung cấp cho CSDL quốc gia về Giáo dục và Đào tạo; và (ii) **CSDL quốc gia về Giáo dục và Đào tạo** là môi trường lưu trữ tập trung (Data Warehouse/Lakehouse) thực hiện việc hợp nhất, chuẩn hóa, làm sạch và tổ

chức dữ liệu tạo thành dữ liệu chủ và dữ liệu danh mục dùng chung duy nhất. Trên cơ sở đó, dữ liệu được sử dụng để cung cấp lại cho nguồn dữ liệu, các lớp ứng dụng, phân tích và chia sẻ để tham chiếu nhằm bảo đảm tính toàn vẹn, nhất quán của dữ liệu. Về luồng chia sẻ điều phối dữ liệu giữa các thành phần trong lớp lưu trữ và trung tâm dữ liệu quốc gia sẽ được mô tả chi tiết tại mục 3.3. *Mô hình logic luồng dữ liệu.*

- **Lớp quản trị và quản lý dữ liệu giáo dục:** Lớp quản trị và quản lý dữ liệu giáo dục đóng vai trò là lớp nền tảng về thể chế, tổ chức và quy trình vận hành, bảo đảm toàn bộ dữ liệu trong mô hình kiến trúc dữ liệu của Bộ Giáo dục và Đào tạo được quản lý thống nhất, minh bạch, an toàn và có trách nhiệm trong suốt vòng đời dữ liệu. Lớp này xác lập cơ chế quản trị dữ liệu xuyên suốt toàn ngành, làm rõ vai trò, trách nhiệm của từng cấp quản lý; đồng thời cung cấp các công cụ và quy trình để kiểm soát chất lượng, chuẩn hóa và khai thác dữ liệu hiệu quả, phục vụ công tác quản lý nhà nước và chỉ đạo điều hành. Các quy trình, nguyên tắc, tiêu chuẩn của lớp này sẽ được định nghĩa chi tiết tại Khung quản trị, quản lý dữ liệu Bộ Giáo dục Đào tạo.

3.2. Mô hình dữ liệu



Hình 2: Mô hình dữ liệu Bộ Giáo dục và Đào tạo

Mô hình dữ liệu tham chiếu của Bộ Giáo dục và Đào tạo được xây dựng nhằm khái quát hóa các miền dữ liệu phát sinh từ hoạt động nghiệp vụ, quản lý, điều hành và cung cấp dịch vụ công trong lĩnh vực giáo dục. Mô hình này thể hiện cấu trúc dữ liệu theo hướng phân miền, bảo đảm tính chuẩn hóa, khả năng tích hợp, chia sẻ và sử dụng lại dữ liệu trong toàn ngành cũng như kết nối liên thông với các hệ thống dữ liệu quốc gia.

Theo mô hình, dữ liệu của Bộ Giáo dục và Đào tạo được tổ chức thành năm miền chính như sau:

- Miền dữ liệu phục vụ người dân và doanh nghiệp, bao gồm các dữ liệu liên quan đến thủ tục hành chính và dịch vụ công trực tuyến trong lĩnh vực giáo dục và đào tạo hoặc các thủ tục hành chính có thành phần hồ sơ có thể thay thế, cắt giảm bằng dữ liệu của Bộ Giáo dục và Đào tạo. Miền dữ liệu này góp phần thực hiện nguyên tắc “Người dân chỉ cung cấp thông tin một lần”, nâng cao chất lượng phục vụ và tính minh bạch trong cung cấp dịch vụ công.

- Miền dữ liệu nghiệp vụ chuyên ngành giáo dục, là miền dữ liệu cốt lõi, phản ánh trực tiếp các hoạt động chuyên môn của ngành giáo dục. Miền này bao gồm dữ liệu về người học, nhà giáo, cán bộ quản lý, cơ sở giáo dục, cơ quan quản lý giáo dục và tổ chức khác, hồ sơ học tập suốt đời của học sinh. Đây là nguồn dữ liệu gốc phục vụ cho công tác quản lý và phân tích toàn ngành.

- Miền dữ liệu phục vụ công tác quản lý, chỉ đạo điều hành và xây dựng chính sách, bao gồm các dữ liệu tổng hợp, dữ liệu báo cáo, dữ liệu phân tích và dự báo. Miền dữ liệu này được hình thành trên cơ sở tổng hợp và xử lý dữ liệu từ CSDL quốc gia về Giáo dục và Đào tạo, nhằm hỗ trợ lãnh đạo các cấp trong việc ra quyết định, hoạch định chiến lược và xây dựng chính sách phát triển giáo dục.

- Miền dữ liệu phục vụ quản trị nội bộ, bao gồm các dữ liệu về tổ chức bộ máy, nhân sự, tài chính và các nguồn lực khác của ngành giáo dục. Miền dữ liệu này phục vụ trực tiếp cho công tác quản lý nội bộ, bảo đảm hiệu quả vận hành của các cơ quan, đơn vị trong toàn ngành.

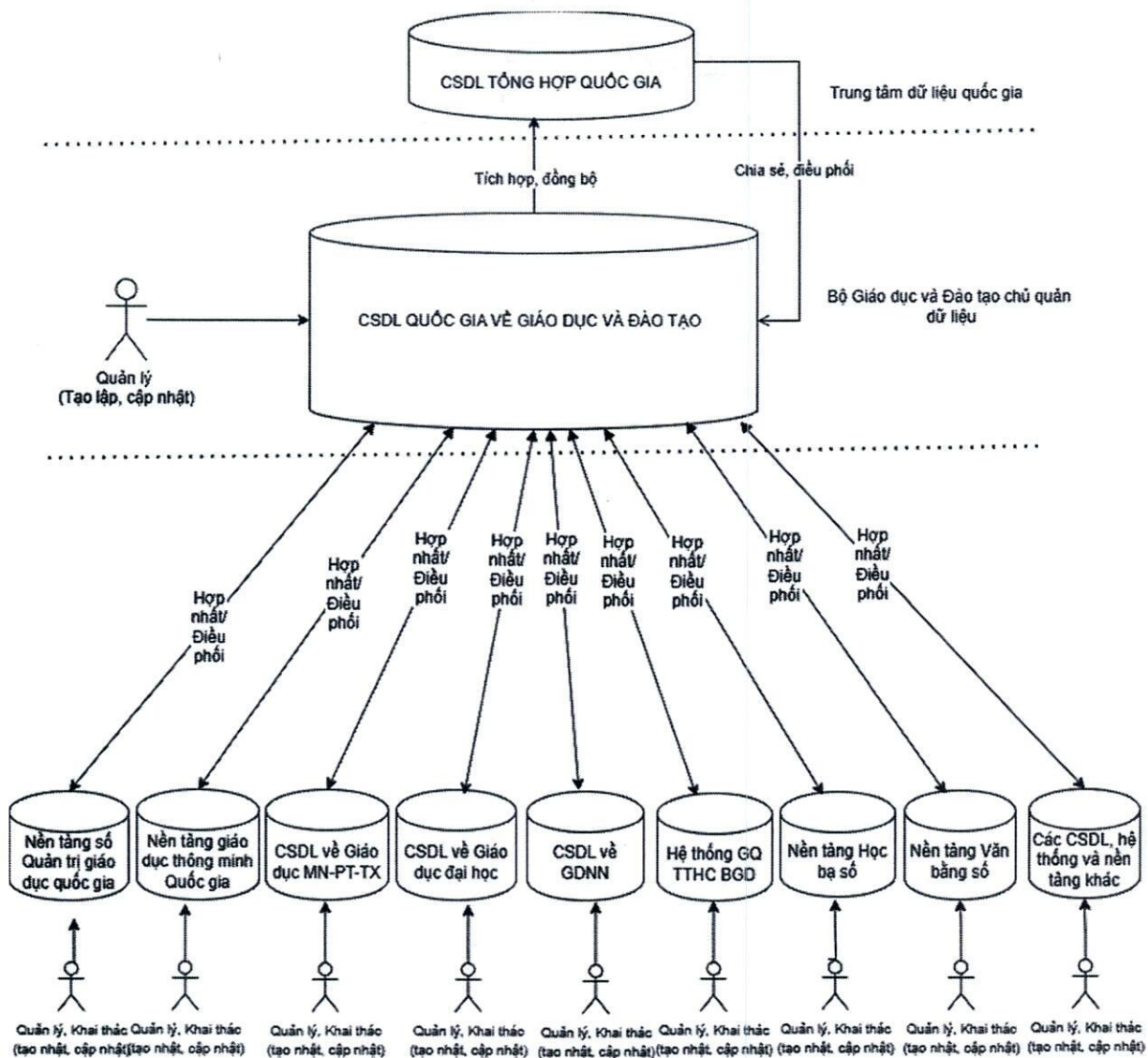
- Miền dữ liệu về thể chế và chính sách, bao gồm các dữ liệu liên quan đến văn bản quy phạm pháp luật, cơ chế, chính sách và các quy định quản lý trong lĩnh vực giáo dục và đào tạo. Đây là miền dữ liệu mang tính nền tảng, làm cơ sở pháp lý cho việc tổ chức và triển khai các hoạt động của ngành.

Dữ liệu dùng chung, dữ liệu chủ và từ điển dữ liệu ngành giáo dục sau khi được hình thành tại Bộ Giáo dục và Đào tạo sẽ được chia sẻ, cập nhật lên Cơ sở dữ liệu dùng chung quốc gia tại Trung tâm Dữ liệu Quốc gia.

Các miền dữ liệu nêu trên có mối quan hệ chặt chẽ, liên thông và bổ trợ lẫn nhau, tạo thành một hệ sinh thái dữ liệu thống nhất. Trong đó, dữ liệu nghiệp vụ là nguồn dữ liệu cốt lõi; dữ liệu tổng hợp và phân tích phục vụ công tác điều hành; dữ liệu thủ tục hành chính phục vụ người dân và doanh nghiệp; dữ liệu quản trị nội bộ bảo đảm vận hành hệ thống; và dữ liệu thể chế, chính sách đóng vai trò định hướng và kiểm soát. Cách tổ chức này phù hợp với yêu cầu của Khung kiến trúc dữ liệu quốc gia, bảo đảm khả năng mở rộng, tích hợp và phát triển Chính phủ số trong lĩnh vực giáo dục và đào tạo.

3.3. Mô hình logic luồng dữ liệu

3.3.1. Mô hình triển khai quản lý dữ liệu dùng chung



Hình 3. Mô hình quản lý dữ liệu dùng chung

Mô hình quản lý phân quyền một phần được xây dựng theo nguyên tắc kết hợp giữa quản lý dữ liệu phân tán tại nguồn và hợp nhất, điều phối tập trung tại Cơ sở dữ liệu quốc gia về Giáo dục và Đào tạo. Theo đó, các dữ liệu được tạo lập trực tiếp tại các cơ sở giáo dục thông qua Nền tảng số Quản trị giáo dục quốc gia và các nền tảng, hệ thống thông tin khác, bảo đảm dữ liệu phát sinh từ thực tiễn nghiệp vụ, hạn chế tối đa việc tổng hợp, nhập liệu lại và giảm sai lệch thông tin. Trong mô hình này, các dữ liệu nền tảng của ngành bao gồm dữ liệu chủ, dữ liệu danh mục dùng chung và siêu dữ liệu được chuẩn hóa, quản lý thống nhất và tham chiếu theo Từ điển dữ liệu dùng chung quốc gia và Từ điển dữ liệu nội ngành. Các dữ liệu này được đồng bộ tự động theo thời gian thực hoặc gần thời gian thực về Nền tảng dữ liệu tập trung của Bộ, bảo đảm tính kịp thời, nhất quán và khả năng khai thác liên thông. Bên cạnh đó, các dữ liệu chuyên ngành, đặc thù được tạo lập, cập nhật và duy trì tại các hệ thống, cơ sở dữ liệu chuyên ngành tương ứng; sau đó được hợp nhất (consolidate) và tích hợp vào cơ sở dữ liệu dùng chung khi đánh giá cần thiết. Ngược lại, các dữ liệu đã được quản lý tập

trung tại Cơ sở dữ liệu quốc gia về Giáo dục và Đào tạo sẽ được phân phối, chia sẻ cho các hệ thống, lĩnh vực có nhu cầu khai thác theo cơ chế phân quyền và kiểm soát thống nhất.

Cách tiếp cận này cho phép vừa phát huy tính chủ động của các đơn vị tạo lập dữ liệu tại nguồn, vừa bảo đảm quản lý tập trung đối với các dữ liệu cốt lõi, từ đó nâng cao chất lượng dữ liệu, giảm trùng lặp, tăng khả năng liên thông, chia sẻ và tái sử dụng dữ liệu trên toàn ngành, đồng thời phục vụ hiệu quả công tác quản lý, điều hành và ra quyết định dựa trên dữ liệu.

Cơ sở dữ liệu quốc gia về Giáo dục và Đào tạo đồng thời thực hiện vai trò cung cấp dữ liệu lên Cơ sở dữ liệu tổng hợp quốc gia thông qua Trung tâm dữ liệu quốc gia, bảo đảm dữ liệu được quản lý thống nhất, có căn cứ pháp lý, sẵn sàng phục vụ công tác quản lý nhà nước, thống kê, phân tích, dự báo và cung cấp dịch vụ công. Nguyên tắc vận hành của mô hình:

- Dữ liệu gốc được tạo lập, cập nhật trên Nền tảng số Quản trị giáo dục quốc gia và các nền tảng, hệ thống thông tin khác do các cơ sở giáo dục - nơi phát sinh nghiệp vụ thực hiện, bảo đảm đúng thẩm quyền và trách nhiệm.

- Dữ liệu đặc thù ngành được tạo lập trên các Cơ sở dữ liệu chuyên ngành, hệ thống thông tin và nền tảng khác sau đó được hợp nhất (consolidate) và tích hợp vào cơ sở dữ liệu dùng chung khi đánh giá cần thiết.

- Dữ liệu dùng chung được hợp nhất, chuẩn hóa tại CSDL quốc gia về Giáo dục và Đào tạo để hình thành dữ liệu thống nhất toàn ngành.

- Dữ liệu dùng chung sau đó điều phối, chia sẻ xuống các CSDL chuyên ngành để phục vụ khai thác nội ngành.

- Kết hợp dữ liệu dùng chung và đặc thù: Cho phép các hệ thống chuyên ngành quản lý thêm dữ liệu riêng, đồng thời tích hợp với dữ liệu dùng chung khi cần.

- Bảo đảm một nguồn dữ liệu tin cậy (Single Source of Truth) đối với các dữ liệu dùng chung.

- Dữ liệu được kiểm soát, giám sát và quản trị tập trung, bảo đảm tính chính xác, nhất quán, an toàn thông tin và tuân thủ quy định chia sẻ dữ liệu quốc gia.

Mô hình được tổ chức theo kiến trúc phân tầng, kết hợp giữa quản lý dữ liệu phân quyền tại nguồn và hợp nhất, điều phối tập trung ở cấp Bộ, với các luồng dữ liệu được kiểm soát và quản trị thống nhất, gồm ba lớp chính như sau:

- **Lớp Dữ liệu chuyên ngành** (Lớp nguồn tạo lập và khai thác):

Đây là lớp dữ liệu nền tảng, bao gồm các nền tảng, cơ sở dữ liệu chuyên ngành, hệ thống thông tin phục vụ trực tiếp hoạt động nghiệp vụ của Bộ Giáo dục và Đào tạo, như: Nền tảng số Quản trị giáo dục quốc gia, Nền tảng giáo dục thông minh quốc gia, CSDL về Giáo dục Mầm non - Phổ thông – Thường xuyên, CSDL về Giáo dục đại học, CSDL về Giáo dục nghề nghiệp, Nền tảng Văn bằng số, Nền tảng Học bạ số, Hệ thống Giải quyết TTHC Bộ Giáo dục và Đào tạo và các nền tảng, CSDL, hệ thống thông tin khác.

Theo nguyên tắc “dữ liệu dùng chung được tạo lập một lần tại nguồn”, các đơn vị chuyên môn, cơ sở giáo dục và cơ quan quản lý trực tiếp được phân quyền tạo lập, cập nhật và chịu trách nhiệm về chất lượng, tính chính xác và tính kịp thời của dữ liệu thuộc phạm vi quản lý. Các nền tảng, hệ thống thông tin, CSDL chuyên ngành cũng được tạo lập, cập nhật các dữ liệu đặc thù phục vụ lĩnh vực, chuyên môn ngành nghề và sau đó được hợp nhất (consolidate) và tích hợp vào cơ sở dữ liệu dùng chung khi đánh giá cần thiết. Các Nền tảng, CSDL chuyên ngành và hệ thống thông tin sẽ nhận dữ liệu được điều phối từ CSDL quốc gia về Giáo dục và Đào tạo để khai thác cũng như vẫn cho phép tạo lập, quản lý các dữ liệu đặc thù theo lĩnh vực, chuyên ngành.

- Lớp Dữ liệu dùng chung tập trung (Lõi hợp nhất và điều phối)

Lớp dữ liệu dùng chung tập trung là hạt nhân của mô hình, hình thành Cơ sở dữ liệu dùng chung ngành Giáo dục và Đào tạo, đóng vai trò hợp nhất, chuẩn hóa và điều phối dữ liệu toàn ngành trên cơ sở kết nối, tích hợp dữ liệu từ các hệ thống và cơ sở dữ liệu chuyên ngành.

Cơ sở dữ liệu quốc gia về Giáo dục và Đào tạo đồng thời thực hiện vai trò điều phối và chia sẻ dữ liệu hai chiều: tiếp nhận, hợp nhất dữ liệu từ các Nền tảng, CSDL và hệ thống thông tin; đồng thời cung cấp dữ liệu đã được làm sạch, chuẩn hóa, danh mục dùng chung phục vụ khai thác, sử dụng trong toàn ngành.

- Lớp Dữ liệu quốc gia (Lớp liên thông và chia sẻ quốc gia)

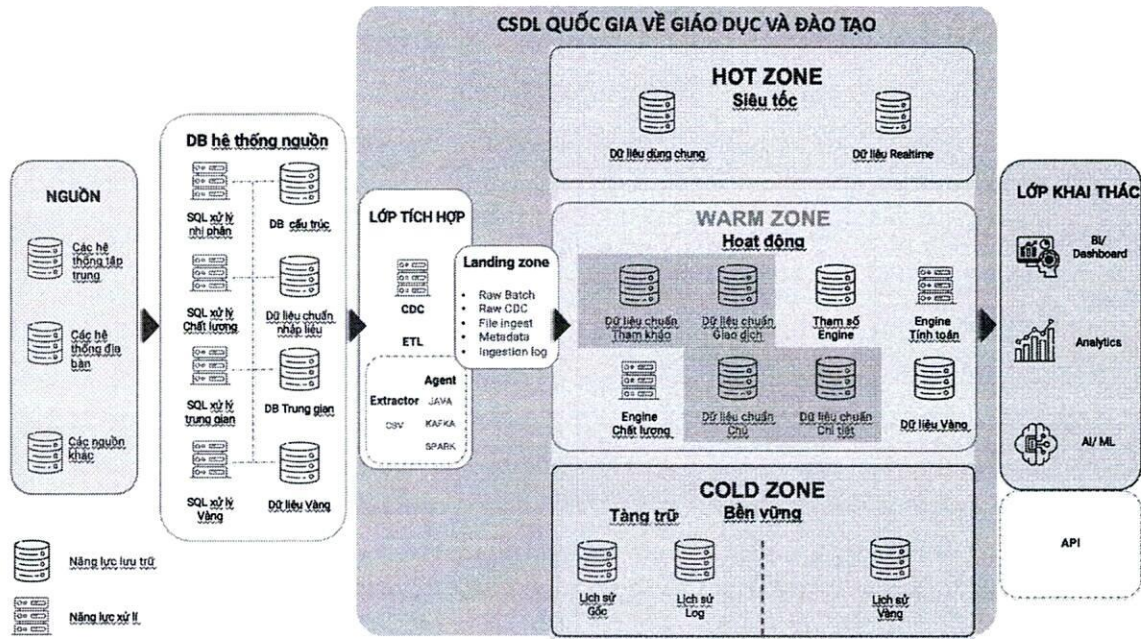
Từ CSDL quốc gia về Giáo dục và Đào tạo, các dữ liệu dùng chung quốc gia được điều phối để kết nối, chia sẻ và đồng bộ với Cơ sở dữ liệu tổng hợp quốc gia đặt tại Trung tâm Dữ liệu quốc gia, thông qua Nền tảng chia sẻ, điều phối dữ liệu quốc gia () cùng các cơ chế kỹ thuật và pháp lý liên quan.

Luồng dữ liệu này bảo đảm kiến trúc dữ liệu ngành giáo dục liên thông, tuân thủ kiến trúc dữ liệu quốc gia, phục vụ yêu cầu quản lý nhà nước liên ngành, thống kê quốc gia, phát triển nguồn nhân lực và hoạch định chính sách ở quy mô quốc gia.

3.3.2. Quy trình khai thác điều phối dữ liệu dùng chung

	Chuẩn hóa dữ liệu	CSDL Quốc gia về GD&ĐT	Tổ chức, cấu trúc dữ liệu nhằm loại bỏ trùng lặp, bảo đảm tính nhất quán, toàn vẹn và giảm thiểu sai sót.
	Lưu trữ dữ liệu và khai thác	CSDL Quốc gia về GD&ĐT	Lưu trữ dữ liệu dùng chung phục vụ tra cứu, thống kê, phân tích, chia sẻ.
3,4,5, (7)	Chia sẻ dữ liệu	CSDL Quốc gia về GD&ĐT	Tích hợp dữ liệu lên LGSP/LDOP, sau đó truyền sang NDOP (C12) thông qua điểm kết nối bảo mật AGN để cung cấp dữ liệu cho Cơ sở dữ liệu dùng chung quốc gia và chia sẻ ra bên ngoài theo quyền truy cập và phạm vi dữ liệu được xác định.
6,8,9 (7)	Phân phối dữ liệu Quốc gia	Cơ sở dữ liệu tổng hợp quốc gia	Phân phối dữ liệu từ Cơ sở dữ liệu tổng hợp quốc gia đến CSDL quốc gia về Giáo dục và Đào tạo thông qua NDOP -> AGN -> LGSP/LDOP Dữ liệu được lọc, đúng theo phân quyền, mục đích sử dụng, tính chất dữ liệu, phân loại dữ liệu.
10, 11	Chia sẻ, phân phối DL nội ngành	CSDL Quốc gia về GD&ĐT	Chia sẻ dữ liệu với các Nền tảng, CSDL chuyên ngành và các Hệ thống thông tin khác theo lĩnh vực, quyền truy cập được xác định và phạm vi dữ liệu của chuyên ngành tương ứng thông qua LGSP/LDOP.

3.4. Mô hình nền tảng dữ liệu



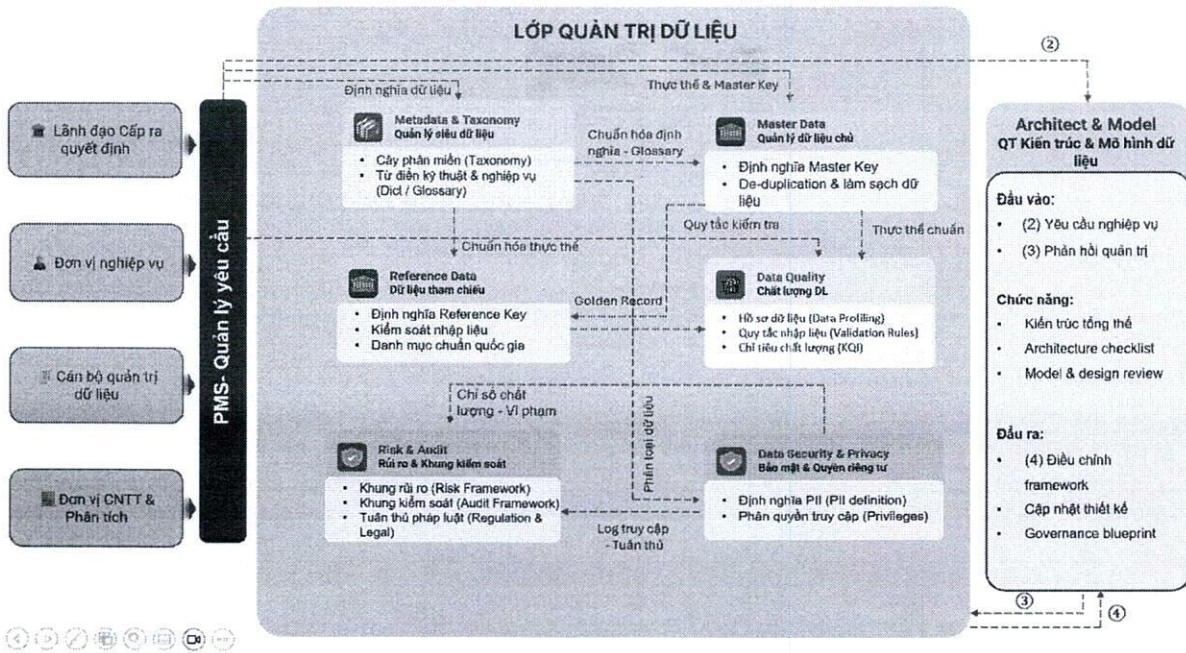
Hình 5: Mô hình nền tảng dữ liệu

Mô hình trên minh họa luồng dữ liệu chi tiết và các thành phần trong nền tảng dữ liệu của Bộ Giáo dục và Đào tạo, được thiết kế theo hướng hiện đại, tích hợp và có khả năng mở rộng. Mô hình bảo đảm đáp ứng đầy đủ các yêu cầu về quản trị, xử lý và khai thác dữ liệu, phục vụ hiệu quả công tác quản lý nhà nước về giáo dục, thúc đẩy chuyển đổi số ngành và phát triển Chính phủ số.

Đồng thời, mô hình được xây dựng bám sát các định hướng, nguyên tắc và thành phần cốt lõi quy định tại Quyết định 2439/QĐ-TTg về ban hành Khung kiến trúc dữ liệu quốc gia, Khung quản trị, quản lý dữ liệu quốc gia và Từ điển dữ liệu dùng chung. Cụ thể, việc tổ chức các lớp dữ liệu từ nguồn, vận hành đến phân tích và lưu trữ lịch sử phù hợp với mô hình phân lớp dữ liệu trong khung kiến trúc; việc hình thành nguồn dữ liệu tin cậy duy nhất (Single Source of Truth) và quản lý dữ liệu chủ, dữ liệu tham chiếu bảo đảm nguyên tắc quản trị dữ liệu thống nhất; các quy trình chuẩn hóa, làm sạch, kiểm soát chất lượng và đối soát dữ liệu đáp ứng yêu cầu về quản lý vòng đời và chất lượng dữ liệu; trong khi đó, việc xây dựng danh mục dữ liệu, siêu dữ liệu và chuẩn hóa cấu trúc dữ liệu bảo đảm khả năng liên thông, kết nối và chia sẻ dữ liệu theo Từ điển dữ liệu dùng chung.

Qua đó, mô hình không chỉ đáp ứng yêu cầu kỹ thuật mà còn bảo đảm tính tuân thủ về pháp lý và chuẩn hóa ở cấp quốc gia, tạo nền tảng vững chắc cho việc kết nối, chia sẻ dữ liệu giữa ngành giáo dục với các bộ, ngành, địa phương và các hệ thống thông tin quốc gia.

Bên cạnh kiến trúc nền tảng dữ liệu được trình bày tại Hình 5, lớp quản trị dữ liệu (Hình 6) đóng vai trò là lớp điều phối trung tâm (governance layer), bảo đảm toàn bộ dữ liệu trong hệ thống được định nghĩa, kiểm soát và vận hành theo các quy tắc thống nhất xuyên suốt vòng đời dữ liệu.



Hình 6: Mô hình quản trị dữ liệu

Để bảo đảm sự thống nhất trong quá trình triển khai, mô hình nền tảng dữ liệu tổ chức theo ba tầng nhiệt độ lưu trữ (HOT – WARM – COLD), mỗi tầng tương ứng trực tiếp với một lớp kỹ thuật xử lý dữ liệu (OLTP – ODS – OLAP/Data Warehouse – Data Lake Historical Zone).

Lưu ý: Vùng tiếp nhận (Landing Zone) không thuộc một tầng nhiệt độ cố định. Đây là vùng đệm trung gian, tiếp nhận dữ liệu đầu vào từ cả lớp OLTP (qua CDC gần thời gian thực) và lớp ODS (qua quy trình ETL/ELT định kỳ), lưu trữ tạm thời từ vài tuần đến tối đa 03 tháng trước khi dữ liệu được phân loại và chuyển vào tầng COLD theo đặc thù nghiệp vụ.

3.4.1. Lớp quản trị dữ liệu

Bên cạnh kiến trúc nền tảng dữ liệu được trình bày tại Hình 5, lớp quản trị dữ liệu (Hình 6) đóng vai trò là lớp điều phối trung tâm (governance layer), bảo đảm toàn bộ dữ liệu trong hệ thống được định nghĩa, kiểm soát và vận hành theo các quy tắc thống nhất xuyên suốt vòng đời dữ liệu.

Cụ thể, luồng vận hành được thiết kế theo hướng chuyển hóa yêu cầu nghiệp vụ thành cấu trúc dữ liệu, quy tắc kiểm soát và triển khai kỹ thuật, bắt đầu từ việc quản lý yêu cầu, nơi tiếp nhận các yêu cầu từ lãnh đạo, đơn vị nghiệp vụ và các bộ phận liên quan. Các yêu cầu này được phân rã thành ba nhóm chính: định nghĩa dữ liệu (metadata), thực thể dữ liệu (master data) và các quy tắc kiểm tra dữ liệu (data quality rules).

Tại lớp quản trị dữ liệu, các cấu phần được liên kết chặt chẽ nhằm bảo đảm tính nhất quán và khả năng kiểm soát:

- Metadata & Taxonomy đóng vai trò nền tảng, xây dựng từ điển dữ liệu, chuẩn hóa định nghĩa và mô tả ngữ nghĩa dữ liệu, từ đó cung cấp ngữ cảnh cho toàn bộ hệ thống;

- Master Data và Reference Data hình thành các tập dữ liệu chuẩn, bảo đảm nguyên tắc "một nguồn dữ liệu tin cậy duy nhất" (Single Source of Truth), đồng thời cung cấp cơ sở cho việc đối soát và chuẩn hóa dữ liệu trong các hệ thống nguồn;

- Data Quality thực hiện chức năng kiểm soát chất lượng thông qua các quy tắc kiểm tra, hồ sơ hóa dữ liệu, giúp phát hiện và xử lý sai lệch dữ liệu;

- Data Security & Privacy và Risk & Audit bảo đảm việc phân loại dữ liệu, kiểm soát truy cập, ghi nhận nhật ký và tuân thủ các quy định pháp lý, tạo thành lớp kiểm soát xuyên suốt hệ thống.

Các kết quả đầu ra từ lớp quản trị dữ liệu, bao gồm mô hình dữ liệu chuẩn, quy tắc kiểm tra, chính sách bảo mật và khung kiểm soát, được chuyển tiếp sang lớp Kiến trúc và Mô hình dữ liệu (Architecture & Model) để triển khai trên nền tảng dữ liệu thực tế.

Trong mối quan hệ với kiến trúc nền tảng dữ liệu, lớp quản trị dữ liệu đóng vai trò định nghĩa và kiểm soát các quy tắc vận hành cho từng tầng dữ liệu:

- Tại tầng HOT (OLTP), các quy tắc về metadata và phân loại dữ liệu được áp dụng để ghi nhận nguồn gốc, cấu trúc và đặc điểm dữ liệu ngay từ giai đoạn tiếp nhận;

- Tại tầng WARM (ODS), các quy tắc về chuẩn hóa, làm sạch, đối soát dữ liệu và quản lý dữ liệu chủ được triển khai, bảo đảm dữ liệu đạt chất lượng trước khi đưa vào khai thác;

- Tại tầng COLD (OLAP/Data Warehouse), các yêu cầu về lưu trữ lịch sử, truy vết dữ liệu (data lineage) và kiểm toán được thực hiện nhằm bảo đảm tính toàn vẹn và khả năng kiểm tra dữ liệu theo thời gian; row-level security và giới hạn xuất dữ liệu cá nhân trực tiếp được áp dụng để bảo đảm an toàn thông tin; khu vực tàng trữ (Data Lake Historical Zone), dữ liệu phải được ẩn danh hóa theo kỹ thuật k-anonymity hoặc differential privacy trước khi cấp quyền truy cập cho mục đích nghiên cứu và huấn luyện AI/ML; toàn bộ hoạt động truy cập phải được Data Owner phê duyệt trước và ghi nhật ký đầy đủ;

- Tại lớp khai thác, dữ liệu đã được chuẩn hóa và kiểm soát chất lượng được cung cấp cho các hệ thống báo cáo, phân tích và trí tuệ nhân tạo, bảo đảm độ tin cậy và tính nhất quán.

Cách tiếp cận này hoàn toàn phù hợp với các nguyên tắc trong Quyết định 2439/QĐ-TTg, đặc biệt là:

- Nguyên tắc quản lý dữ liệu theo vòng đời, từ thu thập, xử lý đến khai thác và lưu trữ;
- Nguyên tắc chuẩn hóa và thống nhất dữ liệu, thông qua việc xây dựng từ điển dữ liệu và danh mục dữ liệu dùng chung;
- Nguyên tắc bảo đảm chất lượng và tính toàn vẹn dữ liệu, thông qua các cơ chế kiểm soát và đánh giá chất lượng;
- Nguyên tắc bảo mật và chia sẻ dữ liệu có kiểm soát, bảo đảm cân bằng giữa khai thác và bảo vệ dữ liệu.

Nhờ đó, mô hình không chỉ bảo đảm tính hiệu quả trong vận hành kỹ thuật mà còn đáp ứng đầy đủ các yêu cầu về quản trị dữ liệu ở cấp quốc gia, tạo nền tảng cho việc kết nối, liên thông và chia sẻ dữ liệu giữa các hệ thống trong và ngoài ngành giáo dục.

3.4.2. Các lớp xử lý dữ liệu

a) Lớp nguồn và xử lý giao dịch dữ liệu (OLTP – Tầng HOT)

Đây là lớp khởi phát dữ liệu, phản ánh trực tiếp các hoạt động nghiệp vụ của ngành giáo dục và đào tạo. Người dùng thực hiện thao tác trên các hệ thống nguồn dữ liệu. Dữ liệu phát sinh được ghi nhận ban đầu dưới dạng dữ liệu gốc (raw data) và lưu trữ trong các hệ thống giao dịch.

Tại lớp này, dữ liệu được xử lý sơ bộ thông qua các công cụ chuyển đổi (data transformation tools) nhằm thực hiện các bước chuẩn hóa ban đầu, bao gồm làm sạch dữ liệu (data cleansing), kiểm tra tính hợp lệ (data validation), loại bỏ trùng lặp (deduplication) và chuẩn hóa cấu trúc (data standardization). Sau đó, dữ liệu được ghi nhận vào cơ sở dữ liệu giao dịch có cấu trúc (structured transactional databases), phục vụ các nhu cầu vận hành nghiệp vụ ngắn hạn.

Bên cạnh luồng xử lý chính, cơ chế thu thập thay đổi dữ liệu (CDC – Change Data Capture) được áp dụng nhằm trích xuất dữ liệu gần thời gian thực (near real-time), phục vụ các nhu cầu báo cáo nhanh, phân tích tức thời và đồng bộ dữ liệu tới các hệ thống phía sau, đồng thời hạn chế tác động đến hiệu năng của hệ thống giao dịch.

Thời gian lưu trữ:

Về thời gian lưu trữ, dữ liệu tại tầng HOT được giữ tối đa 12 tháng kể từ ngày phát sinh. Sau thời hạn này, dữ liệu chỉ tiết được purge sang tầng WARM (ODS) thông qua quy trình ETL/ELT có kiểm soát, nhằm duy trì hiệu năng xử lý giao dịch và tối ưu tài nguyên hạ tầng. Thời gian lưu trữ cụ thể có thể được điều chỉnh linh hoạt tùy theo đặc thù nghiệp vụ, tần suất truy cập và yêu cầu khai thác dữ liệu của từng hệ thống.

Quyền truy cập và bảo mật:

Về quyền truy cập và bảo mật, tầng HOT áp dụng kiểm soát truy cập theo vai trò (RBAC) phân cấp theo đơn vị chủ quản; xác thực đa yếu tố (MFA) bắt buộc; ghi nhật ký kiểm toán (audit log) đầy đủ cho mọi thao tác; mã hóa dữ liệu khi truyền tải (TLS 1.3) và khi lưu trữ (AES-256); nghiêm cấm xuất dữ liệu ra ngoài hệ thống theo phương thức tự phát.

b) Lớp vận hành dữ liệu (ODS – Tầng WARM)

Lớp này đóng vai trò trung gian giữa hệ thống giao dịch (OLTP) và hệ thống phân tích (OLAP), thực hiện chức năng hợp nhất và chuẩn hóa dữ liệu ở mức vận hành. Dữ liệu từ các hệ thống giao dịch và dữ liệu gốc được nạp định kỳ thông qua các quy trình ETL/ELT vào cơ sở dữ liệu vận hành (ODS). Tại đây, dữ liệu tiếp tục được xử lý thông qua các công cụ xử lý dữ liệu nhằm bảo đảm chất lượng dữ liệu (data quality), tính nhất quán (data consistency) và khả năng đối soát (data reconciliation).

Các quy trình xử lý bao gồm chuẩn hóa nâng cao (advanced data standardization), kiểm tra và áp dụng quy tắc nghiệp vụ (business rule validation), tích hợp dữ liệu (data integration) từ nhiều nguồn, kiểm soát trùng lặp (deduplication) và bảo đảm tính toàn vẹn dữ liệu (data integrity). Kết quả của quá trình này là hình thành bộ dữ liệu chuẩn, đóng vai trò là nguồn dữ liệu tin cậy duy nhất (Single Source of Truth – SSOT) trong ngữ cảnh vận hành.

Thời gian lưu trữ:

Về thời gian lưu trữ, dữ liệu tại tầng WARM được lưu giữ từ 12 tháng đến tối đa 24 tháng kể từ ngày tiếp nhận từ tầng HOT, nhằm đáp ứng nhu cầu tra cứu, tổng hợp và đối soát dữ liệu gần thời gian thực. Sau thời hạn này, dữ liệu lịch sử được chuyển dịch sang tầng COLD (OLAP/Data Warehouse). Dữ liệu chủ (master data) và dữ liệu danh mục dùng chung không áp dụng chu kỳ purge theo thời gian mà được quản lý theo phiên bản.

Quyền truy cập và bảo mật:

Về quyền truy cập và bảo mật, tầng WARM được truy cập bởi Data Steward, DBA và kỹ sư pipeline; cán bộ báo cáo vận hành được cấp quyền đọc (read-only) thông qua giao diện được kiểm soát. Dữ liệu cá nhân (PII) bắt buộc phải được che giấu (data masking) khi hiển thị cho nhóm người dùng không có thẩm quyền xem trực tiếp. Toàn bộ luồng xử lý ETL/ELT phải được ghi nhật ký lineage đầy đủ.

c) Lớp phân tích và lưu trữ lịch sử (OLAP/Data Warehouse – Tầng COLD)

Lớp này phục vụ các nhu cầu phân tích dữ liệu, báo cáo tổng hợp và khai thác dữ liệu nâng cao phục vụ công tác chỉ đạo, điều hành và hoạch định chính sách. Dữ liệu sau khi được xử lý và chuẩn hóa tại tầng WARM được tổ chức lưu trữ theo hai hướng chính: (1) lưu trữ trong các cơ sở dữ liệu lịch sử (historical databases) phục vụ phân tích xu hướng và báo cáo dài hạn; (2) nạp vào kho dữ liệu (Data Warehouse) và tổ chức theo các mô hình phân tích (data mart, dimensional model) để phục vụ hệ thống báo cáo thông minh (BI), bảng điều khiển (dashboard) và các ứng dụng phân tích nâng cao.

Kiến trúc tại lớp này cho phép tách biệt giữa dữ liệu phục vụ vận hành và dữ liệu phục vụ phân tích (workload separation), qua đó tối ưu hiệu năng hệ thống và bảo đảm khả năng mở rộng khi khối lượng dữ liệu gia tăng.

Thời gian lưu trữ:

Về thời gian lưu trữ, dữ liệu tại tầng COLD được lưu giữ từ 03 năm đến 10 năm tính từ năm học hoặc năm tài chính liên quan, nhằm phục vụ phân tích xu hướng giáo dục dài hạn, thống kê, dự báo và hoạch định chính sách. Sau thời hạn này, dữ liệu được migrate sang Khu vực tàng trữ (Data Lake Historical Zone) để bảo đảm lưu trữ lâu dài với chi phí tối ưu.

Quyền truy cập và bảo mật:

Về quyền truy cập và bảo mật, tầng COLD được cấp quyền đọc (read-only) cho các nhà phân tích dữ liệu (BI analyst), cán bộ thống kê và lãnh đạo các cấp. Dữ liệu cá nhân không được xuất trực tiếp; mọi báo cáo phải ở dạng tổng hợp, thống kê. Kiểm soát truy cập theo hàng (row-level security) được áp dụng để phân quyền theo đơn vị quản lý. Sao lưu tối thiểu 03 bản sao tại không ít hơn 02 vị trí vật lý; kiểm toán định kỳ hàng năm.

d) Lớp Data Lake và tích hợp mở rộng (Data Lake Historical Zone – Khu vực tàng trữ)

Song song với các lớp xử lý chính, hệ thống triển khai kiến trúc Data Lake nhằm lưu trữ dữ liệu ở quy mô lớn, hỗ trợ đa dạng định dạng (structured, semi-structured, unstructured data). Khu vực tầng trữ tương ứng với vùng lưu trữ lịch sử (Historical Zone) của Data Lake, khác biệt hoàn toàn về mục đích và cơ chế truy cập so với tầng COLD.

Khu vực tầng trữ lưu trữ dữ liệu trong toàn bộ vòng đời hoặc theo yêu cầu pháp lý về lưu trữ, phục vụ ba nhóm mục đích chính: (1) huấn luyện mô hình trí tuệ nhân tạo (AI/ML) và nghiên cứu học máy; (2) phân tích chuyên sâu đa năm học phục vụ nghiên cứu khoa học và đánh giá chính sách giáo dục; (3) tuân thủ nghĩa vụ lưu trữ theo quy định pháp luật hiện hành. Việc kết hợp giữa các hệ quản trị cơ sở dữ liệu có cấu trúc (RDBMS) và Data Lake hình thành kiến trúc lai (hybrid data architecture), vừa bảo đảm hiệu năng cho các hệ thống nghiệp vụ, vừa đáp ứng yêu cầu mở rộng và khai thác dữ liệu lớn trong dài hạn.

Thời gian lưu trữ:

Về thời gian lưu trữ, dữ liệu tại khu vực tầng trữ được lưu giữ không giới hạn thời gian hoặc theo thời hạn tối thiểu do pháp luật quy định. Cơ chế lưu trữ bất biến (immutable storage) được áp dụng để bảo đảm tính toàn vẹn và không thể sửa đổi dữ liệu sau khi lưu trữ.

Quyền truy cập và bảo mật:

Về quyền truy cập và bảo mật, khu vực tầng trữ áp dụng kiểm soát truy cập nghiêm ngặt nhất trong bốn tầng: dữ liệu cá nhân bắt buộc phải được ẩn danh hóa trước khi cung cấp cho bất kỳ mục đích nghiên cứu hoặc huấn luyện AI/ML nào; mọi yêu cầu truy cập phải được Data Owner phê duyệt bằng văn bản trước khi thực hiện; môi trường truy cập phải là môi trường thử nghiệm có kiểm soát, không cho phép xuất dữ liệu thô ra ngoài. Toàn bộ hoạt động truy cập phải được ghi nhật ký kiểm toán đầy đủ và lưu giữ tối thiểu 05 năm.

Tổng thể, mô hình nền tảng dữ liệu được thiết kế theo chu trình xử lý dữ liệu khép kín, bảo đảm dòng chảy dữ liệu xuyên suốt từ khâu phát sinh tại các hệ thống nguồn, qua các bước thu thập, chuẩn hóa, tích hợp, xử lý đến lưu trữ và khai thác. Các luồng dữ liệu được tổ chức linh hoạt, kết hợp giữa xử lý theo lô (batch processing) và xử lý gần thời gian thực (near real-time via CDC), đáp ứng đồng thời yêu cầu vận hành nghiệp vụ và nhu cầu phân tích, báo cáo.

Việc triển khai mô hình góp phần chuyển đổi căn bản từ phương thức quản lý dữ liệu phân tán sang quản lý tập trung, trong đó dữ liệu được chuẩn hóa, kiểm soát chất lượng và

quản trị thống nhất trong toàn ngành giáo dục. Sự phân tầng rõ ràng giữa HOT – WARM – COLD không chỉ tối ưu hóa chi phí hạ tầng mà còn tạo ra khuôn khổ kiểm soát truy cập và bảo mật phù hợp với từng mức độ nhạy cảm và tần suất khai thác của dữ liệu. Qua đó, mô hình nền tảng dữ liệu tạo cơ sở vững chắc để thực hiện chuyển đổi số toàn diện ngành giáo dục, hướng tới xây dựng nền quản trị hiện đại, hiệu quả và dựa trên dữ liệu.

4. Kiến trúc an toàn dữ liệu 8 lớp

Mô hình kiến trúc An toàn thông tin 8 lớp là mô hình phân tầng để bảo vệ toàn diện hệ sinh thái số của Bộ Giáo dục và Đào tạo. Mỗi lớp tập trung vào một lĩnh vực cụ thể, tạo thành các lớp phòng thủ kế tiếp nhau, từ hạ tầng vật lý đến các chính sách quản trị.

Lớp	Tên Lớp	Mục tiêu chính
1	Bảo vệ Dữ liệu gốc	Đảm bảo dữ liệu được chuẩn hóa, phân loại, gắn metadata và quản lý vòng đời đầy đủ ngay từ khi tạo lập.
2	Phân loại và Gán nhãn Dữ liệu	Phân loại dữ liệu theo mức độ nhạy cảm (công khai, nội bộ, cá nhân, nhạy cảm) để áp dụng chính sách bảo vệ phù hợp.
3	Quản lý truy cập và mã hóa dữ liệu.	Kiểm soát truy cập theo vai trò, định danh; mã hóa dữ liệu trong lưu trữ và truyền tải nhằm ngăn chặn truy cập trái phép.
4	Che giấu dữ liệu, tính toán an toàn và truy vết.	Áp dụng masking, tokenization; ghi nhật ký và truy vết toàn bộ hoạt động truy cập và khai thác dữ liệu.
5	Giám sát và đánh giá	Theo dõi, phát hiện, cảnh báo sớm các rủi ro, sự cố; đánh giá mức độ an toàn và tuân thủ.
6	Kiểm soát pháp lý và quản lý vòng đời dữ liệu.	Đảm bảo tuân thủ quy định pháp luật; quản lý lưu trữ, chia sẻ, hủy dữ liệu theo đúng quy định
7	Công cụ kỹ thuật	Triển khai các giải pháp kỹ thuật như DLP, SIEM, IAM, mã hóa, sao lưu, phục hồi để bảo vệ dữ liệu.
8	Báo cáo và đánh giá an toàn dữ liệu định kỳ	Thực hiện báo cáo, kiểm toán, đánh giá định kỳ và cải tiến liên tục hệ thống an toàn dữ liệu.

4.1. Lớp Dữ liệu

- Mục tiêu:

Đảm bảo dữ liệu ngành giáo dục được quản lý thống nhất, chuẩn hóa, có khả năng liên thông và sử dụng lại, đồng thời bảo đảm tính toàn vẹn, chính xác và nhất quán trong toàn bộ vòng đời dữ liệu.

- Giải pháp:

Triển khai chuẩn hóa dữ liệu trên toàn ngành giáo dục theo các bộ danh mục dùng chung, đảm bảo thống nhất về cấu trúc, định dạng và ý nghĩa dữ liệu giữa các hệ thống. Xây dựng và vận hành hệ thống từ điển dữ liệu (Data Dictionary) và metadata nhằm mô tả đầy đủ các thuộc tính dữ liệu, phục vụ việc quản lý, tích hợp và khai thác hiệu quả.

Thiết lập cơ chế quản lý vòng đời dữ liệu (Data Lifecycle Management), bao gồm các giai đoạn tạo lập, cập nhật, lưu trữ, khai thác và hủy dữ liệu, gắn với trách nhiệm cụ thể của từng đơn vị. Áp dụng các quy trình kiểm soát chất lượng dữ liệu (Data Quality) như kiểm tra tính đầy đủ, chính xác, nhất quán và kịp thời, đặc biệt tại các điểm nhập liệu từ hệ thống trường học và địa phương.

Tăng cường cơ chế đồng bộ và đối soát dữ liệu giữa các hệ thống thông tin trong ngành, đảm bảo dữ liệu được cập nhật kịp thời, hạn chế trùng lặp và sai lệch.

4.2. Lớp Phân loại và gán nhãn dữ liệu

- Mục tiêu:

Xác định mức độ quan trọng và nhạy cảm của dữ liệu nhằm áp dụng các chính sách bảo vệ phù hợp, đảm bảo cân bằng giữa khai thác và bảo mật.

- Giải pháp:

Xây dựng khung phân loại dữ liệu thống nhất toàn ngành giáo dục, dựa trên các tiêu chí về mức độ nhạy cảm, phạm vi sử dụng và yêu cầu bảo mật. Các nhóm dữ liệu cần được phân loại rõ ràng như dữ liệu công khai, dữ liệu nội bộ, dữ liệu cá nhân và dữ liệu nhạy cảm.

Triển khai cơ chế gán nhãn dữ liệu tự động dựa trên nội dung, nguồn dữ liệu và mục đích sử dụng, kết hợp với kiểm soát thủ công đối với các dữ liệu quan trọng. Việc gán nhãn cần được tích hợp vào các hệ thống dữ liệu và nền tảng tích hợp để đảm bảo dữ liệu luôn được kiểm soát theo đúng cấp độ bảo mật.

Xây dựng các chính sách kiểm soát truy cập, chia sẻ và khai thác dữ liệu tương ứng với từng mức phân loại, đồng thời áp dụng các biện pháp kỹ thuật nhằm ngăn chặn việc sử dụng dữ liệu sai mục đích hoặc vượt quyền hạn.

4.3. Lớp Quản lý truy cập và mã hóa dữ liệu

- Mục tiêu:

Đảm bảo dữ liệu chỉ được truy cập bởi các đối tượng có thẩm quyền, đồng thời bảo vệ dữ liệu khỏi các nguy cơ truy cập trái phép trong quá trình lưu trữ và truyền tải.

- Giải pháp:

Triển khai hệ thống quản lý định danh và truy cập thống nhất toàn ngành, cho phép kiểm soát truy cập theo vai trò phù hợp với mô hình tổ chức nhiều cấp.

Áp dụng xác thực đa yếu tố (MFA) đối với các tài khoản quản trị và truy cập dữ liệu quan trọng.

Thực hiện mã hóa dữ liệu toàn diện, bao gồm mã hóa dữ liệu khi lưu trữ (data at rest) và khi truyền tải (data in transit), sử dụng các tiêu chuẩn mã hóa hiện đại. Xây dựng hệ thống quản lý khóa mã hóa tập trung (KMS) nhằm đảm bảo việc tạo, lưu trữ và sử dụng khóa được kiểm soát chặt chẽ.

Đồng thời, thiết lập cơ chế kiểm soát và giám sát truy cập dữ liệu theo thời gian thực, nhằm phát hiện và ngăn chặn kịp thời các hành vi truy cập trái phép.

4.4. Lớp Che giấu dữ liệu và truy vết

- Mục tiêu:

Bảo vệ dữ liệu nhạy cảm trong quá trình khai thác và đảm bảo khả năng theo dõi, truy vết toàn bộ hoạt động liên quan đến dữ liệu.

- Giải pháp:

Áp dụng các kỹ thuật bảo vệ dữ liệu nhạy cảm như che giấu dữ liệu (data masking), mã hóa một phần, tokenization đối với các thông tin cá nhân như mã định danh, địa chỉ, thông tin liên hệ. Các kỹ thuật này cần được áp dụng linh hoạt tùy theo mục đích sử dụng dữ liệu (phân tích, báo cáo, chia sẻ).

Thiết lập hệ thống ghi nhật ký (logging) và theo dõi toàn bộ hoạt động truy cập, chỉnh sửa, khai thác dữ liệu. Các nhật ký cần được lưu trữ tập trung, đảm bảo tính toàn vẹn và không thể chỉnh sửa.

Xây dựng cơ chế truy vết (audit trail) cho phép kiểm tra lại toàn bộ lịch sử sử dụng dữ liệu, phục vụ công tác kiểm tra, thanh tra và xử lý sự cố. Đồng thời, áp dụng các công cụ phân tích để phát hiện hành vi bất thường trong truy cập dữ liệu.

4.5. Lớp Giám sát và đánh giá an toàn dữ liệu

- Mục tiêu:

Đảm bảo khả năng giám sát liên tục việc truy cập, khai thác và sử dụng dữ liệu; phát hiện kịp thời các dấu hiệu bất thường, rủi ro mất an toàn dữ liệu; đồng thời đánh giá định kỳ mức độ tuân thủ và hiệu quả của các biện pháp bảo vệ dữ liệu trong toàn ngành giáo dục.

- Giải pháp:

Thiết lập cơ chế giám sát tập trung đối với toàn bộ hoạt động truy cập, khai thác và chia sẻ dữ liệu trên các hệ thống thông tin ngành giáo dục. Các hoạt động này phải được ghi nhận, tổng hợp và phân tích nhằm phát hiện các hành vi bất thường như truy cập vượt quyền, khai thác dữ liệu ngoài phạm vi cho phép hoặc truy cập với tần suất bất thường.

Xây dựng hệ thống thu thập và phân tích nhật ký dữ liệu (data audit log), bảo đảm khả năng theo dõi xuyên suốt từ nguồn dữ liệu, nền tảng tích hợp đến các hệ thống khai thác. Việc giám sát cần tập trung vào luồng dữ liệu, đối tượng truy cập, mục đích sử dụng và phạm vi khai thác dữ liệu.

Áp dụng các phương pháp phân tích và đánh giá rủi ro đối với dữ liệu, bao gồm xác định các điểm có nguy cơ mất an toàn, đánh giá mức độ ảnh hưởng và khả năng xảy ra sự cố, từ đó đưa ra các biện pháp kiểm soát phù hợp.

Thiết lập cơ chế cảnh báo sớm đối với các hành vi có dấu hiệu vi phạm chính sách dữ liệu hoặc tiềm ẩn nguy cơ rò rỉ, lạm dụng dữ liệu; đồng thời xây dựng quy trình xử lý, khắc phục và báo cáo sự cố liên quan đến an toàn dữ liệu.

Thực hiện đánh giá định kỳ về an toàn dữ liệu, bao gồm đánh giá tuân thủ các quy định về phân loại dữ liệu, kiểm soát truy cập, chia sẻ dữ liệu và bảo vệ dữ liệu cá nhân. Kết quả đánh giá được sử dụng để điều chỉnh chính sách, hoàn thiện quy trình và nâng cao hiệu quả quản lý, bảo vệ dữ liệu.

Tăng cường công tác kiểm tra, giám sát của cơ quan quản lý cấp Bộ đối với các đơn vị trực thuộc và địa phương trong việc thực hiện các quy định về an toàn dữ liệu, bảo đảm việc triển khai thống nhất và đồng bộ trên toàn ngành.

4.6. Lớp Kiểm soát pháp lý và vòng đời dữ liệu

- Mục tiêu:

Đảm bảo việc quản lý và khai thác dữ liệu tuân thủ đầy đủ các quy định pháp luật, đồng thời kiểm soát dữ liệu theo toàn bộ vòng đời.

- Giải pháp:

Xây dựng và ban hành các chính sách, quy định nội bộ về quản lý và bảo vệ dữ liệu, đảm bảo phù hợp với các quy định pháp luật hiện hành như Luật An ninh mạng, Luật Bảo vệ dữ liệu cá nhân và các văn bản liên quan.

Thiết lập quy trình quản lý dữ liệu theo vòng đời, quy định rõ thời gian lưu trữ, điều kiện chia sẻ và tiêu chí hủy dữ liệu. Kiểm soát chặt chẽ việc chia sẻ dữ liệu giữa các hệ thống, các đơn vị trong và ngoài ngành.

Phân định rõ trách nhiệm của các đơn vị, cá nhân trong việc quản lý dữ liệu; thiết lập cơ chế giám sát và xử lý vi phạm liên quan đến sử dụng dữ liệu.

4.7. Lớp Công cụ kỹ thuật bảo mật

- Mục tiêu:

Cung cấp các công cụ và giải pháp công nghệ nhằm thực thi hiệu quả các chính sách bảo mật và bảo vệ dữ liệu.

- Giải pháp:

Triển khai đồng bộ các giải pháp công nghệ phục vụ bảo vệ dữ liệu, bao gồm hệ thống quản lý định danh và truy cập, hệ thống phòng chống rò rỉ dữ liệu (DLP), hệ thống giám sát và phân tích an ninh (SIEM).

Thiết lập các giải pháp bảo vệ hạ tầng dữ liệu như tường lửa, hệ thống phát hiện và ngăn chặn xâm nhập (IDS/IPS), bảo vệ ứng dụng và API. Đồng thời, triển khai cơ chế sao lưu và phục hồi dữ liệu (backup & disaster recovery) nhằm đảm bảo tính liên tục của hệ thống.

Tăng cường khả năng tích hợp và liên thông giữa các công cụ bảo mật với nền tảng dữ liệu và các hệ thống nghiệp vụ.

4.8. Lớp Báo cáo và đánh giá định kỳ

- Mục tiêu:

Đánh giá tổng thể hiệu quả của hệ thống an toàn dữ liệu và đảm bảo việc cải tiến liên tục.

- Giải pháp:

Thiết lập cơ chế báo cáo định kỳ về tình trạng an toàn dữ liệu ở các cấp (Bộ, Sở, đơn vị), bao gồm các chỉ số về rủi ro, sự cố và mức độ tuân thủ.

Tổ chức kiểm toán và đánh giá độc lập hệ thống an toàn dữ liệu theo các tiêu chuẩn quốc tế như ISO/IEC 27001, NIST. Thực hiện rà soát, cập nhật các chính sách, quy trình bảo mật dựa trên kết quả đánh giá.

Xây dựng cơ chế cải tiến liên tục, trong đó các bài học kinh nghiệm từ sự cố và đánh giá được đưa vào quy trình nâng cấp hệ thống, đảm bảo kiến trúc an toàn dữ liệu luôn thích ứng với các mối đe dọa mới.

III. NHIỆM VỤ VÀ GIẢI PHÁP

1. Giai đoạn 1 (2026-2027)

Xây dựng nền tảng và thiết lập vận hành

Giai đoạn này tập trung xây dựng nền móng về thể chế, dữ liệu và hạ tầng công nghệ nhằm hình thành Cơ sở dữ liệu quốc gia về Giáo dục và Đào tạo, bảo đảm khả năng quản trị, kết nối, chia sẻ và tái sử dụng dữ liệu trong toàn ngành và chia sẻ dữ liệu với quốc gia, ngành khác.

Nhiệm vụ trọng tâm:

- Hoàn thiện cơ sở pháp lý, khung quản trị, quản lý dữ liệu thống nhất trong toàn ngành giáo dục;
- Xây dựng Từ điển dữ liệu phiên bản 1.0.0;
- Xây dựng hệ thống CSDL quốc gia về Giáo dục và Đào tạo để quản trị, quản lý dữ liệu danh mục dùng chung;
- Thiết lập cơ chế tích hợp, kết nối và chia sẻ dữ liệu với các hệ thống nội bộ, địa phương, các cơ sở dữ liệu quốc gia và cơ sở dữ liệu chuyên ngành liên quan;
- Từng bước công khai, tái sử dụng dữ liệu phục vụ giải quyết thủ tục hành chính.
- Rà soát, đánh giá hiện trạng các hệ thống và cơ sở dữ liệu hiện có để xây dựng lộ trình tái cấu trúc theo kiến trúc dữ liệu thống nhất.

2. Giai đoạn 2 (2027-2030)

Mở rộng, tối ưu và khai thác dữ liệu thông minh

Giai đoạn này tập trung mở rộng phạm vi dữ liệu, hoàn thiện liên thông toàn diện và nâng cao hiệu quả khai thác dữ liệu phục vụ quản lý, điều hành và hoạch định chính sách.

Nhiệm vụ trọng tâm:

- Hoàn thiện liên thông toàn diện giữa hệ thống dữ liệu cấp Bộ, địa phương, cơ sở giáo dục và các cơ sở dữ liệu quốc gia, chuyên ngành;
- Mở rộng phạm vi dữ liệu bao phủ toàn bộ các lĩnh vực giáo dục và đào tạo;
- Tiếp tục đẩy mạnh tái sử dụng dữ liệu số để cắt giảm thành phần hồ sơ giấy trong thủ tục hành chính;
- Tiếp tục ban hành, hoàn thiện Từ điển dữ liệu các phiên bản;
- Xây dựng kho dữ liệu phân tích, báo cáo điều hành và hệ thống hỗ trợ ra quyết định;
- Từng bước ứng dụng trí tuệ nhân tạo, phân tích dự báo và cảnh báo sớm phục vụ quản lý ngành.

IV. KINH PHÍ THỰC HIỆN

Nguồn kinh phí thực hiện được bảo đảm từ các nguồn sau:

1. Kinh phí thực hiện được bố trí từ nguồn ngân sách nhà nước theo phân cấp ngân sách hiện hành, vốn lồng ghép và các nguồn vốn hợp pháp khác (nếu có).
2. Huy động các nguồn kinh phí hợp pháp khác (nếu có)

V. TỔ CHỨC THỰC HIỆN

1. Cục Khoa học, Công nghệ và Thông tin (KHCNTT) đóng vai trò là cơ quan điều phối dữ liệu thống nhất của Bộ Giáo dục và Đào tạo:

a) Chủ trì, phối hợp với các đơn vị duy trì và tổ chức triển khai Kiến trúc dữ liệu ngành giáo dục được phê duyệt tại Quyết định này; cập nhật, điều chỉnh Kiến trúc dữ liệu ngành phù hợp với điều kiện thực tế trong từng thời kỳ;

b) Chủ trì đầu tư, xây dựng và quản trị, vận hành kỹ thuật, bảo đảm an ninh mạng, an toàn thông tin các Hệ thống thông tin, Nền tảng dùng chung quốc gia/toàn ngành;

c) Chủ trì, phối hợp với các đơn vị xây dựng kế hoạch và chi tiết hóa lộ trình phát triển các miền dữ liệu của ngành; Kế hoạch phối hợp với các đơn vị cung cấp dữ liệu, sử dụng dữ liệu trong việc kết xuất, tích hợp, bổ sung, chỉnh sửa Cơ sở dữ liệu quốc gia về Giáo dục và Đào tạo, cũng như hướng dẫn, điều phối các hoạt động truy cập và sử dụng dữ liệu;

d) Hướng dẫn, kiểm tra các đơn vị trong việc xây dựng, triển khai các Cơ sở dữ liệu và hệ thống thông tin đảm bảo tính nhất quán của Kiến trúc dữ liệu ngành Giáo dục được ban hành tại Quyết định này;

đ) Tổng hợp, trình cấp có thẩm quyền bố trí vốn đầu tư công, kinh phí chi thường xuyên trong dự toán ngân sách nhà nước hoặc nguồn thu hợp pháp khác để triển khai các dự án, nhiệm vụ ứng dụng công nghệ thông tin phù hợp với kiến trúc, lộ trình phát triển của Kiến trúc dữ liệu ngành Giáo dục và quy định của pháp luật có liên quan.

2. Các Cục/Vụ chuyên môn

- a) Tuân thủ chặt chẽ Kiến trúc dữ liệu, Khung quản trị, quản lý dữ liệu được phê duyệt tại Quyết định này;
 - b) Thực hiện việc tích hợp, cung cấp dữ liệu nguồn một cách đầy đủ, chính xác, kịp thời theo quy định;
 - c) Chủ động, phối hợp với Cục Khoa học Công nghệ thông tin để xây dựng, triển khai, phát triển, nâng cấp các ứng dụng, hệ thống thông tin, cơ sở dữ liệu chuyên ngành thuộc phạm vi quản lý của đơn vị; liên thông, tích hợp, đồng bộ, chia sẻ dữ liệu nguồn với CSDL quốc gia về Giáo dục và Đào tạo.
- 

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**KHUNG QUẢN TRỊ, QUẢN LÝ DỮ LIỆU
BỘ GIÁO DỤC VÀ ĐÀO TẠO
(PHIÊN BẢN 1.0)**

*(Kèm theo Quyết định số 1866/QĐ-BGDĐT ngày 29 tháng 6 năm 2026
của Bộ trưởng Bộ Giáo dục và Đào tạo)*

NĂM 2026

MỤC LỤC

I. HIỆN TRẠNG VÀ MỤC TIÊU	4
1. Hiện trạng	4
2. Mục tiêu	4
II. PHẠM VI ĐIỀU CHỈNH	5
III. ĐỐI TƯỢNG ÁP DỤNG	5
IV. KHUNG QUẢN TRỊ, QUẢN LÝ DỮ LIỆU GIÁO DỤC VÀ ĐÀO TẠO..	7
1. Lớp quản trị dữ liệu.....	7
1.1. Mô hình quản trị dữ liệu	7
1.2. Chiến lược dữ liệu, Quy định, quy trình và tiêu chuẩn.....	17
1.3. Giám sát các hoạt động dữ liệu	18
1.4. Xây dựng văn hóa dữ liệu.....	19
2. Lớp quản lý dữ liệu	19
2.1. Quản lý Chất lượng dữ liệu	19
2.2. Quản lý dữ liệu chủ và dữ liệu danh mục dùng chung	22
2.2.1. Quản lý dữ liệu chủ.....	22
2.2.2. Quản lý dữ liệu danh mục dùng chung	23
2.2.3. Quản lý Từ điển dữ liệu.....	24
2.2.4. Quản lý quyền riêng tư và an toàn bảo mật dữ liệu	25
1. Quản lý quyền riêng tư dữ liệu.....	25
2. Quản lý an toàn bảo mật dữ liệu.....	26
3. Bảo vệ dữ liệu cốt lõi, dữ liệu quan trọng:	33
2.2.5. Quản lý kiến trúc dữ liệu	34
2.2.6. Quản lý điều phối, chia sẻ, liên thông dữ liệu.....	35
2.2.7. Quản lý nền tảng dữ liệu.....	35
2.2.8. Quản lý giám sát đo lường dữ liệu.....	36
2.2.9. Quản lý kho dữ liệu và phân tích dữ liệu.....	45
2.2.10. Quản lý các dự án CNTT.....	46
2.2.11. Quản lý khai thác gia tăng giá trị dữ liệu	47
2.2.12. Quản lý dịch vụ dữ liệu	47
3. Lớp vận hành dữ liệu	48
3.1. Vận hành vòng đời dữ liệu	48
3.2. Hạ tầng kỹ thuật quản trị dữ liệu.....	48

1. Cục Khoa học, Công nghệ và Thông tin.....	50
2. Các Cục, Vụ chuyên môn thuộc Bộ Giáo dục và Đào tạo.....	50
3. Các Sở Giáo dục và Đào tạo tỉnh, thành phố trực thuộc Trung ương.....	51
4. Các Viện và Trường Đại học, Cao đẳng thuộc Bộ Giáo dục và Đào tạo	51
PHỤ LỤC I.....	53
1. Danh sách thành viên Ban Chỉ đạo về dữ liệu Bộ Giáo dục và Đào tạo	53
2. Phân công vai trò quản trị dữ liệu Cấp 1	54
3. Phân công vai trò quản trị dữ liệu Cấp 2	55

I. HIỆN TRẠNG VÀ MỤC TIÊU

1. Hiện trạng

Thời gian qua, hệ thống cơ sở dữ liệu ngành Giáo dục và Đào tạo đã được quan tâm đầu tư, xây dựng và lưu trữ khối lượng thông tin lớn. Tuy nhiên, công tác quản trị, quản lý dữ liệu vẫn còn những tồn tại, hạn chế mang tính hệ thống, cụ thể như sau:

Một là, về thiết chế chỉ đạo và phân công trách nhiệm: Các hệ thống thông tin nghiệp vụ đã được triển khai nhưng chưa thiết lập được mô hình quản trị dữ liệu đồng bộ. Chưa có quy định cụ thể và phân định rõ ràng về vai trò, trách nhiệm của cơ quan đối với các tập dữ liệu trọng yếu (như dữ liệu cơ quan quản lý, tổ chức giáo dục, dữ liệu người học, đội ngũ/nhà giáo,...). Điều này dẫn đến sự chồng chéo, thiếu thống nhất trong giải trình và xử lý trách nhiệm khi xảy ra sai sót, xung đột dữ liệu giữa các cấp quản lý từ Địa phương đến Trung ương.

Hai là, về tiêu chuẩn hóa và kiểm soát vòng đời dữ liệu: Sự thiếu đồng bộ trong dữ liệu hiện nay chủ yếu do chưa ban hành và áp dụng thống nhất "Từ điển dữ liệu" và "Danh mục dữ liệu" dùng chung cấp Bộ có tính ràng buộc pháp lý. Nhiều dự án công nghệ thông tin được triển khai phân tán, tự định nghĩa cấu trúc dữ liệu dẫn đến khó khăn trong việc liên thông, tích hợp. Bên cạnh đó, công tác hậu kiểm, rà soát, làm sạch dữ liệu định kỳ và cơ chế lưu vết chưa được quy định chặt chẽ, tiềm ẩn rủi ro về an toàn, an ninh thông tin và bảo vệ dữ liệu cá nhân.

Ba là, về nhận thức và phương thức quản lý dữ liệu: Các hệ thống thông tin phần lớn được thiết kế để phục vụ các bài toán nghiệp vụ đơn lẻ (như khảo thí, tuyển sinh, kiểm định chất lượng) thay vì lấy dữ liệu chủ (Master Data) làm nền tảng cốt lõi. Dữ liệu chưa thực sự được nhìn nhận và quản lý như một "tài sản số" độc lập. Hạn chế này làm giảm hiệu quả của công tác liên thông, chia sẻ, tái sử dụng dữ liệu và cản trở việc khai thác giá trị gia tăng từ hệ sinh thái dữ liệu của toàn ngành.

2. Mục tiêu

Nhằm khắc phục các tồn tại hiện có và đáp ứng theo định hướng tại Quyết định số 2439/QĐ-TTg, việc sớm ban hành Khung quản trị, quản lý dữ liệu cấp Bộ là nhiệm vụ cấp bách. Quá trình này sẽ từng bước xóa bỏ tình trạng phân mảnh, chuyển đổi từ quản lý thủ công sang quản trị chủ động bằng công nghệ, coi dữ liệu giáo dục là tài sản số trọng tâm phục vụ quản lý nhà nước. Mục tiêu tổng quát của Khung quản trị, quản lý dữ liệu bao gồm:

- Thiết lập hệ thống quản trị, quản lý dữ liệu ngành Giáo dục và Đào tạo thống nhất, minh bạch và hiệu quả trên toàn ngành, bảo đảm quản trị, quản lý dữ liệu xuyên suốt từ cấp Bộ đến địa phương và cơ sở giáo dục; làm căn cứ triển khai cho các đơn vị thuộc Bộ Giáo dục và Đào tạo và các tổ chức, cá nhân có liên quan đến hoạt động quản lý, khai thác và sử dụng dữ liệu giáo dục;

- Hoàn thiện cơ chế, chính sách và quy định nội bộ ngành giáo dục nhằm: Xác lập rõ quyền quản lý, quyền khai thác và sử dụng dữ liệu; Từng bước coi dữ liệu giáo dục là tài sản quan trọng phục vụ quản lý nhà nước, đổi mới phương thức quản lý, giảng dạy và học tập; Thúc đẩy phát triển các ứng dụng, dịch vụ và hệ sinh thái giáo dục số dựa trên dữ liệu;

- Bảo đảm chất lượng, tính toàn vẹn, an toàn, bảo mật và khả năng chia sẻ của dữ liệu giáo dục, tạo điều kiện thuận lợi cho: Kết nối, liên thông và chia sẻ dữ liệu trong toàn ngành; Kết nối với các cơ sở dữ liệu quốc gia, bộ, ngành và địa phương; Khai thác, tái sử dụng dữ liệu phục vụ quản lý, điều hành, nghiên cứu và phát triển giáo dục.

II. PHẠM VI ĐIỀU CHỈNH

Khung này quy định các nguyên tắc chung, mô hình và yêu cầu cơ bản trong quản trị, quản lý dữ liệu ngành Giáo dục và Đào tạo, làm cơ sở thống nhất cho việc tổ chức các hoạt động quản trị dữ liệu trong toàn ngành. Phạm vi điều chỉnh bao gồm nhưng không giới hạn các nội dung nền tảng của quản trị dữ liệu như: siêu dữ liệu, dữ liệu chủ, chất lượng dữ liệu, an toàn và bảo mật dữ liệu, cùng các hoạt động quản trị dữ liệu có liên quan khác.

III. ĐỐI TƯỢNG ÁP DỤNG

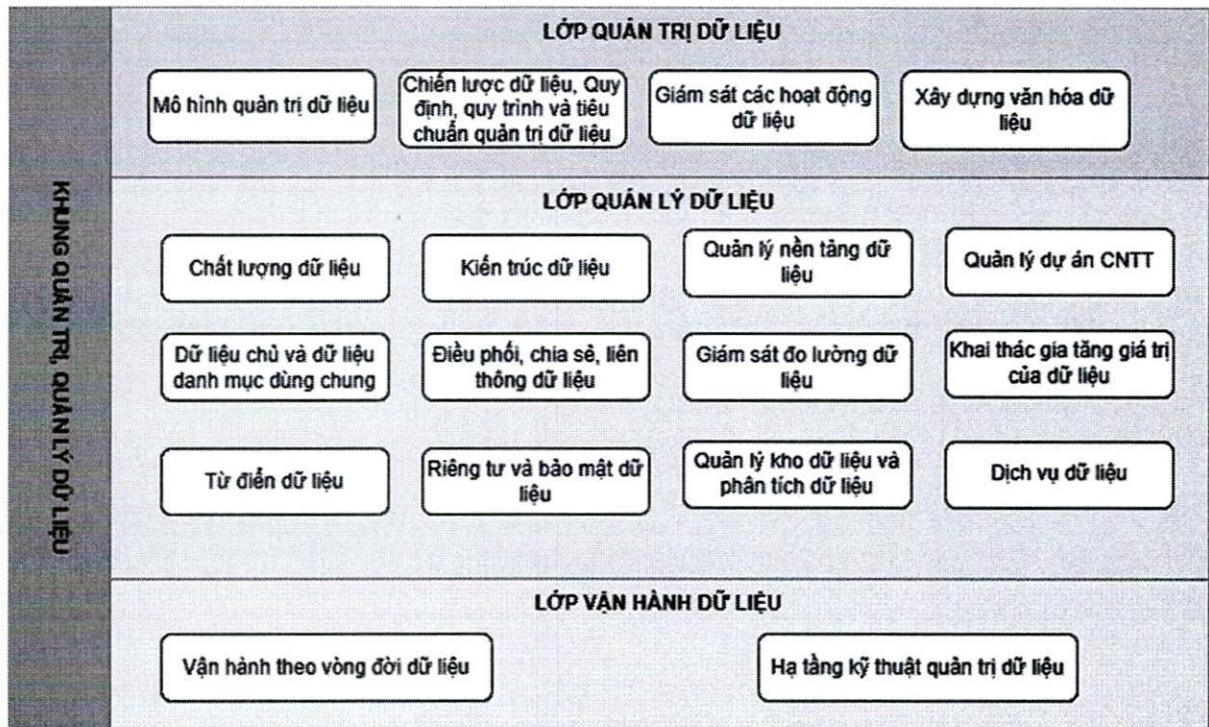
Khung quản trị, quản lý dữ liệu ngành Giáo dục và Đào tạo được áp dụng đối với các cơ quan, tổ chức, cá nhân có liên quan đến hoạt động quản lý, vận hành, khai thác và sử dụng dữ liệu giáo dục, bao gồm:

- a) Các cơ quan quản lý nhà nước về giáo dục:
 - Bộ Giáo dục và Đào tạo; các đơn vị thuộc Bộ (Cục, Vụ, Thanh tra, Văn phòng);
 - Sở Giáo dục và Đào tạo các tỉnh, thành phố trực thuộc Trung ương.
- b) Các cơ sở giáo dục và đơn vị trong hệ thống giáo dục quốc dân:
 - Các cơ sở giáo dục công lập ở các cấp học và trình độ đào tạo thuộc Bộ và Sở Giáo dục và Đào tạo quản lý; các trung tâm giáo dục thường xuyên, trung tâm bồi dưỡng công lập;
 - Các cơ sở giáo dục đại học tự chủ, cơ sở giáo dục ngoài công lập; trung tâm đào tạo ngoài công lập;
 - Các đơn vị sự nghiệp thuộc hệ thống giáo dục quốc dân.
- c) Các cơ quan, tổ chức, cá nhân có liên quan đến dữ liệu giáo dục:
 - Các tổ chức, doanh nghiệp tham gia xây dựng, vận hành, kết nối, tích hợp hệ thống dữ liệu giáo dục;
 - Các tổ chức nghiên cứu, cá nhân tham gia khai thác, sử dụng dữ liệu trong Hệ thống cơ sở dữ liệu giáo dục và đào tạo.

Mức độ và phạm vi áp dụng Khung đối với các đối tượng trên được quy định như sau:

- Áp dụng toàn diện: Các cơ quan quản lý nhà nước về giáo dục quy định tại điểm (a) và các cơ sở giáo dục công lập, đơn vị sự nghiệp công lập thuộc hệ thống giáo dục quốc dân quy định tại điểm (b) có trách nhiệm tuân thủ toàn diện các nguyên tắc, quy định và tiêu chuẩn tại Khung này trong toàn bộ hoạt động quản trị, quản lý và sử dụng dữ liệu.
- Áp dụng trong phạm vi kết nối, báo cáo và chia sẻ dữ liệu: Các cơ sở giáo dục đại học tự chủ, cơ sở giáo dục ngoài công lập, trung tâm đào tạo ngoài công lập, các tổ chức và doanh nghiệp quy định tại điểm (b) và (c) có trách nhiệm tuân thủ các tiêu chuẩn kỹ thuật, quy trình bảo mật, yêu cầu chất lượng và nghĩa vụ báo cáo theo Khung này khi thực hiện bất kỳ hoạt động nào liên quan đến kết nối, tích hợp, báo cáo hoặc chia sẻ dữ liệu với hệ thống cơ sở dữ liệu của ngành Giáo dục và Đào tạo. Yêu cầu tuân thủ Khung này là điều kiện bắt buộc được quy định trong hợp đồng, thỏa thuận cung cấp dịch vụ công nghệ thông tin với cơ quan nhà nước có thẩm quyền trong ngành.
- Áp dụng khi khai thác dữ liệu giáo dục: Các tổ chức nghiên cứu, cá nhân và đối tác trong và ngoài nước khai thác dữ liệu giáo dục, bao gồm dữ liệu mở và dữ liệu được cấp phép, có trách nhiệm tuân thủ các quy định pháp luật hiện hành về bảo vệ dữ liệu cá nhân, an toàn thông tin và sử dụng dữ liệu đúng mục đích được cấp phép; đồng thời tham chiếu các nguyên tắc, tiêu chuẩn của Khung này nhằm bảo đảm tính tương thích và an toàn trong quá trình khai thác, sử dụng dữ liệu giáo dục.

IV. KHUNG QUẢN TRỊ, QUẢN LÝ DỮ LIỆU GIÁO DỤC VÀ ĐÀO TẠO



Hình 1. Khung quản trị, quản lý dữ liệu Giáo dục và Đào tạo

1. Lớp quản trị dữ liệu

Quản trị dữ liệu bao gồm các hoạt động xây dựng chính sách, kế hoạch, chương trình, quy trình và tiêu chuẩn liên quan đến dữ liệu, được thực hiện bởi chủ sở hữu dữ liệu và chủ quản dữ liệu. Mục tiêu là bảo đảm dữ liệu được quản lý liên tục, hiệu quả, đồng thời đáp ứng các yêu cầu về tính đầy đủ, chính xác, toàn vẹn, nhất quán, thống nhất, chuẩn hóa, an toàn, bảo mật và kịp thời.

1.1. Mô hình quản trị dữ liệu

1.1.1. Nguyên tắc quản trị dữ liệu

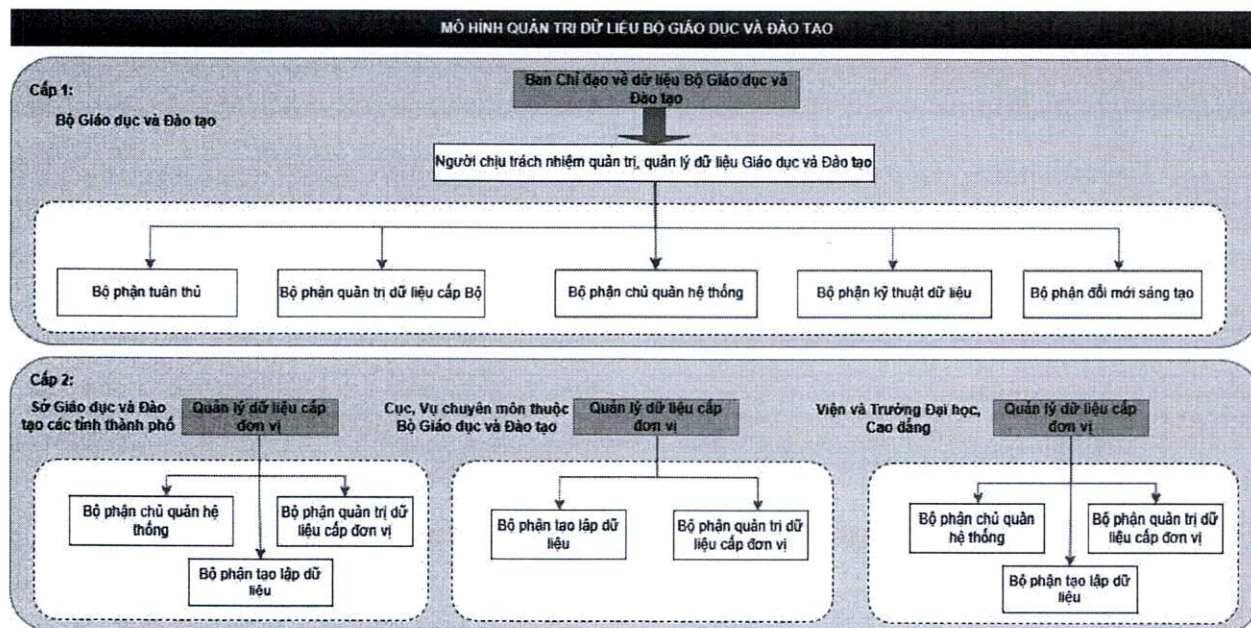
(1) Phân cấp quản trị, quản lý dữ liệu bảo đảm tính thống nhất, xuyên suốt từ trung ương đến địa phương; đồng thời phát huy tính chủ động, linh hoạt và sáng tạo của các cấp quản lý, cơ sở giáo dục trong việc vận hành các hệ thống thông tin;

(2) Bộ Giáo dục và Đào tạo ban hành quy định Khung quản trị, quản lý dữ liệu; đồng thời quản trị, quản lý tập trung các hệ thống lõi, cơ sở dữ liệu nền tảng toàn ngành và dữ liệu các lĩnh vực chuyên ngành (giáo dục mầm non, phổ thông, đại học, thường xuyên...);

(3) Sở Giáo dục và Đào tạo chịu trách nhiệm quản trị dữ liệu cấp đơn vị; thực hiện phân cấp, phối hợp quản lý dữ liệu với Ủy ban nhân dân các cấp và các cơ sở giáo dục, đào tạo trên địa bàn. Sở được quyền mở rộng mô hình quản trị dữ liệu để phục vụ bài toán quản lý đặc thù của địa phương (trong khuôn khổ tương thích với Khung quản trị, quản lý dữ liệu của Bộ), chủ động phát triển các ứng dụng, phần mềm giáo dục số bảo đảm đáp ứng tiêu chuẩn kết nối, tích hợp vào hệ thống chung của ngành;

(4) Các cơ sở giáo dục, đào tạo đóng vai trò là đơn vị tạo lập và sử dụng và đảm bảo chất lượng dữ liệu tại nguồn. Cán bộ quản lý, giáo viên, giảng viên và nhân viên nghiệp vụ tại các cơ sở này được phân quyền tạo lập, cập nhật, tra cứu và khai thác dữ liệu thuộc phạm vi quản lý của đơn vị mình theo đúng quy định về kiểm soát truy cập và bảo mật thông tin. Cơ sở giáo dục đại học được thực hiện quyền tự chủ trong tổ chức quản trị dữ liệu nội bộ theo quy định của pháp luật về giáo dục đại học; bảo đảm tuân thủ tiêu chuẩn kết nối, tích hợp với hệ thống dữ liệu chung của ngành.

1.1.2. Mô hình quản trị dữ liệu



Hình 2. Mô hình quản trị dữ liệu tại Bộ Giáo dục và Đào tạo

Mô hình quản trị dữ liệu tại Bộ Giáo dục và Đào tạo bao gồm 2 cấp:

- **Cấp 1:** Bộ Giáo dục và Đào tạo bao gồm các vai trò quản trị dữ liệu như:

+ Ban Chỉ đạo về dữ liệu Bộ Giáo dục và Đào tạo;

+ Người chịu trách nhiệm quản trị, quản lý dữ liệu tham mưu, tư vấn, thẩm định và giúp việc Ban Chỉ đạo về dữ liệu Bộ Giáo dục và Đào tạo về các nội dung liên quan đến Quản trị dữ liệu được giao thông qua:

- Bộ phận quản trị dữ liệu;
- Bộ phận chủ quản hệ thống;
- Bộ phận kỹ thuật dữ liệu;
- Bộ phận đổi mới sáng tạo.

- **Cấp 2:** Sở giáo dục đào tạo các Tỉnh/Thành phố; các Cục Vụ chuyên môn thuộc Bộ Giáo dục và Đào tạo và Viện, trường Đại học Cao đẳng thuộc Bộ Giáo dục và Đào tạo.

+ Sở Giáo dục và Đào tạo các Tỉnh/Thành phố: Quản lý dữ liệu cấp đơn vị là các Giám đốc Sở thực hiện quản lý dữ liệu của đơn vị thông qua:

- Bộ phận chủ quản hệ thống;

- Bộ phận quản trị dữ liệu cấp đơn vị;
- Bộ phận tạo lập dữ liệu.

+ Các Cục, Vụ chuyên môn:

• Quản lý dữ liệu cấp đơn vị là các Cục trưởng/Vụ trưởng thực hiện quản lý dữ liệu của đơn vị thông qua:

- Bộ phận quản trị dữ liệu cấp đơn vị;
- Bộ phận tạo lập dữ liệu.

+ Viện và trường Đại học, Cao đẳng thuộc Bộ Giáo dục và Đào tạo: Quản lý dữ liệu cấp đơn vị là các Viện trưởng hoặc Hiệu trưởng thực hiện quản lý dữ liệu của đơn vị thông qua:

- Bộ phận chủ quản hệ thống;
- Bộ phận quản trị dữ liệu cấp đơn vị;
- Bộ phận tạo lập dữ liệu.

1.1.3. Chức năng và nhiệm vụ trong Mô hình quản trị dữ liệu

- **Cấp 1:** Bộ Giáo dục và Đào tạo.

(1) Ban Chỉ đạo về dữ liệu Bộ Giáo dục và Đào tạo.

➤ Chức năng:

• Định hướng chiến lược tổng thể về dữ liệu ngành Giáo dục và Đào tạo: xây dựng tầm nhìn, mục tiêu, chiến lược và lộ trình triển khai quản trị dữ liệu, bảo đảm thống nhất với chiến lược chuyển đổi số ngành và định hướng của Ban Chỉ đạo quốc gia về dữ liệu;

• Định hướng và chỉ đạo hoạt động quản trị dữ liệu trong toàn ngành, bảo đảm dữ liệu được quản lý như một tài nguyên chiến lược, phục vụ công tác quản lý nhà nước, chỉ đạo điều hành và hoạch định chính sách giáo dục;

• Chỉ đạo thiết lập mô hình tổ chức quản trị dữ liệu; ban hành hoặc trình ban hành các quy định, quy trình, tiêu chuẩn, chính sách về dữ liệu áp dụng thống nhất trong toàn ngành;

• Chỉ đạo công tác quản trị rủi ro và tuân thủ về dữ liệu: bảo đảm an toàn, an ninh dữ liệu; tuân thủ các quy định pháp luật và yêu cầu của Ban Chỉ đạo quốc gia về dữ liệu;

• Chỉ đạo xây dựng, phát triển và duy trì văn hóa dữ liệu trong toàn ngành Giáo dục và Đào tạo.

➤ Nhiệm vụ:

• Chỉ đạo, phê duyệt việc xây dựng và triển khai các văn bản về quản trị dữ liệu (quy chế, quy định, quy trình, hướng dẫn); phê duyệt các chương trình, đề án, hệ thống, nền tảng và công cụ phục vụ quản trị dữ liệu trong ngành;

• Phê duyệt kế hoạch triển khai các hoạt động quản trị dữ liệu theo từng giai đoạn; quyết định hoặc đề xuất cấp có thẩm quyền bố trí kinh phí, nguồn lực để thực hiện;

• Chỉ đạo xây dựng, phát triển các cơ sở dữ liệu ngành Giáo dục và Đào tạo; tổ chức kết nối, tích hợp, đồng bộ, lưu trữ, chia sẻ, phân tích, khai thác và điều phối dữ liệu, bảo đảm liên thông với Cơ sở dữ liệu quốc gia và Cơ sở dữ liệu tổng hợp quốc gia;

- Giám sát, đánh giá hiệu quả hoạt động quản trị dữ liệu trên phạm vi toàn ngành; xem xét, phê duyệt các báo cáo, khuyến nghị, đề xuất xử lý đối với các vấn đề dữ liệu quan trọng hoặc liên ngành;

- Chỉ đạo, kiểm tra việc thực hiện các chủ trương, chiến lược, cơ chế, chính sách về dữ liệu tại các đơn vị thuộc và trực thuộc; bảo đảm tuân thủ khung quản trị dữ liệu thống nhất;

- Chỉ định đơn vị đầu mối và phân công trách nhiệm cho các tổ chức, cá nhân phụ trách quản trị và quản lý dữ liệu (theo mô hình chuyên trách hoặc kiêm nhiệm); chỉ đạo kiện toàn tổ chức bộ máy và phát triển nguồn nhân lực dữ liệu;

- Chỉ đạo xây dựng, lan tỏa văn hóa dữ liệu; tổ chức truyền thông, đào tạo, nâng cao nhận thức, năng lực về dữ liệu trong toàn ngành.

(2) Người chịu trách nhiệm quản trị, quản lý dữ liệu Giáo dục và Đào tạo

➤ Chức năng:

- Tham mưu, tư vấn, giúp việc Ban Chỉ đạo về dữ liệu Bộ Giáo dục và Đào tạo trong việc xây dựng, tổ chức triển khai và giám sát các nội dung liên quan đến quản trị và quản lý dữ liệu;

- Đảm bảo chất lượng dữ liệu, an toàn và bảo mật dữ liệu trong toàn ngành;
- Xác định và phân giao trách nhiệm quản trị, quản lý dữ liệu cho các đơn vị, tổ chức và cá nhân có liên quan trong ngành Giáo dục và Đào tạo.

➤ Nhiệm vụ:

- Tổ chức thực hiện mọi hoạt động quản trị dữ liệu trong phạm vi ngành Giáo dục và Đào tạo theo chỉ đạo của Ban Chỉ đạo dữ liệu và quy định của cơ quan có thẩm quyền;

- Đề xuất xây dựng, cập nhật và tổ chức thực hiện chiến lược dữ liệu ngành; đề xuất các giải pháp phát triển, khai thác và sử dụng dữ liệu phù hợp với mục tiêu quản lý, điều hành và phát triển giáo dục trong từng giai đoạn;

- Giám sát và báo cáo định kỳ, đột xuất về tình hình quản trị dữ liệu trong toàn ngành, bao gồm: Việc xây dựng và vận hành các quy định, quy trình dữ liệu; Phân công và thực hiện trách nhiệm quản trị dữ liệu; Các hoạt động quản trị dữ liệu như: Từ điển dữ liệu, Chất lượng dữ liệu, Riêng tư và an toàn bảo mật dữ liệu; Kết quả khai thác, sử dụng dữ liệu;

- Đề xuất các giải pháp nâng cao hiệu quả quản trị dữ liệu; kiến nghị xử lý các vấn đề phát sinh, rủi ro, sự cố liên quan đến dữ liệu;

- Thực hiện các công việc khác theo phân công.

(3) Bộ phận tuân thủ

➤ Chức năng:

- Tham mưu, giúp việc cho Người chịu trách nhiệm quản trị, quản lý dữ liệu thực hiện thống nhất quản lý nhà nước về dữ liệu theo quy định của pháp luật;

- Thực hiện giám sát và đánh giá việc tuân thủ đối với toàn bộ các hoạt động liên quan đến dữ liệu trong ngành, bao gồm: xây dựng, phát triển, bảo vệ, quản trị, xử lý, sử dụng và đánh giá chất lượng dữ liệu;

- Giám sát và đảm bảo các hoạt động dữ liệu được thực hiện phù hợp với quy định pháp luật về an ninh, an toàn và bảo vệ dữ liệu, đặc biệt đối với dữ liệu cá nhân, dữ liệu quan trọng và dữ liệu cốt lõi.

➤ **Nhiệm vụ:**

- Giám sát độc lập và đánh giá việc tuân thủ các quy định pháp luật, chính sách và tiêu chuẩn về dữ liệu trong toàn bộ vòng đời dữ liệu trong ngành Giáo dục và Đào tạo;
- Giám sát và đánh giá việc tuân thủ các quy định về riêng tư và an toàn bảo mật dữ liệu, đặc biệt đối với dữ liệu cá nhân, dữ liệu quan trọng và dữ liệu cốt lõi của ngành;
- Chủ trì hoặc phối hợp thực hiện công tác thanh tra, kiểm tra việc chấp hành pháp luật và quy định về dữ liệu; tham gia giải quyết khiếu nại, tố cáo và kiến nghị xử lý vi phạm liên quan đến dữ liệu theo thẩm quyền;
- Tổ chức đánh giá, giám sát việc tuân thủ đối với các hoạt động liên quan đến dữ liệu cốt lõi, dữ liệu quan trọng, dữ liệu cá nhân trong toàn ngành; kịp thời phát hiện rủi ro và đề xuất biện pháp xử lý;
- Tổng hợp, báo cáo định kỳ hoặc đột xuất về tình hình tuân thủ dữ liệu trong toàn bộ hoạt động dữ liệu của ngành lên cấp có thẩm quyền; đề xuất các giải pháp nâng cao hiệu quả quản lý, giám sát và tuân thủ dữ liệu trong ngành. Bộ phận tuân thủ thực hiện chế độ báo cáo kép: báo cáo hành chính lên Người chịu trách nhiệm quản trị, quản lý dữ liệu và báo cáo chức năng về tuân thủ lên trực tiếp Ban Chỉ đạo dữ liệu nhằm bảo đảm tính độc lập trong hoạt động giám sát;
- Thực hiện các công việc khác theo phân công.

(4) Bộ phận quản trị dữ liệu cấp Bộ

➤ **Chức năng:**

- Tham mưu, giúp việc cho Người chịu trách nhiệm quản trị, quản lý dữ liệu thiết lập và điều phối triển khai các hoạt động quản trị dữ liệu theo Khung quản trị quản lý dữ liệu Giáo dục và Đào tạo;
- Xây dựng, duy trì và quản lý các chính sách, quy trình, tiêu chuẩn dữ liệu nhằm bảo đảm dữ liệu được quản lý thống nhất, nhất quán và có thể khai thác hiệu quả;
- Điều phối, hướng dẫn các đơn vị triển khai các hoạt động quản trị dữ liệu theo Khung quản trị, quản lý dữ liệu;
- Thúc đẩy nâng cao nhận thức, năng lực và văn hóa quản trị dữ liệu trong toàn ngành.

➤ **Nhiệm vụ:**

- Hỗ trợ điều phối và vận hành mô hình quản trị dữ liệu trong toàn ngành;
- Xây dựng, cập nhật, ban hành hoặc trình ban hành các quy định, quy trình, tiêu chuẩn và hướng dẫn về quản trị dữ liệu trong ngành Giáo dục và Đào tạo; tổ chức triển khai và hướng dẫn áp dụng thống nhất trong toàn ngành;
- Chủ trì xây dựng và quản lý Từ điển dữ liệu đảm bảo quản lý tập trung, thống nhất trong toàn ngành; theo dõi, hướng dẫn và đôn đốc việc cập nhật, thay đổi;
- Tổ chức đánh giá và theo dõi chất lượng dữ liệu; phối hợp điều phối triển khai các biện pháp cải tiến, xử lý sai lệch dữ liệu theo yêu cầu quản lý;

- Phối hợp thực hiện các biện pháp bảo đảm an ninh, an toàn dữ liệu; tuân thủ các quy định về riêng tư và an toàn bảo mật dữ liệu, đặc biệt là dữ liệu cá nhân và dữ liệu quan trọng, dữ liệu cốt lõi của ngành;
- Điều phối quản lý dữ liệu chủ, dữ liệu tham chiếu, đảm bảo được cập nhật và quản lý nhất quán;
- Phối hợp với các bộ, ngành, địa phương và các đơn vị liên quan trong hoạt động chia sẻ, kết nối, khai thác dữ liệu; tổng hợp, phản ánh khó khăn, vướng mắc và đề xuất giải pháp hoàn thiện chính sách;
- Là đầu mối tổng hợp nhu cầu, đề xuất giải pháp và công cụ phục vụ quản trị dữ liệu.

(5) Bộ phận chủ quản hệ thống

> Chức năng:

- Chịu trách nhiệm quản lý, vận hành hệ thống dữ liệu; đảm bảo hệ thống hoạt động thông suốt, ổn định, an toàn và đạt hiệu quả tối ưu trong quá trình khai thác;
- Phê duyệt và giám sát việc triển khai các biện pháp bảo mật hạ tầng; chủ động ngăn chặn, xử lý các hành vi truy cập trái phép hoặc các nguy cơ gây mất an toàn hệ thống.
- Phê duyệt và kiểm soát việc phân quyền truy cập dữ liệu đối với các nền tảng ứng dụng công nghệ thuộc phạm vi quản lý;
- Đảm bảo toàn bộ hệ thống và hoạt động vận hành đáp ứng đúng các tiêu chuẩn kỹ thuật, chính sách quản lý và quy định pháp lý về dữ liệu do Bộ Giáo dục và Đào tạo ban hành.

> Nhiệm vụ:

- Phê duyệt thiết lập, vận hành, thay đổi và nâng cấp hệ thống dữ liệu theo yêu cầu nghiệp vụ và kỹ thuật;
- Phê duyệt và kiểm soát việc phân quyền truy cập hệ thống và dữ liệu;
- Phê duyệt và giám sát việc triển khai các biện pháp bảo mật, bảo đảm an toàn hệ thống và dữ liệu;
- Giám sát việc vận hành hệ thống, bảo đảm tính ổn định, sẵn sàng và hiệu quả khai thác;
- Chỉ đạo và quyết định phương án xử lý các sự cố hệ thống; giám sát việc khắc phục và khôi phục hoạt động;
- Phê duyệt kế hoạch sao lưu, khôi phục và bảo đảm tính liên tục của hệ thống;
- Phê duyệt các thay đổi liên quan đến cấu trúc dữ liệu, tích hợp, luồng dữ liệu và chức năng hệ thống;
- Giám sát việc tuân thủ các tiêu chuẩn, quy chuẩn kỹ thuật, chính sách và quy định pháp luật về dữ liệu;
- Điều phối, ra quyết định trong các vấn đề liên quan giữa các bên (quản trị dữ liệu, kỹ thuật dữ liệu, đơn vị nghiệp vụ);
- Theo dõi, đánh giá hiệu suất hệ thống; định hướng và phê duyệt các giải pháp tối ưu, nâng cấp hệ thống;
- Phê duyệt tài liệu kỹ thuật, hướng dẫn sử dụng và kế hoạch đào tạo người dùng;
- Báo cáo cấp có thẩm quyền về tình trạng vận hành, rủi ro và hiệu quả hệ thống.

(6) Bộ phận kỹ thuật dữ liệu

➤ Chức năng:

- Tổ chức triển khai, quản lý và vận hành hạ tầng kỹ thuật, hệ thống dữ liệu và nền tảng dữ liệu tập trung của Bộ Giáo dục và Đào tạo, bảo đảm dữ liệu được lưu trữ an toàn, sẵn sàng và khai thác hiệu quả;
- Triển khai các biện pháp kỹ thuật về bảo mật, kiểm soát truy cập và bảo vệ dữ liệu; phòng ngừa các rủi ro mất mát, rò rỉ hoặc truy cập trái phép;
- Thực hiện quản lý kỹ thuật đối với lưu trữ, sao lưu, khôi phục, luồng và xử lý dữ liệu; bảo đảm tính toàn vẹn và khả năng phục hồi của dữ liệu;
- Hỗ trợ kỹ thuật cho các hoạt động quản trị dữ liệu (Tư điển dữ liệu, chất lượng dữ liệu, tích hợp, khai thác dữ liệu);
- Phối hợp với các đơn vị liên quan nhằm bảo đảm hệ thống và dữ liệu đáp ứng yêu cầu nghiệp vụ, tiêu chuẩn kỹ thuật và quy định.

➤ Nhiệm vụ:

- Vận hành hệ thống lưu trữ dữ liệu, cơ sở dữ liệu và nền tảng dữ liệu;
- Thực hiện thiết lập, cấu hình và duy trì môi trường dữ liệu theo yêu cầu kỹ thuật;
- Thực hiện cấp quyền truy cập dữ liệu theo phê duyệt; kiểm soát truy cập ở mức kỹ thuật;
- Triển khai các biện pháp bảo mật kỹ thuật nhằm bảo vệ dữ liệu khỏi truy cập trái phép, mất mát hoặc rò rỉ;
- Thực hiện sao lưu dữ liệu định kỳ, vận hành phương án khôi phục và bảo đảm tính liên tục của dữ liệu;
- Giám sát luồng dữ liệu, quá trình xử lý dữ liệu và hiệu suất hệ thống dữ liệu;
- Triển khai các kiểm soát chất lượng dữ liệu trên hệ thống theo quy định;
- Quản lý, cập nhật và cung cấp siêu dữ liệu kỹ thuật phục vụ quản trị dữ liệu;
- Hỗ trợ triển khai tích hợp, chia sẻ, chuyển đổi và lưu trữ dữ liệu;
- Hỗ trợ kỹ thuật cho việc xây dựng tư điển dữ liệu, truy xuất nguồn dữ liệu và các hoạt động quản trị dữ liệu;
- Tham gia phân tích, phát hiện và xử lý các vấn đề về dữ liệu từ góc độ hệ thống và công nghệ;
- Tiếp nhận và xử lý các yêu cầu khai thác dữ liệu; hỗ trợ người dùng sử dụng dữ liệu đúng kỹ thuật;
- Phát hiện, xử lý sự cố kỹ thuật liên quan đến dữ liệu; phối hợp khắc phục theo chỉ đạo;
- Theo dõi, báo cáo tình trạng vận hành, an toàn và chất lượng dữ liệu.

(7) Bộ phận đổi mới sáng tạo

➤ Chức năng:

- Tham mưu, nghiên cứu và tổ chức triển khai các hoạt động đổi mới sáng tạo, phát triển sản phẩm và dịch vụ dữ liệu phục vụ quản lý, điều hành và cung cấp dịch vụ trong ngành Giáo dục và Đào tạo;
- Thúc đẩy phát triển và ứng dụng các mô hình, công nghệ mới trong phân tích, khai thác, chia sẻ và kinh tế dữ liệu, phù hợp với định hướng chuyển đổi số ngành và quốc gia;

- Tổ chức triển khai các hoạt động thử nghiệm có kiểm soát (sandbox), nghiên cứu khoa học công nghệ, đổi mới sáng tạo và chuyển đổi số liên quan đến dữ liệu theo quy định của pháp luật.

➤ **Nhiệm vụ:**

- Nghiên cứu, đề xuất và tổ chức triển khai các sản phẩm, dịch vụ dữ liệu trong ngành Giáo dục và Đào tạo, bao gồm: Dịch vụ trung gian dữ liệu; Phân tích, tổng hợp và khai thác dữ liệu; Xác thực, xác nhận dữ liệu; Các nền tảng, sàn giao dịch/chia sẻ dữ liệu;
- Chủ trì hoặc phối hợp xây dựng, phát triển và vận hành các nền tảng dữ liệu, công cụ phân tích, hệ thống hỗ trợ ra quyết định phục vụ quản lý và điều hành giáo dục;
- Tổ chức triển khai các chương trình thử nghiệm có kiểm soát (sandbox) đối với các mô hình, công nghệ, sản phẩm dữ liệu mới; đánh giá hiệu quả và đề xuất nhân rộng;
- Tham gia các hoạt động nghiên cứu khoa học, phát triển công nghệ, đổi mới sáng tạo và chuyển đổi số liên quan đến dữ liệu trong ngành;
- Phối hợp với các đơn vị liên quan trong việc khai thác, sử dụng và thương mại hóa dữ liệu (nếu có), bảo đảm tuân thủ quy định pháp luật;
- Theo dõi, đánh giá xu hướng công nghệ và mô hình dữ liệu mới; đề xuất áp dụng vào ngành Giáo dục và Đào tạo;
- Phối hợp với Bộ phận quản trị dữ liệu, Bộ phận tạo lập dữ liệu và Bộ phận kỹ thuật dữ liệu để bảo đảm các sản phẩm, dịch vụ dữ liệu tuân thủ tiêu chuẩn, quy chuẩn và yêu cầu kỹ thuật;
- Tổ chức đào tạo, nâng cao năng lực đổi mới sáng tạo, khai thác và sử dụng dữ liệu cho các đơn vị trong ngành;
- Báo cáo định kỳ hoặc đột xuất về tình hình triển khai các hoạt động đổi mới sáng tạo dữ liệu; đề xuất cơ chế, chính sách thúc đẩy phát triển dữ liệu;
- Thực hiện các nhiệm vụ khác theo phân công;
- Thúc đẩy hệ sinh thái đổi mới sáng tạo dựa trên dữ liệu giáo dục; hợp tác với doanh nghiệp, tổ chức nghiên cứu trong và ngoài nước.

- **Cấp 2:** Sở giáo dục đào tạo các Tỉnh/Thành phố; các Cục Vụ chuyên môn thuộc Bộ Giáo dục và Đào tạo và Viện, trường Đại học Cao đẳng thuộc Bộ Giáo dục và Đào tạo.

(1) Quản lý dữ liệu cấp đơn vị

➤ **Chức năng:**

- Chỉ đạo, phê duyệt và giám sát toàn bộ các hoạt động liên quan đến miền dữ liệu thuộc phạm vi quản lý; bảo đảm thực hiện thống nhất, đầy đủ và kịp thời các quy định, chính sách về dữ liệu.
- Chịu trách nhiệm toàn diện về chất lượng dữ liệu trên phạm vi dữ liệu được giao; bảo đảm tính chính xác, đầy đủ, nhất quán và kịp thời, đáp ứng yêu cầu quản lý, điều hành và báo cáo.
- Bảo đảm tuân thủ các chính sách, quy định về bảo mật và an toàn dữ liệu trong phạm vi đơn vị; tổ chức triển khai các biện pháp bảo vệ dữ liệu phù hợp với cấp độ phân loại của từng nhóm dữ liệu theo quy định của ngành.
- Là đầu mối phối hợp và điều phối với các đơn vị khác trong và ngoài ngành để giải quyết các vấn đề liên quan đến dữ liệu; đại diện đơn vị trong các hoạt động trao đổi, chia sẻ và liên thông dữ liệu theo phân công của cấp Bộ.

➤ **Nhiệm vụ trọng tâm:**

- Thực hiện các nhiệm vụ được Ban Chỉ đạo về dữ liệu Bộ Giáo dục và Đào tạo phân công về các hoạt động quản trị dữ liệu
- Đảm bảo nguồn lực cần thiết (nhân sự, tài chính, công cụ) để giải quyết các vấn đề liên quan đến miền dữ liệu quản lý
- Phê duyệt các hoạt động quản trị dữ liệu, các thay đổi quy trình nghiệp vụ, nguồn dữ liệu chuẩn hóa của dữ liệu thuộc phạm vi quản lý
- Tuyên bố các yêu cầu, quy tắc nghiệp vụ, ngưỡng về chất lượng dữ liệu cần phải đáp ứng
- Chỉ đạo thực hiện công tác làm sạch, làm giàu dữ liệu, các vấn đề chất lượng dữ liệu của miền dữ liệu quản lý
- Phối hợp với các đơn vị khác để giải quyết các vấn đề liên quan đến dữ liệu
- Báo cáo định kỳ lên Ban Chỉ đạo và đề xuất cải tiến
- Đào tạo nâng cao năng lực dữ liệu và xây dựng văn hóa dữ liệu

(2) Bộ phận chủ quản hệ thống

➤ **Chức năng:**

- Chịu trách nhiệm quản lý, vận hành và khai thác hiệu quả các hệ thống thông tin, cơ sở dữ liệu trong phạm vi đơn vị; bảo đảm hệ thống hoạt động ổn định, an toàn và đáp ứng yêu cầu nghiệp vụ.
- Điều phối các hoạt động liên quan đến hệ thống giữa các bộ phận trong đơn vị; là đầu mối kết nối với Bộ phận kỹ thuật dữ liệu và Bộ phận chủ quản hệ thống cấp Bộ.
- Chỉ đạo triển khai và giám sát việc thực hiện các yêu cầu về an toàn, an ninh hệ thống và dữ liệu tại đơn vị, phù hợp với quy định của ngành.

➤ **Nhiệm vụ:**

- Tổ chức vận hành, giám sát các hệ thống thông tin, cơ sở dữ liệu trong phạm vi đơn vị; bảo đảm tính ổn định, sẵn sàng và hiệu quả khai thác.
- Phê duyệt và triển khai phân quyền truy cập hệ thống và dữ liệu cho cán bộ trong đơn vị theo nguyên tắc đặc quyền tối thiểu; kiểm soát việc sử dụng và chia sẻ dữ liệu trên hệ thống theo quy định.
- Chỉ đạo triển khai các biện pháp bảo mật kỹ thuật; bảo đảm hệ thống tuân thủ các tiêu chuẩn, quy chuẩn kỹ thuật và chính sách an toàn dữ liệu của ngành.
- Phối hợp với Bộ phận quản trị dữ liệu cấp đơn vị cũng như các Bộ phận liên quan cấp Bộ để triển khai các hoạt động quản trị dữ liệu trên hệ thống thuộc phạm vi quản lý đáp ứng yêu cầu nghiệp vụ và yêu cầu chất lượng.
- Báo cáo định kỳ hoặc đột xuất về tình trạng vận hành, an toàn hệ thống và các vấn đề phát sinh; đề xuất giải pháp nâng cao hiệu quả quản lý, vận hành hệ thống tại đơn vị.
- Thực hiện các nhiệm vụ khác theo phân công.

(3) Bộ phận quản trị dữ liệu cấp đơn vị

➤ **Chức năng:**

- Đầu mỗi triển khai quản trị dữ liệu tại đơn vị; tham mưu, giúp việc cho Quản lý dữ liệu cấp đơn vị trong các hoạt động quản lý và khai thác dữ liệu thuộc phạm vi phụ trách.

- Bảo đảm dữ liệu trong đơn vị được định nghĩa rõ ràng, chuẩn hóa, nhất quán và phản ánh đúng nghiệp vụ giáo dục.

- Tham gia xây dựng, duy trì và áp dụng các quy tắc nghiệp vụ, tiêu chuẩn dữ liệu, từ điển dữ liệu và siêu dữ liệu; giám sát và cải thiện chất lượng dữ liệu, bảo đảm dữ liệu được sử dụng đúng mục đích, đúng quy trình trong phạm vi đơn vị.

➤ **Nhiệm vụ:**

- Hỗ trợ Quản lý dữ liệu cấp đơn vị trong tổ chức các hoạt động quản trị dữ liệu; điều phối các bộ phận nghiệp vụ trong đơn vị trong việc cung cấp, cập nhật và quản lý dữ liệu.

- Phối hợp với Bộ phận quản trị dữ liệu cấp Bộ để bảo đảm tuân thủ tiêu chuẩn, quy chuẩn, quy định và chính sách dữ liệu của ngành tại đơn vị.

- Thu thập, đề xuất và cập nhật định nghĩa dữ liệu, thuật ngữ nghiệp vụ, danh mục dữ liệu tại đơn vị; tham gia xây dựng và cập nhật từ điển dữ liệu, sơ đồ luồng dữ liệu.

- Xác định và đề xuất nguồn dữ liệu chuẩn, quy tắc nghiệp vụ của dữ liệu trong phạm vi đơn vị.

- Theo dõi, phát hiện và phân tích nguyên nhân các vấn đề về chất lượng dữ liệu tại đơn vị (liên quan đến quy trình, biểu mẫu, nhập liệu...); điều phối thực hiện làm sạch, chuẩn hóa dữ liệu.

- Đề xuất và cập nhật quy tắc nghiệp vụ, quy tắc dữ liệu và ngưỡng chất lượng dữ liệu; phối hợp xử lý các vấn đề chất lượng dữ liệu theo yêu cầu của các bên liên quan.

- Tham gia triển khai phân loại, gắn nhãn và bảo vệ dữ liệu theo quy định trong phạm vi nghiệp vụ của đơn vị.

- Giám sát việc tuân thủ các quy định về dữ liệu trong phạm vi đơn vị; kịp thời phát hiện và báo cáo các sai lệch, vi phạm.

- Báo cáo định kỳ hoặc đột xuất về tình trạng dữ liệu, chất lượng dữ liệu và các vấn đề phát sinh cho Quản lý dữ liệu cấp đơn vị.

- Đề xuất cải tiến quy trình nghiệp vụ, biểu mẫu và phương thức quản lý dữ liệu nhằm nâng cao chất lượng và hiệu quả sử dụng dữ liệu trong đơn vị.

- Thực hiện các công việc khác theo phân công.

(4) Bộ phận tạo lập dữ liệu

➤ **Chức năng:**

- Tổ chức, hướng dẫn, đôn đốc và giám sát toàn bộ quá trình thu thập, tạo lập, cập nhật dữ liệu do các đơn vị trực thuộc và tuyến cơ sở (phòng giáo dục, trường học, khoa, trung tâm, giáo viên, người học...) trực tiếp thực hiện.

- Đóng vai trò là đầu mối quản lý khâu đầu vào; chịu trách nhiệm trước Quản lý dữ liệu cấp đơn vị về tiến độ, độ bao phủ và tính trung thực của nguồn dữ liệu được thu thập từ cơ sở.

- Bảo đảm công tác tạo lập dữ liệu tại tuyến cơ sở tuân thủ đúng các tiêu chuẩn, quy chuẩn, từ điển dữ liệu và quy trình nghiệp vụ do ngành ban hành.

➤ **Nhiệm vụ:**

- Lập kế hoạch, xây dựng các văn bản hướng dẫn chi tiết và tổ chức triển khai công tác thu thập, nhập liệu (định kỳ hoặc đột xuất) đến các đơn vị, cá nhân trực tiếp tạo lập dữ liệu.
- Theo dõi, đôn đốc và giám sát thường xuyên tiến độ, tỷ lệ hoàn thành việc cập nhật dữ liệu của các cơ sở giáo dục, đơn vị trực thuộc trên địa bàn hoặc trong phạm vi quản lý.
- Tổ chức rà soát, đối soát bước đầu (tiền kiểm) dữ liệu được đẩy lên từ tuyến cơ sở; chủ động phát hiện các điểm bất thường và yêu cầu cơ sở kịp thời điều chỉnh, làm sạch, bổ sung dữ liệu sai lệch, thiếu hụt.
- Tiếp nhận, tổng hợp và trực tiếp hỗ trợ, giải đáp các vướng mắc về mặt nghiệp vụ nhập liệu, quy định biểu mẫu cho đội ngũ trực tiếp tạo lập dữ liệu tại cơ sở.
- Lưu trữ, quản lý hồ sơ, tài liệu hoặc các biên bản nghiệm thu dữ liệu (nếu có) từ tuyến cơ sở chuyển lên để phục vụ công tác thanh tra, hậu kiểm.
- Phối hợp chặt chẽ với Bộ phận quản trị dữ liệu trong việc triển khai áp dụng các tiêu chuẩn, quy chuẩn, danh mục và từ điển dữ liệu xuống tuyến cơ sở; phản hồi các vướng mắc từ thực tiễn nhập liệu để kịp thời tối ưu hóa biểu mẫu, quy tắc nghiệp vụ; và phối hợp xử lý dứt điểm các vấn đề chất lượng dữ liệu được phát hiện qua công tác giám sát, hậu kiểm.
- Phối hợp với Bộ phận chủ quản hệ thống để đề xuất cấp phát, quản lý tài khoản nhập liệu cho tuyến cơ sở; phối hợp xử lý các lỗi kỹ thuật để bảo đảm quá trình đồng bộ dữ liệu từ cơ sở lên hệ thống cấp 2 được thông suốt.
- Tổng hợp, báo cáo định kỳ hoặc đột xuất về tình hình thu thập, tiến độ tạo lập và các vấn đề về chất lượng dữ liệu đầu vào cho Quản lý dữ liệu cấp đơn vị.
- Chủ trì tổ chức các hội nghị, khóa tập huấn hướng dẫn nghiệp vụ nhập liệu và nâng cao nhận thức về tầm quan trọng của dữ liệu cho đội ngũ cán bộ chuyên trách tại các cơ sở.

1.1.4. Phân công các vai trò trong Mô hình quản trị dữ liệu (*Chi tiết xem thêm tại Phụ lục*)

1.2. Chiến lược dữ liệu, Quy định, quy trình và tiêu chuẩn

1.2.1. Chiến lược dữ liệu

Chiến lược dữ liệu của Bộ Giáo dục và Đào tạo xác định rõ tầm nhìn, mục tiêu và định hướng tổng thể về phát triển dữ liệu trong phạm vi toàn ngành giáo dục. Đây là cơ sở để tổ chức, chỉ đạo thống nhất các hoạt động liên quan đến dữ liệu, bao gồm: xây dựng, thu thập, chuẩn hóa, tích hợp, chia sẻ, khai thác, bảo vệ và quản trị dữ liệu giáo dục. Chiến lược hướng tới hình thành hệ sinh thái dữ liệu giáo dục đồng bộ, liên thông từ trung ương đến địa phương, phục vụ hiệu quả công tác quản lý nhà nước, nâng cao chất lượng dạy và học, thúc đẩy chuyển đổi số toàn ngành giáo dục.

1.2.2. Quy định, quy trình và tiêu chuẩn

Hệ thống quy định, quy trình và tiêu chuẩn là nền tảng pháp lý và kỹ thuật cốt lõi bảo đảm cho việc triển khai đồng bộ, thống nhất các hoạt động quản trị, vận hành, kết nối và chia sẻ dữ liệu trên toàn ngành Giáo dục. Hệ thống văn bản này cần được xây dựng, ban hành và liên tục cập nhật, bao gồm các nhóm trọng tâm sau:

- Nhóm quản lý vòng đời dữ liệu: Các quy định, quy trình bao quát toàn bộ các giai đoạn của dữ liệu từ khi thu thập/tạo lập, xử lý, lưu trữ, khai thác/chia sẻ, cho đến khi lưu trữ lịch sử (archiving) và tiêu hủy dữ liệu;

- Nhóm an toàn, bảo mật và quyền riêng tư: Các quy định, tiêu chuẩn phân loại cấp độ an toàn dữ liệu, bảo vệ dữ liệu cá nhân, quản lý quyền truy cập và các biện pháp bảo mật nhằm phòng chống rủi ro không gian mạng;

- Nhóm quản lý chất lượng dữ liệu: Các quy định, quy trình và tiêu chuẩn thiết lập tiêu chí đánh giá, đo lường ngưỡng chất lượng; hướng dẫn kiểm soát, rà soát và làm sạch dữ liệu nhằm bảo đảm tính chính xác, đầy đủ, nhất quán và kịp thời;

- Nhóm quản lý dữ liệu chủ và danh mục dùng chung: Các quy định, tiêu chuẩn định nghĩa, khởi tạo, duy trì, phân phối và đồng bộ các tập dữ liệu cốt lõi (Master Data) và hệ thống danh mục dùng chung trên toàn ngành;

- Nhóm kiến trúc và từ điển dữ liệu: Các quy định, tiêu chuẩn về kiến trúc tổng thể, mô hình dữ liệu, siêu dữ liệu (metadata) và hệ thống từ điển dữ liệu, bảo đảm tính thống nhất về ngữ nghĩa và khả năng tương tác, liên thông giữa các hệ thống;

- Các nhóm quy định, quy trình, tiêu chuẩn khác: Được chủ động nghiên cứu, bổ sung và ban hành để đáp ứng các yêu cầu nghiệp vụ mới phát sinh trong quá trình vận hành hệ sinh thái dữ liệu giáo dục.

1.3. Giám sát các hoạt động dữ liệu

Hoạt động giám sát dữ liệu nhằm theo dõi toàn diện các hoạt động liên quan đến dữ liệu trong ngành giáo dục đào tạo, bảo đảm việc thu thập, lưu trữ, xử lý, chia sẻ và sử dụng dữ liệu tuân thủ quy định pháp luật, đồng thời phục vụ hiệu quả công tác quản lý, điều hành và ra quyết định.

Hoạt động giám sát được thực hiện theo nguyên tắc liên tục, khách quan, minh bạch, có thể kiểm tra, truy vết và đánh giá, đảm bảo bao phủ toàn bộ các cấp quản lý.

Hoạt động giám sát trong ngành Giáo dục và Đào tạo bao gồm:

- Giám sát chất lượng dữ liệu;
- Giám sát quản lý dữ liệu chủ và danh mục dùng chung;
- Giám sát quản lý từ điển dữ liệu (metadata);
- Giám sát kiến trúc tổng thể dữ liệu;
- Giám sát điều phối, chia sẻ và liên thông dữ liệu;
- Giám sát quyền riêng tư và an toàn bảo mật dữ liệu;
- Giám sát nền tảng dữ liệu;
- Giám sát đo lường KPI, KQI;
- Giám sát kho dữ liệu và phân tích dữ liệu;
- Giám sát các dự án công nghệ thông tin;
- Giám sát khai thác và gia tăng giá trị dữ liệu;
- Giám sát dịch vụ dữ liệu.

1.4. Xây dựng văn hóa dữ liệu

Văn hóa dữ liệu trong ngành Giáo dục và Đào tạo được hình thành khi dữ liệu được coi là tài sản chiến lược, được quản lý thống nhất và sử dụng làm cơ sở cho quản lý, điều hành và ra quyết định. Văn hóa dữ liệu bảo đảm việc quản lý, chia sẻ và khai thác dữ liệu minh bạch, có trách nhiệm và tuân thủ quy định pháp luật.

Việc xây dựng văn hóa dữ liệu được thực hiện thông qua triển khai khung quản trị dữ liệu thống nhất trong toàn ngành, bao gồm các quy định và tiêu chuẩn về dữ liệu chủ, dữ liệu dùng chung, chất lượng dữ liệu, bảo mật và chia sẻ dữ liệu; đồng thời nâng cao năng lực cho đội ngũ trong phân tích và sử dụng dữ liệu, thúc đẩy ứng dụng dữ liệu trong quản lý, giảng dạy và khuyến khích đổi mới sáng tạo dựa trên dữ liệu.

Văn hóa dữ liệu được gắn với cơ chế đánh giá, trong đó các tiêu chí về chất lượng và mức độ sử dụng dữ liệu được đưa vào hệ thống đo lường, bảo đảm triển khai đồng bộ, liên tục và hiệu quả trong toàn ngành.

2. Lớp quản lý dữ liệu

Quản lý dữ liệu là việc tổ chức thực hiện có hệ thống các hoạt động nhằm hiện thực hóa các chính sách, quy trình và tiêu chuẩn trong khuôn khổ quản trị dữ liệu, bao gồm:

2.1. Quản lý Chất lượng dữ liệu

Mục tiêu của quản lý chất lượng dữ liệu trong ngành Giáo dục và Đào tạo là bảo đảm dữ liệu được duy trì ở mức độ chính xác, đầy đủ, kịp thời và nhất quán, phục vụ hiệu quả công tác quản lý, điều hành, thống kê và phân tích. Bộ tiêu chuẩn chất lượng dữ liệu giáo dục và đào tạo bao gồm:

(1) Các tiêu chí chất lượng dữ liệu:

- Tính đầy đủ: Đo lường sự hiện diện của các giá trị dữ liệu với các tiêu chuẩn nội bộ, bên ngoài hoặc của toàn ngành;
- Tính hợp lệ: Đo lường mức độ tuân thủ của dữ liệu với các tiêu chuẩn nội bộ, bên ngoài hoặc của toàn ngành và các quy định của pháp luật;
- Tính duy nhất: Tính duy nhất đề cập đến tính đơn lẻ của các bản ghi hoặc các giá trị dữ liệu nhằm mục đích đánh giá bản ghi có bị trùng lặp hay không, đảm bảo mỗi bản ghi là duy nhất;
- Tính chính xác: Đo lường độ chuẩn xác của dữ liệu, có thể đo lường dựa trên so sánh với tài liệu gốc hoặc nguồn ủy quyền và xác thực với các quy tắc nghiệp vụ đã xác định;
- Tính nhất quán: Tính nhất quán cung cấp sự đảm bảo rằng giá trị, định dạng và định nghĩa của dữ liệu trong một tổng thể là đồng nhất với các tổng thể khác;
- Tính kịp thời: Đo lường mức độ cập nhật của dữ liệu cũng như mức độ sẵn sàng của dữ liệu để sử dụng khi cần thiết;

- Tính toàn vẹn: Tính toàn vẹn dữ liệu đề cập đến việc dữ liệu không bị thay đổi, mất mát trong kho lưu trữ dữ liệu hay khi được truyền tải. Nói cách khác tính toàn vẹn là tính không bị hiệu chỉnh của dữ liệu.

(2) Quy tắc đánh giá chất lượng dữ liệu phải được xác định rõ ràng, bảo đảm khả năng áp dụng thống nhất, theo dõi và so sánh giữa các đơn vị.

- Các quy tắc nghiệp vụ đánh giá chất lượng dữ liệu được xây dựng theo yêu cầu của các Quản lý dữ liệu cấp đơn vị, Bộ phận quản trị dữ liệu cấp đơn vị và các bên liên quan đến dữ liệu;

- Khi các quy tắc nghiệp vụ về đánh giá chất lượng dữ liệu được xây dựng, các quy tắc này có thể chuyển thành các quy tắc kỹ thuật và áp dụng vào nền tảng dữ liệu tập trung nhằm đảm bảo tính khả thi trong triển khai.

(3) KPI chất lượng dữ liệu:

- KPI chất lượng dữ liệu: Là một tập hợp những nguyên tắc được sử dụng để đo lường và đánh giá chất lượng dữ liệu, cụ thể:

+ Chỉ tiêu kiểm tra định lượng dữ liệu;

+ Quy đổi ra điểm;

+ Các ngưỡng ghi nhận kết quả;

+ Chu kỳ kiểm soát định kỳ;

+ Các cấp phê duyệt;

+ Các cấp chịu trách nhiệm khắc phục.

- Các dữ liệu được thực hiện kiểm soát cần phải xây dựng hồ sơ chất lượng (Data Profiling) để tự động phát hiện ra những thiếu sót cơ bản.

- Việc thực hiện quét các KPI chất lượng dữ liệu được tiến hành tự động và có nhiều mức kiểm soát:

+ Đánh dấu sai và tự động thông qua dữ liệu đúng;

+ Yêu cầu phê duyệt những trường hợp sai;

+ Ưu tiên sử dụng dữ liệu nếu đáp ứng chỉ tiêu chất lượng theo quy định.

- Thứ tự thực hiện ưu tiên của các KPI dữ liệu phụ thuộc vào các yếu tố

+ Tầm quan trọng của phần tử dữ liệu;

+ Mức độ ảnh hưởng nghiêm trọng của sai sót dữ liệu (nếu xảy ra);

+ Tính khả thi của việc tính toán xử lý chất lượng dữ liệu với tính đúng hạn theo yêu cầu của nghiệp vụ;

+ Trong trường hợp xử lý chất lượng realtime/streaming, cần làm rõ những chỉ tiêu KPI được phép bỏ qua hoặc thực hiện xử lý sau đó.

- Có hai loại tạo lập dữ liệu cần những cách thức riêng biệt trong việc kiểm soát và xử lý lỗi chất lượng:

+ Dữ liệu tự động: Những sai sót do việc tính toán hoặc tự động cập nhật dữ liệu của hệ thống;

+ Dữ liệu nhập tay: Những sai sót do Phụ trách nhập liệu thực hiện.

Hoạt động kiểm soát, giám sát chất lượng dữ liệu phải được thiết lập và tổ chức thực hiện thống nhất trong toàn ngành, bao gồm việc kiểm tra, theo dõi, phát hiện và khắc phục các vấn đề về dữ liệu tại các cơ sở giáo dục và các đơn vị quản lý. Các đơn vị có trách nhiệm chủ động rà soát, xử lý sai lệch dữ liệu và phối hợp với các bên liên quan để bảo đảm các tiêu chí chất lượng dữ liệu, bao gồm nhưng không giới hạn các nội dung:

(1) Kiểm soát chất lượng dữ liệu là quy tắc, chốt kiểm soát được xác định và thiết lập trên các hệ thống, ứng dụng công nghệ và trong quá trình trích xuất, biến đổi và tải dữ liệu nhằm đảm bảo chất lượng dữ liệu được kiểm soát theo yêu cầu nhất định. Các kiểm soát chất lượng dữ liệu phải được xây dựng trên cơ sở các tiêu chí đánh giá chất lượng dữ liệu tương ứng.

(2) Chất lượng dữ liệu phải được đo lường và giám sát dựa trên ngưỡng chất lượng dữ liệu đã thiết lập thống nhất tương ứng.

(3) Báo cáo chất lượng dữ liệu được tổng hợp và xây dựng dựa trên kết quả đánh giá chất lượng dữ liệu thực tế và các ngưỡng chất lượng đã thống nhất trước đó nhằm mục đích trình bày một cách trực quan các thông tin về chất lượng dữ liệu.

Xử lý các vấn đề chất lượng dữ liệu:

(1) Xử lý các vấn đề về chất lượng dữ liệu bao gồm việc rà soát các hệ thống và quy trình thủ công, nghiệp vụ có liên quan nhằm xác định các nguyên nhân cốt lõi để xử lý các vấn đề về chất lượng dữ liệu theo thời gian một cách có hệ thống, giảm thiểu các rủi ro phát sinh do chất lượng dữ liệu không đảm bảo.

(2) Xử lý các vấn đề về chất lượng dữ liệu bao gồm các giải pháp đề xuất và kế hoạch xử lý phù hợp với các nguyên nhân tương ứng. Việc xử lý các vấn đề về chất lượng dữ liệu đòi hỏi sự tham gia phối hợp của tất cả các bên liên quan.

(3) Các vấn đề nghiêm trọng về chất lượng dữ liệu (chất lượng dữ liệu vi phạm và việc chậm tiến độ trong việc xử lý các vấn đề về chất lượng dữ liệu) được ghi nhận bởi bộ phận quản trị dữ liệu phải được trình lên cấp có thẩm quyền để xử lý.

(4) Bảo đảm tuân thủ quy định của pháp luật liên quan đến dữ liệu trong việc xử lý các vấn đề về chất lượng dữ liệu.

Các công cụ và hệ thống hỗ trợ quản lý chất lượng dữ liệu phải được xây dựng và vận hành nhằm tự động hóa việc kiểm soát chất lượng, phát hiện và cảnh báo kịp thời các

vấn đề phát sinh; đồng thời hỗ trợ theo dõi, đánh giá và nâng cao chất lượng dữ liệu trong toàn ngành Giáo dục và Đào tạo.

2.2. Quản lý dữ liệu chủ và dữ liệu danh mục dùng chung

2.2.1. Quản lý dữ liệu chủ

Mục tiêu của quản lý dữ liệu chủ trong ngành Giáo dục và Đào tạo là bảo đảm dữ liệu chủ được quản lý thống nhất, chính xác và dùng chung trong toàn ngành; hình thành nguồn dữ liệu tin cậy phục vụ công tác quản lý, điều hành, thống kê và phân tích; đồng thời hạn chế trùng lặp và sai lệch dữ liệu giữa các hệ thống.

(1) Dữ liệu chủ phải được thiết lập hoặc khởi tạo trên nguyên tắc nguồn dữ liệu tin cậy duy nhất (Single Source of Truth – SSOT): Một phần tử dữ liệu chỉ có một nguồn duy nhất. Nếu có nhiều nguồn, phải xác định cụ thể 01 nguồn được coi là nguồn tin cậy duy nhất;

(2) Dữ liệu chủ phải được thiết lập hoặc khởi tạo từ bộ khóa định danh (Data key). Khóa định danh là những phần tử dữ liệu đặc biệt có đặc tính cơ bản như sau:

- Là những phần tử dữ liệu giúp định danh chính xác một đối tượng cụ thể;
- Khóa có thể là một phần tử dữ liệu (Single key) hoặc kết hợp từ nhiều phần tử dữ liệu (Combined key);
- Khóa là các phần tử dữ liệu sau khi thực hiện chuẩn hóa về chất lượng và khuôn dạng;
- Khóa phải đạt một số tiêu chuẩn nhất định khi thực hiện hồ sơ dữ liệu (Data profiling);
- Bộ dữ liệu khóa định danh thường trùng hợp với những dữ liệu định danh cá nhân. Việc phân quyền, bảo mật phải tuân thủ nghiêm túc theo quy định của pháp luật, của Bộ Giáo dục và Đào tạo về bảo vệ dữ liệu cá nhân (Data privacy).

(3) Mỗi dữ liệu chủ phải được gắn với một Mã số định danh bản ghi vàng (GRK), là mã số chuẩn hóa do hệ thống quản trị dữ liệu tự động sinh ra, đảm bảo một GRK tương ứng với một và chỉ một đối tượng. GRK là kết quả sau khi hoàn thiện quá trình Quản trị dữ liệu chủ.

(4) Dữ liệu chủ phải được đảm bảo chất lượng bằng giải pháp chống trùng lặp dữ liệu (De-duplicate), là tập hợp các nguyên tắc và công cụ thực hiện để phát hiện ra sự trùng lặp của các khóa định danh.

- Quy tắc kiểm tra do Quản lý dữ liệu cấp đơn vị quy định, có sự hỗ trợ chuyên môn từ Bộ phận quản trị dữ liệu;

- Tiêu chuẩn, độ chính xác, quy trình phát hiện, xác nhận, chỉnh sửa, hậu kiểm dữ liệu chủ là sự phối hợp giữa Quản lý dữ liệu cấp đơn vị, Bộ phận quản trị dữ liệu, Quản lý các cấp và Bộ phận tạo lập sẽ được quy định cụ thể trong các quy trình chất lượng dữ liệu.

(5) Quá trình thu thập tạo lập dữ liệu chủ cần có những cơ chế riêng biệt trong việc kiểm soát, xử lý lỗi và có biện pháp sửa đổi, bổ sung, điều chỉnh để khắc phục theo đúng quy định và quy trình, có trách nhiệm cụ thể của các đơn vị và cá nhân tham gia. Bao gồm các công đoạn:

- Chỉnh sửa sai sót dữ liệu phát sinh trong quá trình cập nhật dữ liệu tự động từ hệ thống nguồn hoặc trong quá trình tính toán dữ liệu tự động. Các thay đổi về dữ liệu chủ từ nguồn đều phải gửi thông tin đến nền tảng dữ liệu tập trung trước khi thực hiện;

- Phương pháp phòng ngừa những sai sót phát sinh do việc tính toán hoặc tự động cập nhật dữ liệu của hệ thống được thực hiện thông qua: Xây dựng hồ sơ dữ liệu kết hợp với Kiểm soát chỉ tiêu KPI chất lượng dữ liệu để phát hiện sớm và chỉnh sửa hệ thống; tiến hành chỉnh sửa và khắc phục hậu quả;

- Giải pháp khắc phục khi có sai sót xảy ra và được phản ánh bởi người dùng cuối: chỉnh sửa hệ thống nhập liệu, đưa ra các cơ chế kỹ thuật hỗ trợ và hạn chế nhập sai (nhập liệu từ danh sách chọn sẵn (List of values), tiêu chí kiểm tra dữ liệu khi nhập liệu (Validation rules));

- Việc thực hiện khắc phục sai sót dữ liệu xảy ra trong quá trình nhập liệu, do việc tính toán hoặc tự động cập nhật dữ liệu của hệ thống cần được phê duyệt theo quy định.

(6) Quy trình quản lý dữ liệu chủ phải được xây dựng, ban hành và áp dụng thống nhất, bao gồm các hoạt động tạo mới, cập nhật, chỉnh sửa, lưu trữ và hủy/xóa dữ liệu.

(7) Dữ liệu chủ phải được chuẩn hóa theo mô hình dữ liệu thống nhất toàn ngành và được định nghĩa đầy đủ về khái niệm, cấu trúc và cách sử dụng; đồng thời được quản lý trong Từ điển dữ liệu ngành nhằm bảo đảm thống nhất về thuật ngữ và hỗ trợ tích hợp, chia sẻ dữ liệu.

(8) Danh mục và nội dung dữ liệu chủ phải được rà soát, cập nhật định kỳ, phù hợp với yêu cầu quản lý và thực tiễn phát sinh trong ngành giáo dục. Việc quản lý phiên bản dữ liệu phải được thực hiện nhằm ghi nhận đầy đủ lịch sử thay đổi, bảo đảm khả năng truy vết và phục vụ công tác kiểm tra, đối soát và khai thác dữ liệu một cách tin cậy.

(9) Cơ chế kiểm soát và giám sát dữ liệu chủ phải được thiết lập để theo dõi chất lượng, việc cập nhật và sử dụng dữ liệu, phục vụ đánh giá mức độ tuân thủ và nâng cao hiệu quả quản lý. Trách nhiệm được phân công rõ ràng theo cấp: Bộ Giáo dục và Đào tạo quản lý tổng thể và ban hành tiêu chuẩn; Sở Giáo dục và Đào tạo quản lý trong phạm vi địa phương; cơ sở giáo dục tạo lập, cập nhật và bảo đảm chất lượng. Mỗi loại dữ liệu chủ phải có đơn vị chịu trách nhiệm chính và đơn vị quản lý, vận hành tương ứng.

2.2.2. Quản lý dữ liệu danh mục dùng chung

Mục tiêu của quản lý dữ liệu danh mục dùng chung trong ngành Giáo dục và Đào tạo là bảo đảm các danh mục, bảng mã được xây dựng thống nhất, sử dụng đồng bộ trong toàn ngành, làm cơ sở chuẩn hóa dữ liệu, hỗ trợ tích hợp, chia sẻ và liên thông dữ liệu giữa các hệ thống.

Việc quản lý dữ liệu danh mục dùng chung được thực hiện theo các nguyên tắc bảo đảm tính thống nhất, dùng chung toàn ngành; mỗi danh mục có một nguồn quản lý chính;

được chuẩn hóa về cấu trúc, mã hóa và định nghĩa; được sử dụng thống nhất trong các hệ thống thông tin; đồng thời bảo đảm khả năng mở rộng, cập nhật và truy vết thay đổi.

(1) Dữ liệu danh mục dùng chung phải được thiết lập hoặc khởi tạo từ nguồn dữ liệu chuẩn hóa của dữ liệu và được áp dụng nhất quán trên các hệ thống, ứng dụng công nghệ.

(2) Dữ liệu danh mục dùng chung phải có mã định danh duy nhất và rõ ràng, các cơ chế kiểm soát phù hợp phải được thiết lập nhằm tránh sự trùng lặp. Bộ Giáo dục và Đào tạo có trách nhiệm chủ trì xây dựng, ban hành và quản lý danh mục dữ liệu dùng chung, bảo đảm phù hợp và thống nhất với các danh mục quốc gia và liên ngành.

(3) Dữ liệu danh mục dùng chung phải được rà soát định kỳ tối thiểu hàng năm để đảm bảo tính phù hợp trong việc được ghi nhận là dữ liệu danh mục dùng chung.

(4) Quyền truy cập, sử dụng và chia sẻ dữ liệu danh mục dùng chung phải được triển khai một cách phù hợp theo các yêu cầu nghiệp vụ, các yêu cầu của pháp luật, của Bộ Giáo dục và Đào tạo về bảo vệ dữ liệu.

(5) Sử dụng công cụ có tính năng lưu trữ và hỗ trợ truy cập vào dữ liệu danh mục dùng chung trong quá khứ. Đảm bảo việc có thể truy cập vào phiên bản trước của dữ liệu danh mục dùng chung và thể hiện trực quan các thay đổi đối với dữ liệu danh mục dùng chung theo thời gian.

(6) Tất cả các thay đổi đối với dữ liệu danh mục dùng chung cần được xác định nội dung, thời hạn bắt đầu thay đổi để truyền thông rộng rãi tới tất cả người dùng để hiểu được các ảnh hưởng có thể xảy ra và chủ động trong việc sử dụng và khai thác.

Các hệ thống thông tin trong ngành giáo dục phải sử dụng thống nhất các danh mục dùng chung đã được ban hành, không tự ý tạo mới hoặc thay đổi mã danh mục khi chưa được phê duyệt. Trường hợp phát sinh nhu cầu bổ sung hoặc điều chỉnh danh mục, các đơn vị phải thực hiện theo quy trình đề xuất, thẩm định và phê duyệt thống nhất.

Dữ liệu danh mục dùng chung phải được định nghĩa đầy đủ về ý nghĩa, cấu trúc, mã hóa và phạm vi sử dụng; được quản lý trong Từ điển dữ liệu ngành nhằm bảo đảm thống nhất về khái niệm và hỗ trợ tích hợp, chia sẻ dữ liệu

Trách nhiệm quản lý dữ liệu danh mục dùng chung được phân công rõ ràng, trong đó Bộ Giáo dục và Đào tạo chịu trách nhiệm quản lý tổng thể và ban hành danh mục; các Sở Giáo dục và Đào tạo và cơ sở giáo dục có trách nhiệm tuân thủ, sử dụng đúng và phản hồi các nhu cầu cập nhật.

2.2.3. Quản lý Từ điển dữ liệu

Từ điển dữ liệu dùng chung trong ngành Giáo dục và Đào tạo là công cụ chuẩn hóa và quản lý thống nhất siêu dữ liệu trong toàn ngành, làm nền tảng cho việc tích hợp, chia sẻ, liên thông và khai thác dữ liệu giữa các cơ quan quản lý giáo dục, các cơ sở giáo dục và kết nối với các hệ thống dữ liệu quốc gia. Các nguyên tắc quản lý Từ điển dữ liệu bao gồm:

(1) Từ điển dữ liệu phải được tài liệu hóa tập trung, bao gồm siêu dữ liệu nghiệp vụ, siêu dữ liệu kỹ thuật và siêu dữ liệu vận hành, nhằm bảo đảm mô tả đầy đủ về ý nghĩa, cấu

trúc, nguồn gốc, vòng đời và cách thức sử dụng của từng yếu tố dữ liệu. Toàn bộ nội dung liên quan, bao gồm định nghĩa, mẫu biểu thu thập, quy trình và cơ chế kiểm soát phải được lưu trữ tập trung tại kho siêu dữ liệu dùng chung của ngành.

(2) Từ điển dữ liệu cần được rà soát, xây dựng và cập nhật liên tục, bảo đảm tính đồng bộ và chuẩn hóa; đồng thời phải được triển khai song song ngay khi có yêu cầu phát triển dữ liệu hoặc khi xây dựng mới, nâng cấp, kết nối, tích hợp và nghiệm thu hệ thống thông tin. Kho siêu dữ liệu tập trung phải được rà soát định kỳ ít nhất mỗi năm một lần.

(3) Các định nghĩa trong Từ điển dữ liệu phải rõ ràng, được hiểu theo một nghĩa duy nhất, và được công khai cho tất cả người dùng thông qua Từ điển dữ liệu.

(4) Từ điển dữ liệu phải được áp dụng bắt buộc và thống nhất trong toàn bộ hệ thống thông tin và cơ sở dữ liệu của ngành, bảo đảm tính nhất quán trong định nghĩa, cấu trúc và cách sử dụng dữ liệu.

(5) Việc xây dựng và sử dụng từ điển dữ liệu ngành Giáo dục và Đào tạo phải bảo đảm tính thống nhất, kế thừa và không mâu thuẫn với Từ điển dữ liệu quốc gia, đồng thời phản ánh đầy đủ đặc thù nghiệp vụ của ngành.

(6) Mọi thay đổi đối với Từ điển dữ liệu phải được thẩm định, phê duyệt và ghi nhận đầy đủ, bảo đảm tính minh bạch, ổn định và khả năng truy vết.

(7) Kho siêu dữ liệu tập trung phải cho phép truy cập trên phạm vi toàn ngành theo chức năng, nhiệm vụ được giao; bảo đảm người dùng có thể tra cứu dữ liệu một cách dễ dàng, kịp thời trên cơ sở phân quyền phù hợp và tuân thủ quy định về an toàn, bảo mật thông tin.

(8) Việc xây dựng, quản lý và khai thác từ điển dữ liệu phải được phân công trách nhiệm rõ ràng theo các vai trò trong mô hình quản trị dữ liệu; từ điển dữ liệu phải được tích hợp xuyên suốt trong các quy trình quản trị, bao gồm: thiết kế hệ thống, tích hợp dữ liệu, kiểm soát chất lượng, phân tích và khai thác dữ liệu.

2.2.4. Quản lý quyền riêng tư và an toàn bảo mật dữ liệu

1. Quản lý quyền riêng tư dữ liệu

Quản lý quyền riêng tư dữ liệu là tập hợp các chính sách, quy trình và biện pháp nhằm bảo vệ quyền riêng tư của dữ liệu, đảm bảo dữ liệu cá nhân của chủ thể dữ liệu được thu thập, lưu trữ, xử lý và chia sẻ đúng mục đích, minh bạch, an toàn và tuân thủ pháp luật. Việc xử lý dữ liệu cá nhân phải được thực hiện trên cơ sở có căn cứ pháp lý hợp lệ, đúng mục đích, trong phạm vi cần thiết và bảo đảm tôn trọng quyền riêng tư của chủ thể dữ liệu theo quy định của pháp luật.

Chủ thể dữ liệu có quyền kiểm soát dữ liệu của mình, bao gồm quyền được biết, truy cập, chỉnh sửa, xóa và rút lại sự đồng ý; đơn vị quản lý dữ liệu có trách nhiệm thiết lập cơ chế tiếp nhận, xử lý và bảo đảm thực hiện các quyền này.

Bộ phận quản trị dữ liệu phải áp dụng các biện pháp tổ chức và kỹ thuật để bảo vệ dữ liệu cá nhân, thực hiện giám sát, truy vết, đánh giá rủi ro và đánh giá tác động khi cần thiết, đồng thời chịu trách nhiệm giải trình và xử lý sự cố theo quy định.

Trong trường hợp sử dụng công nghệ trí tuệ nhân tạo (AI) để xử lý dữ liệu giáo dục, các đơn vị phải tuân thủ nguyên tắc AI có trách nhiệm, bao gồm: dữ liệu huấn luyện phải được thu thập và sử dụng hợp pháp; mô hình phải được kiểm soát và giảm thiểu thiên lệch có thể gây bất lợi cho người học; kết quả đầu ra phải có khả năng giải thích phù hợp với từng ngữ cảnh; con người phải được bảo đảm quyền giám sát và can thiệp trong suốt quá trình vận hành. Nghiêm cấm sử dụng AI để ra quyết định tự động ảnh hưởng trực tiếp đến quyền lợi của người học khi chưa thiết lập cơ chế cho phép kiểm tra lại kết quả, tiếp nhận khiếu nại và quy định rõ trách nhiệm giải trình của đơn vị triển khai.

2. Quản lý an toàn bảo mật dữ liệu

Tất cả các cơ sở dữ liệu thuộc phạm vi quản lý của Bộ Giáo dục và Đào tạo 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 về Khung an toàn dữ liệu 8 lớp được quy định trong Quyết định 2439/QĐ-TTg. 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, Khung quản trị quản lý dữ liệu quốc gia nhưng vẫn bảo đảm nguyên tắc thống nhất về kiểm soát rủi ro và bảo vệ an toàn thông tin. Khung an toàn dữ liệu 8 lớp bao gồm:

a) Lớp 1 – Dữ liệu gốc

(1) Dữ liệu gốc là dữ liệu được tạo lập ban đầu trong các hoạt động:

- Quản lý người học;
- Quản lý nhà giáo, cán bộ quản lý giáo dục;
- Quản lý các cơ sở giáo dục đào tạo;
- Quản lý các cơ quan quản lý giáo dục;
- Khảo thí, kiểm định chất lượng;
- Chương trình đào tạo, học liệu số và kết quả học tập;
- Các hoạt động quản lý giáo dục đào tạo và nghiệp vụ chuyên ngành khác.

(2) Mỗi tập dữ liệu cần được gắn mã định danh duy nhất và được quản lý siêu dữ liệu đầy đủ, chuẩn hóa, gắn thông tin về đơn vị sở hữu, đơn vị quản trị quản lý, đơn vị quản lý kỹ thuật và lịch sử cập nhật.

b) Lớp 2 – Phân loại và gán nhãn dữ liệu

(1) Tất cả dữ liệu phải được kiểm kê, phân loại và gán nhãn. Việc gán nhãn có thể được thực hiện thủ công hoặc bằng công cụ. Các tập dữ liệu đều phải có nhãn dữ liệu được quản lý qua Từ điển dữ liệu và phục vụ tự động hóa kiểm soát quyền truy cập, lập nhật ký sử dụng và đánh giá rủi ro. Việc kiểm kê và phân loại dữ liệu được thực hiện:

- Khi dữ liệu được tạo lập hoặc tiếp nhận lần đầu;
- Định kỳ theo kế hoạch;
- Đột xuất trong các trường hợp sau:

- + Phát sinh dữ liệu mới hay có sự thay đổi, yêu cầu từ văn bản pháp luật;
- + Có thay đổi trong hoạt động nghiệp vụ, mô hình tổ chức, hệ thống công nghệ thông tin của ngành Giáo dục và Đào tạo;
- + Phát hiện bất cập, rủi ro liên quan đến mức độ bảo mật của dữ liệu đã được phân loại.

(2) Việc phân loại dữ liệu phải căn cứ trên một hoặc nhiều tiêu chí sau:

- Quy định của pháp luật có liên quan;
- Mức độ nhạy cảm của dữ liệu;
- Mức độ ảnh hưởng, rủi ro có thể phát sinh trong trường hợp dữ liệu bị truy cập trái phép, sử dụng sai mục đích, bị mất, bị lộ, bị thay đổi hoặc phá hủy;
- Giá trị, tầm quan trọng của dữ liệu đối với hoạt động giáo dục.

(3) Dữ liệu ngành GD&ĐT được phân loại thành 05 nhóm cấp độ bảo mật tăng dần:

- Nhóm dữ liệu công khai: là dữ liệu được chia sẻ công khai, công bố rộng rãi trên các phương tiện truyền thông đại chúng hoặc có thể gửi cho tất cả các đối tượng. Ví dụ: thông tin tuyển sinh, chương trình đào tạo công bố.

- Nhóm dữ liệu nội bộ: là dữ liệu phục vụ cho hoạt động quản lý nội bộ ngành Giáo dục và Đào tạo, mọi cán bộ nhân viên được phép tiếp cận, khai thác, sử dụng nhưng không được phép chia sẻ ra bên ngoài phạm vi của cơ quan quản lý nhà nước về giáo dục. Ví dụ: kế hoạch giảng dạy, báo cáo chuyên môn, báo cáo nội bộ.

- Nhóm dữ liệu hạn chế: là dữ liệu có nội dung quan trọng, mang tính chuyên môn, nghiệp vụ. Dữ liệu hạn chế chỉ được chia sẻ, sử dụng trong phạm vi các đơn vị, bộ phận hoặc nhóm đối tượng được xác định, phù hợp với chức năng, nhiệm vụ được giao, không được phép cung cấp, chia sẻ ra bên ngoài khi chưa được cấp có thẩm quyền phê duyệt. Ví dụ: đề thi chưa công bố, ngân hàng câu hỏi, kế hoạch kiểm tra.

- Nhóm dữ liệu nhạy cảm: là dữ liệu mà việc lộ, mất hoặc bị sử dụng trái phép có thể gây ảnh hưởng đáng kể đến quyền và lợi ích hợp pháp của tổ chức, cá nhân hoặc hoạt động của cơ quan. Bao gồm:

+ Dữ liệu cá nhân: là dữ liệu mà việc lộ, mất hoặc bị sử dụng trái phép có thể gây ảnh hưởng trực tới quyền, lợi ích hợp pháp của tổ chức, cá nhân. Dữ liệu cá nhân bao gồm dữ liệu cá nhân cơ bản, dữ liệu cá nhân nhạy cảm được quy định tại Nghị định 356/2025/NĐ-CP Quy định chi tiết một số điều và biện pháp thi hành Luật bảo vệ dữ liệu cá nhân.

+ Dữ liệu cốt lõi, dữ liệu quan trọng: là những dữ liệu được xác định theo các tiêu chí trong Nghị định 165/2025/NĐ-CP Quy định chi tiết một số điều và biện pháp thi hành Luật dữ liệu; Quyết định 20/2025/QĐ-TTg Ban hành Danh mục dữ liệu quan trọng, dữ liệu cốt lõi. Ví dụ: Thông tin cá nhân học sinh, sinh viên; kết quả học tập, học bạ...

- Nhóm dữ liệu bí mật: là dữ liệu có nội dung đặc biệt quan trọng, có mức độ ảnh hưởng nghiêm trọng hoặc đặc biệt nghiêm trọng đối với an ninh quốc gia, trật tự an toàn xã hội, lợi ích của Nhà nước hoặc quyền và lợi ích hợp pháp của tổ chức, cá nhân nếu bị lộ, bị mất hoặc bị sử dụng trái phép.

+ Dữ liệu bí mật chỉ được tiếp cận, khai thác, sử dụng bởi các tổ chức, cá nhân có thẩm quyền và được phê duyệt theo quy định.

+ Dữ liệu bí mật bao gồm các dữ liệu thuộc danh mục Tối mật, Mật theo quy định của pháp luật về bảo vệ bí mật nhà nước và các dữ liệu có tính chất tương đương theo quy định của Bộ.

Ví dụ: Đề thi quốc gia, dữ liệu chấm thi; dữ liệu điều hành chiến lược ngành; dữ liệu thuộc danh mục bí mật nhà nước.

(4) Mỗi dữ liệu chỉ được phân loại vào 1 cấp độ bảo mật. Trường hợp dữ liệu được tổng hợp, tích hợp từ nhiều nguồn dữ liệu có cấp độ bảo mật khác nhau thì cấp độ bảo mật của dữ liệu tổng hợp được xác định theo cấp độ cao nhất của các dữ liệu thành phần.

c) Lớp 3 – Quản lý truy cập và mã hóa dữ liệu

(1) Kiểm soát truy cập dữ liệu:

- Việc truy cập, khai thác và xử lý dữ liệu trong các hệ thống thuộc phạm vi quản lý của Bộ Giáo dục và Đào tạo phải được kiểm soát thống nhất, chặt chẽ và xuyên suốt theo nguyên tắc bảo đảm đúng chủ thể, đúng thẩm quyền, đúng mục đích, đúng phạm vi và trong thời gian được phép, đồng thời tuân thủ nguyên tắc đặc quyền tối thiểu và phân tách nhiệm vụ nhằm hạn chế tối đa rủi ro lạm dụng quyền truy cập, truy cập trái phép hoặc sử dụng dữ liệu sai mục đích.

- Đơn vị quản lý hệ thống có trách nhiệm tổ chức thiết lập, vận hành và duy trì cơ chế kiểm soát truy cập dựa trên mô hình phân quyền theo vai trò hoặc phân quyền theo thuộc tính, trong đó việc xác định quyền truy cập phải căn cứ vào chức năng, nhiệm vụ được giao của từng cá nhân, đơn vị công tác, cấp độ nhạy cảm của dữ liệu và mục đích sử dụng đã được phê duyệt theo quy định. Việc cấp, thay đổi, gia hạn hoặc thu hồi quyền truy cập phải được thực hiện theo quy trình kiểm soát nội bộ, có phê duyệt của cấp có thẩm quyền, có xác định thời hạn cụ thể và được rà soát định kỳ hoặc khi phát sinh thay đổi về tổ chức, nhân sự, chức năng nhiệm vụ hoặc mức độ rủi ro.

- Mỗi cá nhân sử dụng hệ thống phải được cấp một định danh truy cập duy nhất, bảo đảm khả năng nhận diện, truy vết và chịu trách nhiệm đối với toàn bộ hoạt động truy cập và xử lý dữ liệu. Nghiêm cấm việc sử dụng chung tài khoản dưới mọi hình thức, trừ trường hợp đặc biệt phục vụ yêu cầu kỹ thuật hoặc nhiệm vụ đặc thù, phải được cấp có thẩm quyền phê duyệt và áp dụng cơ chế giám sát tăng cường. Các tài khoản không còn nhu cầu sử dụng hoặc không phát sinh hoạt động trong thời gian theo quy định phải được tự động khóa hoặc thu hồi.

- Đối với tài khoản có quyền quản trị hệ thống, quản trị cơ sở dữ liệu hoặc quyền truy cập tới dữ liệu nhạy cảm, dữ liệu quan trọng, đơn vị quản lý hệ thống phải thiết lập cơ chế kiểm soát đặc biệt, bao gồm việc tách biệt tài khoản quản trị với tài khoản sử dụng thông thường, kiểm soát truy cập theo từng phiên làm việc hoặc theo từng nhiệm vụ cụ thể, ghi nhận đầy đủ nhật ký hoạt động và tổ chức giám sát liên tục nhằm phát hiện, cảnh báo và ngăn chặn kịp thời các hành vi truy cập trái phép hoặc vượt quyền.

- Toàn bộ hoạt động truy cập dữ liệu phải được ghi nhận, lưu trữ và bảo vệ dưới dạng nhật ký truy cập, bảo đảm phản ánh đầy đủ thông tin về định danh người truy cập,

thời điểm truy cập, nội dung truy cập, thao tác thực hiện và tập dữ liệu liên quan. Nhật ký truy cập phải được bảo vệ khỏi việc chỉnh sửa, xóa bỏ trái phép, được lưu trữ trong thời hạn phù hợp với cấp độ dữ liệu và yêu cầu quản lý, đồng thời phải sẵn sàng cung cấp phục vụ công tác kiểm tra, giám sát, kiểm toán và điều tra khi có yêu cầu của cơ quan có thẩm quyền.

- Bộ phận chủ quản hệ thống có trách nhiệm tổ chức rà soát, đánh giá quyền truy cập dữ liệu định kỳ hoặc đột xuất nhằm bảo đảm sự phù hợp giữa quyền được cấp và chức năng, nhiệm vụ thực tế của người sử dụng, kịp thời phát hiện và xử lý các trường hợp cấp quyền không phù hợp, vượt thẩm quyền hoặc tiềm ẩn rủi ro đối với an toàn dữ liệu.

(2) Xác thực và mã hóa dữ liệu:

- Việc xác thực người dùng và bảo vệ dữ liệu bằng các biện pháp mã hóa phải được triển khai thống nhất, đồng bộ và tương ứng với mức độ nhạy cảm của dữ liệu, bảo đảm ngăn chặn truy cập trái phép, bảo vệ tính bí mật, toàn vẹn và khả năng sẵn sàng của dữ liệu trong toàn bộ quá trình xử lý.

- Hệ thống thông tin phải áp dụng cơ chế xác thực người dùng với mức độ bảo đảm phù hợp với cấp độ dữ liệu và rủi ro an toàn thông tin, trong đó bắt buộc áp dụng xác thực đa yếu tố (MFA) đối với tài khoản quản trị hệ thống, quản trị dữ liệu, người dùng truy cập vào dữ liệu thuộc cấp độ bảo mật cao hoặc truy cập từ môi trường không tin cậy. Các yếu tố xác thực phải được thiết kế theo nguyên tắc kết hợp ít nhất hai trong ba nhóm yếu tố bao gồm thông tin bí mật do người dùng nắm giữ, phương tiện xác thực do người dùng sở hữu và đặc điểm sinh trắc học, bảo đảm độ tin cậy và khả năng chống giả mạo.

- Thông tin xác thực của người dùng phải được quản lý chặt chẽ, trong đó mật khẩu phải đáp ứng yêu cầu về độ phức tạp, được thay đổi định kỳ và không được trùng lặp với các thông tin định danh dễ đoán; toàn bộ thông tin xác thực phải được lưu trữ dưới dạng mã hóa hoặc băm bảo mật theo tiêu chuẩn kỹ thuật phù hợp, nghiêm cấm lưu trữ hoặc truyền tải dưới dạng rõ. Việc cấp phát, thay đổi và thu hồi thông tin xác thực phải được kiểm soát theo quy trình nội bộ, có ghi nhận và truy vết đầy đủ.

- Dữ liệu trong các hệ thống thông tin phải được mã hóa phù hợp với cấp độ nhạy cảm và bối cảnh sử dụng, trong đó bắt buộc thực hiện mã hóa đối với dữ liệu trong quá trình truyền qua mạng, dữ liệu lưu trữ trên hệ thống, dữ liệu sao lưu và dữ liệu được lưu trữ hoặc xử lý trên các môi trường không bảo đảm an toàn vật lý hoặc logic. Việc lựa chọn thuật toán, giao thức và giải pháp mã hóa phải tuân thủ các tiêu chuẩn, quy chuẩn kỹ thuật do cơ quan có thẩm quyền ban hành, không được sử dụng các thuật toán hoặc phương thức mã hóa đã bị đánh giá là không còn an toàn.

- Khóa mã hóa phải được quản lý theo cơ chế kiểm soát tập trung hoặc phân vùng logic phù hợp với mô hình tổ chức và phạm vi quản lý, bảo đảm tách biệt giữa dữ liệu và khóa mã hóa, kiểm soát chặt chẽ quyền truy cập và sử dụng khóa, đồng thời thiết lập đầy đủ các quy trình tạo lập, phân phối, lưu trữ, luân chuyển, thay thế, sao lưu và hủy khóa nhằm phòng ngừa rủi ro lộ lọt hoặc mất kiểm soát khóa mã hóa.

- Bộ phận chủ quản hệ thống có trách nhiệm tổ chức kiểm tra, đánh giá định kỳ hoặc đột xuất việc tuân thủ các quy định về xác thực và mã hóa dữ liệu, kịp thời phát hiện, khắc phục các tồn tại, lỗ hổng hoặc vi phạm, đồng thời lập và lưu trữ hồ sơ, báo cáo phục vụ công tác kiểm tra, thanh tra và giám sát của cơ quan có thẩm quyền.

d) Lớp 4 – Che giấu dữ liệu và truy vết

(1) Che giấu dữ liệu:

- Việc truy cập, khai thác và hiển thị dữ liệu phải được kiểm soát theo nguyên tắc tiết lộ tối thiểu cần thiết, trong đó đơn vị quản lý hệ thống có trách nhiệm áp dụng các biện pháp kỹ thuật phù hợp để che giấu, làm mờ, thay thế hoặc hạn chế hiển thị thông tin nhằm bảo đảm người sử dụng chỉ được tiếp cận phần dữ liệu phù hợp với chức năng, nhiệm vụ và mục đích đã được phê duyệt.

- Đối với dữ liệu cá nhân, dữ liệu cốt lõi và dữ liệu quan trọng, việc cung cấp, chia sẻ hoặc sử dụng dữ liệu phải được thực hiện trên cơ sở đã áp dụng biện pháp che giấu hoặc giảm thiểu khả năng nhận dạng, trừ trường hợp phục vụ nhiệm vụ theo quy định của pháp luật; mọi trường hợp truy cập đầy đủ dữ liệu gốc phải được kiểm soát, ghi nhận và có căn cứ pháp lý rõ ràng.

(2) Truy vết dữ liệu:

- Dữ liệu trong cơ sở dữ liệu phải được quản lý theo cơ chế truy vết xuyên suốt, trong đó bộ phận chủ quản hệ thống có trách nhiệm ghi nhận, lưu trữ và bảo đảm khả năng truy xuất các thông tin về nguồn gốc dữ liệu, quá trình thu thập, xử lý, biến đổi và sử dụng dữ liệu nhằm phục vụ công tác kiểm tra, giám sát, kiểm toán và xử lý sự cố.

- Thông tin truy vết dữ liệu phải được bảo vệ an toàn, không bị sửa đổi trái phép, được lưu trữ tập trung và tích hợp với hệ thống giám sát để phát hiện kịp thời các hành vi truy cập trái phép, sử dụng sai mục đích hoặc bất thường trong quá trình xử lý dữ liệu, đồng thời xác định trách nhiệm của các tổ chức, cá nhân có liên quan.

e) Lớp 5 – Giám sát và đánh giá theo khung “5 yếu tố an toàn”

Các hệ thống và cơ sở dữ liệu phải được tổ chức giám sát và đánh giá việc khai thác, xử lý, lưu trữ và chia sẻ dữ liệu trên cơ sở tuân thủ đầy đủ 05 yếu tố an toàn, bao gồm người sử dụng, mục đích sử dụng, dữ liệu, môi trường xử lý và kết quả đầu ra, nhằm bảo đảm dữ liệu được sử dụng đúng thẩm quyền, đúng mục đích, đúng phạm vi và trong điều kiện an toàn.

Bộ phận chủ quản hệ thống có trách nhiệm thiết lập cơ chế kiểm soát, giám sát và đánh giá phù hợp đối với từng yếu tố an toàn; bảo đảm việc kiểm tra được thực hiện trước, trong và sau quá trình xử lý dữ liệu; kịp thời phát hiện, ngăn chặn và xử lý các hành vi vi phạm hoặc nguy cơ gây mất an toàn dữ liệu.

05 yếu tố an toàn:

- 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.

- 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.

f) 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.

Dữ liệu phải được quản lý xuyên suốt theo vòng đời, từ khi được thu thập hoặc tạo lập đến khi lưu trữ, khai thác, chia sẻ và xóa, hủy, bảo đảm phù hợp với mục đích ban đầu, cấp độ phân loại và yêu cầu bảo vệ dữ liệu trong từng giai đoạn xử lý.

Bộ phận quản trị dữ liệu có trách nhiệm xây dựng và tổ chức thực hiện quy trình quản lý vòng đời dữ liệu, trong đó bao gồm các nội dung về thu thập dữ liệu theo nguyên tắc hợp pháp, tối thiểu và minh bạch; lưu trữ, cập nhật và khai thác dữ liệu phù hợp với cấp độ nhạy cảm; kiểm soát việc chia sẻ dữ liệu; và thực hiện xóa, hủy dữ liệu khi hết thời hạn lưu trữ hoặc không còn cần thiết cho mục đích xử lý.

Việc rà soát dữ liệu phải được thực hiện định kỳ nhằm phát hiện và xử lý dữ liệu trùng lặp, lỗi thời hoặc không còn giá trị sử dụng; việc xóa, hủy dữ liệu phải bảo đảm không thể khôi phục, có ghi nhận nhật ký và xác định rõ trách nhiệm của tổ chức, cá nhân thực hiện.

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.

g) Lớp 7 – Công cụ kỹ thuật bắt buộc

(1) Yêu cầu chung về công cụ kỹ thuật:

Các hệ thống thông tin thuộc phạm vi quản lý của ngành giáo dục phải được trang bị và vận hành các công cụ kỹ thuật nhằm bảo đảm năng lực giám sát, kiểm soát, phát hiện

và ngăn ngừa rủi ro đối với dữ liệu trong toàn bộ quá trình thu thập, lưu trữ, khai thác và chia sẻ.

Đơn vị vận hành hệ thống có trách nhiệm lựa chọn, triển khai và duy trì các công cụ kỹ thuật phù hợp với quy mô hệ thống, cấp độ nhạy cảm của dữ liệu và yêu cầu quản lý, bảo đảm khả năng tích hợp, vận hành tập trung và phục vụ công tác giám sát, kiểm tra, kiểm toán và xử lý sự cố theo quy định.

Khuyến khích triển khai các công cụ kỹ thuật theo mô hình tập trung. Đối với các cơ sở giáo dục, cơ quan quản lý quy mô nhỏ, ưu tiên kết nối và khai thác các dịch vụ, nền tảng công cụ an toàn dữ liệu dùng chung do Bộ Giáo dục và Đào tạo cung cấp để bảo đảm tính đồng bộ, tối ưu hóa nguồn lực và tiết kiệm chi phí đầu tư.

(2) Danh mục công cụ kỹ thuật bắt buộc:

- Công cụ quản lý siêu dữ liệu và truy vết dữ liệu: phải bảo đảm khả năng quản lý metadata, 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, phục vụ kiểm soát chất lượng dữ liệu và truy vết trong toàn bộ vòng đời.

- Công cụ giám sát sự kiện bảo mật: phải cho phép ghi nhận, phân tích và cảnh báo các sự kiện liên quan đến truy cập hệ thống, truy cập dữ liệu và hành vi bất thường, bảo đảm khả năng giám sát liên tục đối với các hệ thống giáo dục.

- Công cụ ngăn ngừa rò rỉ dữ liệu phải có khả năng phát hiện, kiểm soát và ngăn chặn các hành vi làm lộ dữ liệu thông qua các kênh truyền thông tin, thiết bị lưu trữ hoặc hành vi sao chép trái phép, đặc biệt đối với dữ liệu liên quan đến người học, nhà giáo và cán bộ, nhân viên.

- Công cụ giám sát hoạt động cơ sở dữ liệu phải cho phép theo dõi, ghi nhận và phân tích toàn bộ hoạt động truy cập, truy vấn, chỉnh sửa và khai thác dữ liệu trong các hệ quản trị cơ sở dữ liệu, phục vụ kiểm tra, kiểm toán và phát hiện hành vi bất thường.

- Công cụ bảo vệ dữ liệu trong quá trình khai thác phải hỗ trợ việc che giấu, làm mờ hoặc thay thế dữ liệu nhận dạng, đồng thời tích hợp với cơ chế kiểm soát truy cập nhằm bảo đảm dữ liệu chỉ được hiển thị và sử dụng theo đúng thẩm quyền và mục đích.

- Hệ thống bảng điều khiển tuân thủ phải cung cấp khả năng tổng hợp, hiển thị và theo dõi trạng thái tuân thủ các quy định về quản lý và bảo vệ dữ liệu, hỗ trợ cảnh báo sớm các rủi ro và phục vụ công tác quản lý, giám sát ở các cấp.

h) Lớp 8 – Báo cáo và đánh giá định kỳ

(1) Báo cáo và đánh giá an toàn dữ liệu của cơ sở giáo dục:

- Các cơ sở giáo dục có trách nhiệm thực hiện chế độ báo cáo và đánh giá an toàn dữ liệu nhằm bảo đảm kiểm soát rủi ro và tuân thủ quy định về quản lý, bảo vệ dữ liệu trong phạm vi quản lý.

- Cơ sở giáo dục phải thực hiện báo cáo định kỳ về tình trạng an toàn dữ liệu, bao gồm tối thiểu các nội dung về mức độ tuân thủ quy định, tình hình sự cố an toàn dữ liệu, kết quả triển khai các biện pháp bảo vệ và các rủi ro, tồn tại cần khắc phục; đồng thời phải

tổ chức đánh giá rủi ro dữ liệu tối thiểu một lần trong năm hoặc khi có thay đổi lớn về hệ thống, dữ liệu hoặc hoạt động xử lý dữ liệu.

- Kết quả báo cáo và đánh giá rủi ro phải được lập thành hồ sơ, lưu trữ đầy đủ và gửi cơ quan quản lý có thẩm quyền theo quy định; đồng thời phải làm căn cứ để xây dựng, cập nhật kế hoạch bảo đảm an toàn dữ liệu và khắc phục các điểm yếu, rủi ro đã được xác định.

(2) Trách nhiệm giám sát, kiểm tra của cơ quan quản lý:

- Bộ Giáo dục và Đào tạo có trách nhiệm tổ chức thiết lập và vận hành hệ thống giám sát an toàn dữ liệu tập trung trong toàn ngành giáo dục, bảo đảm khả năng tổng hợp, phân tích, cảnh báo và theo dõi tình trạng tuân thủ của các hệ thống thông tin và cơ sở giáo dục.

- Bộ Giáo dục và Đào tạo thực hiện kiểm tra, thanh tra định kỳ hoặc đột xuất đối với việc tuân thủ các quy định về quản lý và bảo vệ dữ liệu; kịp thời phát hiện, chấn chỉnh và xử lý các hành vi vi phạm, đồng thời hướng dẫn, đôn đốc các đơn vị thực hiện đầy đủ nghĩa vụ báo cáo, đánh giá và khắc phục rủi ro theo quy định.

3. Bảo vệ dữ liệu cốt lõi, dữ liệu quan trọng:

- Dữ liệu cốt lõi, dữ liệu quan trọng phải được áp dụng chế độ bảo vệ ở mức cao nhất trong toàn bộ vòng đời, từ khi thu thập, tạo lập, lưu trữ, xử lý, chia sẻ đến khi xóa, hủy, bảo đảm kiểm soát chặt chẽ rủi ro, ngăn ngừa lộ lọt, mất mát, sai lệch hoặc sử dụng trái phép dữ liệu.

- Bộ phận quản trị dữ liệu có trách nhiệm xây dựng, ban hành và tổ chức thực hiện đầy đủ các quy trình quản lý đối với dữ liệu cốt lõi, dữ liệu quan trọng, bao gồm quy trình thu thập, tạo lập, kiểm tra tính xác thực, giám sát chất lượng, truy xuất nguồn gốc dữ liệu, lưu trữ, sao lưu, phục hồi, khai thác, chia sẻ và xóa, hủy dữ liệu; bảo đảm dữ liệu được kiểm soát xuyên suốt, có khả năng truy vết và đáp ứng yêu cầu kiểm tra, giám sát.

- Việc lưu trữ dữ liệu phải được thực hiện trong hệ thống lưu trữ an toàn, có áp dụng các biện pháp kỹ thuật và tổ chức phù hợp nhằm bảo vệ dữ liệu; phải thiết lập cơ chế sao lưu và phục hồi dữ liệu tự động, bảo đảm dữ liệu quan trọng, dữ liệu cốt lõi được lưu trữ tối thiểu 03 bản sao, tại ít nhất 02 vị trí vật lý khác nhau và có ít nhất 01 bản sao lưu độc lập, tách biệt với hệ thống vận hành.

- Việc truy cập, khai thác và xử lý dữ liệu phải được kiểm soát thông qua cơ chế quản lý truy cập và nhận dạng thống nhất, bảo đảm tuân thủ nguyên tắc đặc quyền tối thiểu; mọi hoạt động truy cập, xử lý dữ liệu phải được ghi nhận đầy đủ dưới dạng nhật ký và được lưu trữ trong thời hạn tối thiểu theo quy định, bảo đảm khả năng kiểm tra, truy vết và phục vụ công tác kiểm toán, điều tra khi cần thiết.

- Đơn vị quản lý dữ liệu có trách nhiệm tổ chức đánh giá rủi ro đối với hoạt động xử lý dữ liệu cốt lõi, dữ liệu quan trọng định kỳ hàng năm hoặc khi có thay đổi lớn; kết quả đánh giá phải được lập thành hồ sơ, lưu trữ đầy đủ và sẵn sàng cung cấp phục vụ hoạt động kiểm tra, thanh tra của cơ quan có thẩm quyền.

- Việc chia sẻ, chuyển giao hoặc xử lý dữ liệu cốt lõi, dữ liệu quan trọng, đặc biệt là trong trường hợp chuyển dữ liệu xuyên biên giới, phải được thực hiện trên cơ sở có đánh giá tác động, xác định rõ mục đích, phương thức, phạm vi xử lý và được cơ quan có thẩm

quyền phê duyệt trước khi thực hiện, trừ trường hợp pháp luật có quy định khác đối với tình huống khẩn cấp.

- Trường hợp ủy thác xử lý dữ liệu, đơn vị quản lý dữ liệu phải thực hiện đánh giá năng lực của bên nhận ủy thác, ký kết thỏa thuận hoặc hợp đồng quy định rõ trách nhiệm, phạm vi xử lý, nghĩa vụ bảo vệ dữ liệu và cơ chế giám sát; đồng thời phải theo dõi, kiểm tra việc thực hiện của bên nhận ủy thác và lưu trữ hồ sơ liên quan theo quy định.

- Trong quá trình xử lý và chia sẻ dữ liệu, phải áp dụng các biện pháp bảo vệ phù hợp như mã hóa, ký số và các biện pháp kỹ thuật khác nhằm bảo đảm tính bí mật, toàn vẹn và khả năng xác thực dữ liệu; bên tiếp nhận dữ liệu có trách nhiệm sử dụng dữ liệu đúng mục đích, phạm vi và thực hiện đầy đủ nghĩa vụ bảo vệ dữ liệu theo thỏa thuận và quy định của pháp luật.

- Dữ liệu phải được xóa, hủy khi hết thời hạn lưu trữ hoặc không còn cần thiết cho mục đích xử lý, bảo đảm không thể khôi phục; việc xóa, hủy phải được ghi nhận, lưu vết và có tài liệu chứng minh phục vụ kiểm tra, giám sát.

- Cán bộ, công chức, viên chức và người lao động được giao nhiệm vụ quản trị, quản lý, bảo vệ và xử lý dữ liệu cốt lõi, dữ liệu quan trọng phải ký cam kết bảo mật và chịu trách nhiệm trước pháp luật về việc thực hiện nghĩa vụ bảo vệ dữ liệu theo quy định.

2.2.5. Quản lý kiến trúc dữ liệu

Mục tiêu của quản lý kiến trúc dữ liệu trong ngành Giáo dục và Đào tạo là bảo đảm xây dựng, duy trì và phát triển kiến trúc dữ liệu thống nhất, đồng bộ, phục vụ hiệu quả việc tích hợp, chia sẻ và khai thác dữ liệu trong toàn ngành.

Kiến trúc dữ liệu phải được mô tả đầy đủ, bao quát các thành phần tối thiểu, bao gồm danh mục các nguồn dữ liệu đầu vào; các quy trình nghiệp vụ có sử dụng dữ liệu; các hệ thống lưu trữ, xử lý, phân tích và cung cấp dữ liệu; đồng thời thể hiện rõ mô hình kiến trúc dữ liệu theo các luồng thu nhận, xử lý và phân phối dữ liệu.

Kiến trúc dữ liệu phải được thẩm định và phê duyệt trước khi triển khai, bảo đảm phù hợp với yêu cầu quản lý, định hướng phát triển và các quy định, tiêu chuẩn liên quan trong toàn ngành.

Việc giám sát và cập nhật kiến trúc dữ liệu phải được thực hiện thường xuyên, đặc biệt trong các trường hợp có thay đổi về cấu trúc dữ liệu, phương thức tích hợp hoặc chia sẻ dữ liệu, nguồn dữ liệu hoặc hệ thống lưu trữ; bảo đảm kiến trúc dữ liệu luôn phản ánh đúng thực trạng và đáp ứng yêu cầu vận hành.

Tài liệu kiến trúc dữ liệu phải được lưu trữ tập trung, bảo đảm khả năng truy cập, khai thác và sử dụng cho các đơn vị liên quan; việc truy cập phải được phân quyền phù hợp với từng nhóm đối tượng và tuân thủ các quy định về an toàn, bảo mật thông tin.

Mọi thay đổi đối với kiến trúc dữ liệu phải được quản lý chặt chẽ thông qua cơ chế quản lý phiên bản, bảo đảm việc theo dõi, kiểm soát và ghi nhận đầy đủ lịch sử thay đổi; đáp ứng yêu cầu truy vết, phục hồi và phục vụ công tác kiểm tra, kiểm toán khi cần thiết.

2.2.6. Quản lý điều phối, chia sẻ, liên thông dữ liệu

Mục tiêu của quản lý điều phối, chia sẻ và liên thông dữ liệu trong ngành Giáo dục và Đào tạo là bảo đảm dữ liệu được kết nối, trao đổi và khai thác hiệu quả giữa các hệ thống trong toàn ngành và với các hệ thống bên ngoài; đồng thời bảo đảm an toàn, bảo mật, kiểm soát truy cập và khả năng giám sát, truy vết trong toàn bộ quá trình sử dụng dữ liệu.

Cơ chế điều phối dữ liệu phải được thiết lập theo hướng tập trung thông qua nền tảng tích hợp, chia sẻ dữ liệu, bảo đảm kết nối với các nền tảng tích hợp, chia sẻ dữ liệu quốc gia và của các bộ, ngành, địa phương; đồng thời bảo đảm khả năng giám sát, truy vết, đo lường hiệu quả khai thác dữ liệu và hỗ trợ xử lý sự cố khi phát sinh.

Việc kết nối, chia sẻ và liên thông dữ liệu phải áp dụng các giao thức, tiêu chuẩn kỹ thuật mở; bảo đảm thống nhất về định dạng, cấu trúc và mô hình dữ liệu; tạo điều kiện thuận lợi cho việc kết nối giữa các hệ thống thông tin của các cơ quan, tổ chức trong và ngoài ngành giáo dục.

Hoạt động chia sẻ dữ liệu phải được thực hiện trên cơ sở phân quyền, phân cấp rõ ràng; có mục đích sử dụng cụ thể, hợp pháp; được kiểm soát chặt chẽ và ghi nhận đầy đủ thông tin phục vụ truy vết, kiểm tra và đánh giá.

Tất cả các hoạt động điều phối, chia sẻ và liên thông dữ liệu phải được giám sát tập trung và tự động; bảo đảm ghi nhận đầy đủ lịch sử truy cập, khai thác và chia sẻ dữ liệu; phục vụ công tác kiểm tra, đánh giá mức độ tuân thủ và nâng cao hiệu quả quản lý dữ liệu trong toàn ngành Giáo dục và Đào tạo.

2.2.7. Quản lý nền tảng dữ liệu

Mục tiêu của quản lý nền tảng dữ liệu trong ngành Giáo dục và Đào tạo là bảo đảm hạ tầng và hệ thống dữ liệu đáp ứng yêu cầu về tính sẵn sàng, khả năng mở rộng, hiệu năng và chất lượng dữ liệu, phục vụ hiệu quả các hoạt động nghiệp vụ, điều hành, phân tích và chia sẻ dữ liệu trong toàn ngành.

Nền tảng dữ liệu ngành giáo dục phải được xây dựng, quản lý và vận hành theo hướng tập trung, bảo đảm kết nối đồng bộ giữa các hệ thống thông tin, hỗ trợ tích hợp, chia sẻ và khai thác dữ liệu thống nhất trong toàn ngành.

Quy trình tích hợp dữ liệu phải được thiết lập và vận hành theo các luồng tích hợp có kiểm soát đầu vào, bao gồm việc xác thực nguồn dữ liệu, kiểm tra định dạng và cấu trúc, phát hiện và xử lý dữ liệu trùng lặp, đồng thời gắn siêu dữ liệu (metadata) để phục vụ nhận dạng, quản lý và truy vết dữ liệu.

Dữ liệu sau tích hợp phải được phân loại và tổ chức lưu trữ theo mô hình các tầng dữ liệu, bao gồm dữ liệu thô, dữ liệu đã xử lý, dữ liệu chuẩn hóa và dữ liệu phục vụ phân tích, nhằm đáp ứng các mục đích khai thác khác nhau và nâng cao hiệu quả sử dụng dữ liệu.

Việc tổ chức lưu trữ dữ liệu phải được thực hiện theo các tầng truy cập phù hợp với mục đích sử dụng, bảo đảm tối ưu hiệu năng và chi phí lưu trữ, đồng thời đáp ứng yêu cầu về khả năng mở rộng, an toàn, bảo mật và truy xuất dữ liệu thuận tiện.

Chính sách quản lý vòng đời dữ liệu và cơ chế di chuyển dữ liệu giữa các tầng lưu trữ phải được thiết lập rõ ràng, có áp dụng tự động hóa và ghi nhận nhật ký theo dõi phục vụ kiểm tra, giám sát.

Dữ liệu phải được chuẩn hóa theo mô hình dữ liệu thống nhất toàn ngành, bao gồm các nội dung về định nghĩa, cấu trúc, kiểu dữ liệu, đơn vị đo lường, mã hóa, định danh và siêu dữ liệu; việc chuẩn hóa phải tuân thủ các tiêu chuẩn, danh mục quốc gia (nếu có) hoặc theo quy định thống nhất trong toàn ngành.

Các công cụ và quy trình kiểm soát chất lượng dữ liệu phải được thiết lập và vận hành theo hướng tự động, tích hợp trong toàn bộ quy trình xử lý dữ liệu; bảo đảm khả năng phát hiện và cảnh báo kịp thời các sai lệch dữ liệu, hỗ trợ cách ly dữ liệu không đạt chất lượng, khôi phục dữ liệu khi cần thiết và phục vụ phân tích nguyên nhân sai lệch nhằm nâng cao chất lượng dữ liệu trong toàn ngành.

2.2.8. Quản lý giám sát đo lường dữ liệu

Mục tiêu của giám sát và đo lường trong quản lý dữ liệu ngành Giáo dục và Đào tạo là theo dõi, đánh giá và cải tiến liên tục hiệu quả quản trị dữ liệu; đồng thời xác định mức độ trưởng thành về quản lý dữ liệu của các đơn vị trên các phương diện như chính sách, quy trình, tổ chức, chất lượng dữ liệu, công nghệ và năng lực vận hành.

Hệ thống chỉ số giám sát và đo lường phải được xây dựng và áp dụng thống nhất trong toàn ngành, bao gồm các chỉ số đánh giá hiệu quả (KPI) và chỉ số chất lượng (KQI), làm cơ sở theo dõi, đánh giá và cải thiện hoạt động quản trị dữ liệu.

a) Bộ chỉ số đánh giá bao gồm:

(1) *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

			duyet 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 (%)	Phân 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 liệu được cập nhật (%)	Phần trăm chính sách, tiêu chuẩn quản trị dữ liệu được ban hành, duy trì và tuân thủ	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
			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 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

(2) 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			0: <10% yêu cầu

	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)	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 cập nhật định kỳ (%)	Phần trăm bản ghi metadata được duy trì, cập nhật đầy đủ, đúng kỳ	0: <10% metadata
			1: 10-30% cập nhật
			2: 30-50% cập nhật
			3: 50-70% cập nhật
			4: 70-99% cập nhật tự động

			5: 100% đồng bộ tự động, cảnh báo lỗi tự động
--	--	--	---

(3) 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, giảm 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 được tự động hóa (%)	Phần trăm các quy trình quản trị, vận hành, báo cáo được thực hiện tự động, tự giám sát	0: 0% tự động
			1: 10-30% tự động
			2: 30-50% tự động
			3: 50-70% tự động
			4: 70-99% tự động
			5: 100% tự động và tối ưu liên tục

b) 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

Việc giám sát và đo lường phải được thực hiện thường xuyên, liên tục, bảo đảm khả năng theo dõi, so sánh và đánh giá giữa các đơn vị; kết quả đo lường là cơ sở để xác định mức độ trưởng thành về quản lý dữ liệu, phục vụ công tác chỉ đạo, điều hành và cải tiến hoạt động quản trị dữ liệu trong toàn ngành.

Kết quả giám sát và đo lường phải được tổng hợp, phân tích và sử dụng làm căn cứ để điều chỉnh chính sách, quy trình, nâng cao chất lượng dữ liệu, hoàn thiện hệ thống công nghệ và nâng cao năng lực quản trị dữ liệu của các đơn vị.

2.2.9. Quản lý kho dữ liệu và phân tích dữ liệu

Mục tiêu của quản lý kho dữ liệu và phân tích dữ liệu trong ngành Giáo dục và Đào tạo là xây dựng nền tảng dữ liệu tập trung phục vụ tích hợp, tổng hợp và khai thác dữ liệu từ nhiều nguồn khác nhau, hỗ trợ hiệu quả công tác báo cáo, phân tích, dự báo và ra quyết định.

Kho dữ liệu ngành giáo dục phải được xây dựng theo mô hình phục vụ phân tích, cho phép tích hợp dữ liệu từ các hệ thống thông tin khác nhau; bảo đảm tính nhất quán, toàn vẹn và khả năng mở rộng nhằm đáp ứng nhu cầu khai thác dữ liệu trong toàn ngành.

Mô hình dữ liệu và phân tích nghiệp vụ phải được thiết kế đồng bộ với kho dữ liệu, bảo đảm phản ánh đúng các yêu cầu quản lý; việc khai thác dữ liệu phải sử dụng các công cụ, kỹ thuật và dịch vụ phân tích phù hợp nhằm tạo ra thông tin có giá trị phục vụ công tác lãnh đạo, quản lý và vận hành.

Các chỉ tiêu phân tích, cấu trúc báo cáo và logic nghiệp vụ phải được thiết lập thống nhất, 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.

Quy trình xác định, tính toán và đối soát các chỉ tiêu phân tích phải được thiết lập và thực hiện trên cơ sở dữ liệu tập trung, tuân thủ mô hình logic thống nhất; việc tổng hợp

chỉ tiêu phải dựa trên các quy tắc nghiệp vụ chuẩn, có luồng xử lý dữ liệu minh bạch và khả năng kiểm soát.

Các chỉ tiêu phân tích sau khi tổng hợp phải được lưu trữ theo các chiều thời gian, đơn vị tổ chức và lĩnh vực chuyên ngành, hỗ trợ khai thác và phân tích đa chiều phục vụ công tác điều hành và đánh giá.

Hệ thống chỉ tiêu phân tích phải được tổ chức thành danh mục thống nhất, bao gồm mã định danh, định nghĩa nghiệp vụ, đơn vị tính, tần suất cập nhật và quy tắc tính toán; được phân loại theo các nhóm mục tiêu như điều hành, giám sát, đánh giá hiệu quả, dự báo và cảnh báo sớm. Mọi thay đổi liên quan đến cấu trúc hoặc logic chỉ tiêu phải được kiểm soát thông qua quy trình quản lý thay đổi, bảo đảm tính thống nhất trong toàn hệ thống.

Việc ứng dụng AI trong phân tích dữ liệu giáo dục phải tuân thủ nguyên tắc AI có trách nhiệm quy định tại mục 2.2.4

Các kho dữ liệu chuyên biệt (data mart) phải được tổ chức nhằm phục vụ nhu cầu phân tích theo từng lĩnh vực hoặc cấp quản lý; việc triển khai phải bảo đảm tuân thủ mô hình dữ liệu logic và vật lý thống nhất, đồng thời bảo đảm khả năng tích hợp, mở rộng và khai thác linh hoạt trong toàn ngành Giáo dục và Đào tạo.

2.2.10. Quản lý các dự án CNTT

Mục tiêu của quản lý dự án công nghệ thông tin liên quan đến dữ liệu trong ngành Giáo dục và Đào tạo là bảo đảm các chương trình, đề án, dự án được triển khai đồng bộ, hiệu quả, phù hợp với định hướng quản trị dữ liệu; đồng thời kiểm soát rủi ro, bảo đảm chất lượng và an toàn dữ liệu trong suốt quá trình thực hiện.

Quản lý dự án công nghệ thông tin được thực hiện xuyên suốt các giai đoạn từ lập kế hoạch, triển khai, giám sát đến đánh giá kết quả đối với các chương trình, đề án, dự án có liên quan đến dữ liệu trong toàn ngành Giáo dục và Đào tạo.

Mọi dự án công nghệ thông tin có cấu phần dữ liệu phải được đánh giá tuân thủ Khung kiến trúc dữ liệu và Khung quản trị dữ liệu; thiết kế cơ sở dữ liệu phải tuân thủ nghiêm ngặt Từ điển dữ liệu và Danh mục dùng chung của ngành Giáo dục và Đào tạo.

Trong quá trình triển khai dự án, quy trình kiểm soát thay đổi phải được thiết lập và thực hiện thống nhất nhằm quản lý các thay đổi về phạm vi, yêu cầu, kiến trúc, dữ liệu và giải pháp kỹ thuật; bảo đảm các thay đổi được đánh giá tác động, phê duyệt và cập nhật đầy đủ trước khi thực hiện.

Các rủi ro liên quan đến dữ liệu trong dự án phải được xác định, đánh giá và quản lý, bao gồm các rủi ro về rò rỉ dữ liệu, sai lệch dữ liệu, lỗi tích hợp hoặc gián đoạn hệ thống; đồng thời phải xây dựng và triển khai các phương án phòng ngừa, phát hiện và xử lý phù hợp.

Việc giám sát và đánh giá dự án phải gắn với các tiêu chí về chất lượng dữ liệu, mức độ tuân thủ kiến trúc dữ liệu, khả năng tích hợp, chia sẻ và khai thác dữ liệu; bảo đảm kết quả dự án đáp ứng yêu cầu quản lý dữ liệu của ngành.

Kết quả triển khai dự án phải được kiểm tra, đánh giá và nghiệm thu trên cơ sở các tiêu chí đã được xác định. Khi nghiệm thu, bắt buộc phải bàn giao đầy đủ siêu dữ liệu (metadata), nhật ký vòng đời dữ liệu (data lineage) và tài liệu quản trị dữ liệu liên quan để tích hợp vào hệ thống quản lý chung của Bộ. Việc này đồng thời phục vụ công tác tổng kết, rút kinh nghiệm và nâng cao hiệu quả quản lý các dự án công nghệ thông tin liên quan đến dữ liệu trong toàn ngành Giáo dục và Đào tạo.

2.2.11. Quản lý khai thác gia tăng giá trị dữ liệu

Mục tiêu của khai thác và phát triển giá trị dữ liệu trong ngành Giáo dục và Đào tạo là thúc đẩy việc sử dụng dữ liệu một cách hiệu quả, minh bạch; tạo điều kiện cho người dân, doanh nghiệp tiếp cận dữ liệu; đồng thời hình thành các sản phẩm, dịch vụ dữ liệu phục vụ phát triển kinh tế số và hệ sinh thái giáo dục số.

Danh mục dữ liệu mở phải được xây dựng, công bố và cập nhật theo quy định, bảo đảm phù hợp với danh mục dữ liệu mở đã được phê duyệt; đáp ứng yêu cầu công khai, minh bạch và khả năng khai thác, sử dụng của các tổ chức, cá nhân.

Dữ liệu mở phải được cung cấp thông qua các nền tảng công bố dữ liệu theo quy định, bảo đảm khả năng truy cập thuận tiện, định dạng tiêu chuẩn, dễ sử dụng và hỗ trợ khai thác hiệu quả cho người dân, doanh nghiệp và các tổ chức có nhu cầu.

Việc khai thác dữ liệu phải được tổ chức theo hướng tạo giá trị gia tăng, thúc đẩy hình thành các sản phẩm, dịch vụ dữ liệu trong lĩnh vực giáo dục và đào tạo; khuyến khích đổi mới sáng tạo, phát triển các ứng dụng, giải pháp sử dụng dữ liệu phục vụ quản lý, giảng dạy, học tập và nhu cầu của xã hội.

Việc ứng dụng AI để khai thác gia tăng giá trị của dữ liệu giáo dục phải tuân thủ nguyên tắc AI có trách nhiệm quy định tại mục 2.2.4.

Hoạt động khai thác và sử dụng dữ liệu phải tuân thủ các quy định về bảo vệ dữ liệu, an toàn, bảo mật thông tin và quyền riêng tư; đồng thời bảo đảm sử dụng dữ liệu đúng mục đích, hiệu quả và bền vững trong toàn ngành Giáo dục và Đào tạo.

2.2.12. Quản lý dịch vụ dữ liệu

Mục tiêu của quản lý dịch vụ dữ liệu trong ngành Giáo dục và Đào tạo là bảo đảm cung cấp các dịch vụ dữ liệu ổn định, an toàn, chất lượng và đáp ứng nhu cầu khai thác, sử dụng dữ liệu của các cơ quan, tổ chức và cá nhân.

Quy trình quản lý dịch vụ dữ liệu phải được xây dựng và vận hành thống nhất, bao gồm kiểm soát truy cập, giám sát chất lượng dữ liệu và bảo vệ thông tin; bảo đảm việc cung cấp và sử dụng dịch vụ dữ liệu tuân thủ các quy định về an toàn, bảo mật và quyền riêng tư.

Hạ tầng kỹ thuật phục vụ dịch vụ dữ liệu phải được duy trì ổn định, an toàn, có khả năng mở rộng linh hoạt theo nhu cầu khai thác; bảo đảm tính sẵn sàng và hiệu năng của các dịch vụ dữ liệu trong toàn ngành.

Mô hình dữ liệu và danh mục dữ liệu phải được rà soát, cập nhật định kỳ nhằm bảo đảm tính nhất quán, đồng bộ và khả năng liên thông giữa các hệ thống cung cấp dịch vụ dữ liệu.

Hoạt động hỗ trợ người dùng phải được tổ chức đầy đủ, bao gồm hướng dẫn truy cập, sử dụng và khai thác dịch vụ dữ liệu; đồng thời thiết lập cơ chế tiếp nhận, xử lý sự cố và phản hồi các yêu cầu cải tiến dịch vụ nhằm nâng cao chất lượng cung cấp dịch vụ dữ liệu trong toàn ngành Giáo dục và Đào tạo.

3. Lớp vận hành dữ liệu

3.1. Vận hành vòng đời dữ liệu

Quản lý vòng đời dữ liệu trong ngành Giáo dục và Đào tạo phải được thực hiện theo nguyên tắc quản lý tập trung, xuyên suốt, bảo đảm dữ liệu được kiểm soát từ khi hình thành, thu thập, xử lý, lưu trữ, khai thác đến khi xóa, hủy theo quy trình thống nhất và chuẩn hóa trong toàn ngành.

Toàn bộ quá trình quản lý vòng đời dữ liệu phải bảo đảm an toàn, bảo mật và tuân thủ các quy định pháp luật có liên quan, đặc biệt là các quy định về an ninh mạng và bảo vệ dữ liệu cá nhân; áp dụng các biện pháp kiểm soát phù hợp với mức độ nhạy cảm của dữ liệu.

Dữ liệu phải được quản lý theo hướng tối ưu hóa giá trị, thông qua việc chuẩn hóa, làm sạch, tích hợp, phân tích và chia sẻ dữ liệu, nhằm phục vụ hiệu quả công tác quản lý, điều hành và phát triển các dịch vụ số trong ngành Giáo dục và Đào tạo.

Trách nhiệm giải trình trong quản lý dữ liệu phải được xác định rõ ràng, theo đó các cơ quan, đơn vị tham gia quản lý và khai thác dữ liệu chịu trách nhiệm về tính chính xác, đầy đủ, kịp thời và an toàn của dữ liệu trong phạm vi quản lý.

Cơ chế giám sát, phân quyền truy cập và ghi nhận nhật ký phải được thiết lập và vận hành đầy đủ, bảo đảm khả năng kiểm soát, truy vết toàn bộ các hoạt động liên quan đến dữ liệu; đồng thời các quy định về quản lý vòng đời dữ liệu phải được ban hành và tổ chức thực hiện thống nhất trong toàn ngành.

3.2. Hạ tầng kỹ thuật quản trị dữ liệu

Mục tiêu của quản lý hạ tầng kỹ thuật và công cụ hỗ trợ dữ liệu trong ngành Giáo dục và Đào tạo là bảo đảm vận hành ổn định, an toàn, hiệu quả các hệ thống công nghệ phục vụ quản trị, quản lý và khai thác dữ liệu trong toàn ngành.

Hạ tầng kỹ thuật và công nghệ phải được tổ chức vận hành bảo đảm tính sẵn sàng, an toàn, bảo mật; thực hiện cấp phát và quản lý tài nguyên hợp lý, đáp ứng nhu cầu xử lý,

lưu trữ và khai thác dữ liệu; đồng thời hỗ trợ khả năng mở rộng linh hoạt theo yêu cầu thực tiễn.

Việc ứng dụng các công nghệ và giải pháp kỹ thuật phải được tăng cường nhằm nâng cao hiệu quả quản trị dữ liệu, bao gồm các công nghệ hỗ trợ tích hợp, xử lý, phân tích, giám sát và bảo vệ dữ liệu trong toàn ngành.

Các công cụ hỗ trợ quản trị, quản lý và vận hành dữ liệu phải được trang bị và triển khai theo hướng tự động hóa, bảo đảm hỗ trợ hiệu quả cho các hoạt động kiểm soát, giám sát, phân tích và điều hành dữ liệu; góp phần nâng cao năng lực quản trị dữ liệu và tối ưu hóa vận hành hệ thống.

V. NHIỆM VỤ TRỌNG TÂM

Để triển khai hiệu quả Khung quản trị, quản lý dữ liệu vào thực tiễn, Bộ Giáo dục và Đào tạo xác định các nhóm nhiệm vụ trọng tâm theo từng giai đoạn căn cứ trên các lớp chức năng của Khung như sau:

(1) Giai đoạn 1 (2026-2027) : Kiện toàn tổ chức và thiết lập nền tảng:

- Kiện toàn Mô hình tổ chức quản trị dữ liệu 2 cấp:

+ Ở Cấp 1 (Bộ GDĐT): Phát huy vai trò của Ban Chỉ đạo dữ liệu ngành, chỉ định Người chịu trách nhiệm quản trị, quản lý dữ liệu. Thiết lập và giao nhiệm vụ cho 05 bộ phận chuyên trách: Bộ phận tuân thủ, Bộ phận quản trị dữ liệu, Bộ phận chủ quản hệ thống, Bộ phận kỹ thuật dữ liệu, Bộ phận đổi mới sáng tạo.

+ Ở Cấp 2 (Sở GDĐT/Cục/Vụ/Viện/Trường): Xác định rõ vai trò Quản lý dữ liệu cấp đơn vị; thiết lập các Bộ phận chủ quản hệ thống, Bộ phận quản trị dữ liệu cấp đơn vị và Bộ phận tạo lập dữ liệu tại đơn vị.

- Xây dựng và tiếp tục hoàn thiện danh mục dữ liệu dùng chung thống nhất và từ điển nghiệp vụ giáo dục đào tạo làm cơ sở để tổ chức, phân loại, làm sạch dữ liệu.

(2) Giai đoạn 2 (2027-2030): Tối ưu hóa, đổi mới sáng tạo và xây dựng văn hóa dữ liệu:

- Xây dựng và lan tỏa văn hóa dữ liệu: Chuyển đổi nhận thức để dữ liệu thực sự trở thành tài sản chiến lược làm cơ sở ra quyết định. Gắn tiêu chí chất lượng và mức độ sử dụng dữ liệu vào hệ thống đo lường, đánh giá năng lực của các đơn vị toàn ngành.

- Mở rộng hệ thống giám sát và liên thông: Hoàn thiện liên thông dữ liệu hai chiều từ cấp cơ sở đến trung ương. Triển khai các chuyên đề giám sát nâng cao: giám sát kho dữ liệu phân tích, giám sát các dự án CNTT và giám sát việc khai thác, gia tăng giá trị thặng dư từ dữ liệu.

- Mục tiêu đo lường: Đánh giá định kỳ hàng năm mức độ trưởng thành quản trị dữ liệu theo Quyết định số 2439/QĐ-TTg; phấn đấu đến năm 2030, Bộ Giáo dục và Đào tạo, các Sở Giáo dục và Đào tạo đạt mức độ trưởng thành từ Mức 4 trở lên; mặt bằng chung toàn ngành đạt từ Mức 3 trở lên.

VI. TỔ CHỨC THỰC HIỆN

1. Cục Khoa học, Công nghệ và Thông tin

- Chủ trì, phối hợp với các đơn vị duy trì và tổ chức triển khai Khung quản trị, quản lý dữ liệu Giáo dục và Đào tạo được phê duyệt tại Quyết định này; cập nhật, điều chỉnh Khung quản trị, quản lý dữ liệu phù hợp với điều kiện thực tế trong từng thời kỳ;

- Chủ trì, phối hợp với các đơn vị liên quan tổ chức xây dựng, phát triển các cơ sở dữ liệu ngành Giáo dục và Đào tạo theo Khung quản trị, quản lý dữ liệu và Khung kiến trúc dữ liệu đã được phê duyệt; bảo đảm liên thông với Cơ sở dữ liệu quốc gia, Cơ sở dữ liệu tổng hợp quốc gia và các hệ thống thông tin liên quan;

- Chủ trì, phối hợp với các đơn vị xây dựng các tiêu chuẩn kỹ thuật, quy chế, quy trình chi tiết về quản trị, quản lý dữ liệu; chủ trì xây dựng, trình cấp có thẩm quyền ban hành hoặc ban hành theo thẩm quyền các văn bản quy định, hướng dẫn, tiêu chuẩn kỹ thuật liên quan đến kết nối, liên thông, chia sẻ dữ liệu giữa các hệ thống thông tin trong và ngoài ngành Giáo dục và Đào tạo;

- Chủ trì, phối hợp với các đơn vị xây dựng kế hoạch và chi tiết hóa lộ trình phát triển từng miền dữ liệu của ngành Giáo dục và Đào tạo; xây dựng kế hoạch phối hợp với các đơn vị cung cấp và sử dụng dữ liệu trong việc truy xuất, tích hợp, bổ sung, chỉnh sửa dữ liệu; hướng dẫn, điều phối các hoạt động truy cập và sử dụng dữ liệu;

- Chủ trì phối hợp và hướng dẫn các Cục, Vụ chuyên môn và Sở Giáo dục và Đào tạo trong việc triển khai các hoạt động quản trị, quản lý dữ liệu theo Khung quản trị, quản lý dữ liệu Giáo dục và Đào tạo được phê duyệt tại Quyết định này;

- Tổng hợp, trình cấp có thẩm quyền bố trí vốn đầu tư công, kinh phí chi thường xuyên trong dự toán ngân sách nhà nước hoặc nguồn thu hợp pháp khác để triển khai các dự án, nhiệm vụ ứng dụng công nghệ thông tin phục vụ quản trị, quản lý dữ liệu ngành Giáo dục và Đào tạo, phù hợp với Khung quản trị, quản lý dữ liệu, lộ trình phát triển dữ liệu ngành và quy định của pháp luật có liên quan;

2. Các Cục, Vụ chuyên môn thuộc Bộ Giáo dục và Đào tạo

- Tuân thủ chặt chẽ Khung quản trị, quản lý dữ liệu Giáo dục và Đào tạo được phê duyệt tại Quyết định này; các quy định của Bộ về quản trị dữ liệu; các tiêu chuẩn, quy trình dữ liệu do Cục Khoa học, Công nghệ và Thông tin ban hành hoặc hướng dẫn;

- Thực hiện cung cấp dữ liệu trong miền dữ liệu được phân công một cách đầy đủ, chính xác, kịp thời theo quy định; chịu trách nhiệm về tính pháp lý và chất lượng dữ liệu thuộc phạm vi quản lý;

- Chủ động phối hợp với Cục Khoa học, Công nghệ và Thông tin trong xây dựng, triển khai, nâng cấp các hệ thống thông tin, cơ sở dữ liệu chuyên ngành thuộc phạm vi quản lý; bảo đảm liên thông, tích hợp, đồng bộ và chia sẻ dữ liệu nguồn với hệ thống dữ liệu tập trung của Bộ;

- Chủ động phối hợp với Cục Khoa học, Công nghệ và Thông tin trong việc triển khai các hoạt động quản trị dữ liệu tại đơn vị;

- Phân công cán bộ phụ trách và hình thành bộ phận quản trị dữ liệu, bộ phận chủ quản hệ thống tại Cục/Vụ theo mô hình quy định tại Khung quản trị, quản lý dữ liệu; thông báo thông tin phân công về Cục Khoa học, Công nghệ và Thông tin để tổng hợp, theo dõi;

- Chủ động tham gia các chương trình đào tạo, bồi dưỡng về quản trị, quản lý dữ liệu do Bộ tổ chức; định kỳ báo cáo tình hình quản lý, chất lượng và khai thác dữ liệu thuộc phạm vi phụ trách; kịp thời phản ánh khó khăn, vướng mắc trong quá trình thực hiện.

3. Các Sở Giáo dục và Đào tạo tỉnh, thành phố trực thuộc Trung ương

- Tổ chức triển khai Khung quản trị, quản lý dữ liệu Giáo dục và Đào tạo tại địa phương; tuân thủ các quy định, tiêu chuẩn, quy trình do Bộ Giáo dục và Đào tạo ban hành; phối hợp chặt chẽ với Cục Khoa học, Công nghệ và Thông tin trong quá trình triển khai. Chủ động tham mưu, phối hợp với Ủy ban nhân dân các cấp và các cơ sở giáo dục, đào tạo trên địa bàn để thống nhất chỉ đạo, hướng dẫn trong công tác quản trị và tổ chức tạo lập dữ liệu. Bảo đảm cung cấp dữ liệu đầy đủ, chính xác, kịp thời thuộc phạm vi địa phương quản lý theo yêu cầu của Bộ; chịu trách nhiệm về nội dung, chất lượng và tính pháp lý của dữ liệu được cung cấp;

- Phân công cán bộ phụ trách và hình thành bộ phận quản trị dữ liệu, bộ phận chủ quản hệ thống tại Sở theo mô hình quản trị dữ liệu quy định tại Khung quản trị, quản lý dữ liệu; gửi thông tin phân công về Bộ (qua Cục Khoa học, Công nghệ và Thông tin) để theo dõi, tổng hợp. Trường hợp thay đổi người phụ trách, phải thông báo kịp thời về Bộ;

- Chủ động triển khai các hệ thống thông tin, cơ sở dữ liệu tại địa phương phù hợp với kiến trúc tổng thể và Khung quản trị, quản lý dữ liệu của ngành; bảo đảm kết nối, liên thông, đồng bộ và chia sẻ dữ liệu với hệ thống dữ liệu tập trung cấp Bộ;

- Chủ động bố trí nguồn lực, tổ chức đào tạo, bồi dưỡng nâng cao năng lực quản trị, quản lý dữ liệu cho cán bộ tại đơn vị; xây dựng và phát triển văn hóa dữ liệu tại địa phương theo định hướng của Bộ Giáo dục và Đào tạo;

- Định kỳ hoặc đột xuất báo cáo tình hình triển khai Khung quản trị, quản lý dữ liệu tại địa phương theo yêu cầu của Bộ; chủ động đề xuất, kiến nghị Bộ Giáo dục và Đào tạo xem xét, xử lý các vấn đề phát sinh trong quá trình thực hiện.

4. Các Viện và Trường Đại học, Cao đẳng thuộc Bộ Giáo dục và Đào tạo

- Tổ chức triển khai Khung quản trị, quản lý dữ liệu Giáo dục và Đào tạo tại đơn vị; tuân thủ các quy định, tiêu chuẩn, quy trình do Bộ Giáo dục và Đào tạo ban hành; phối hợp chặt chẽ với Cục Khoa học, Công nghệ và Thông tin trong quá trình triển khai. Bảo đảm cung cấp dữ liệu đầy đủ, chính xác, kịp thời thuộc phạm vi đơn vị quản lý theo yêu cầu của Bộ; chịu trách nhiệm về nội dung, chất lượng và tính pháp lý của dữ liệu được cung cấp;

- Phân công cán bộ phụ trách và hình thành bộ phận quản trị dữ liệu, bộ phận chủ quản hệ thống tại đơn vị theo mô hình quản trị dữ liệu quy định tại Khung quản trị, quản lý dữ liệu; gửi thông tin phân công về Bộ (qua Cục Khoa học, Công nghệ và Thông tin) để theo dõi, tổng hợp. Trường hợp thay đổi người phụ trách, phải thông báo kịp thời về Bộ;

- Chủ động triển khai các hệ thống thông tin, cơ sở dữ liệu tại đơn vị phù hợp với kiến trúc tổng thể và Khung quản trị, quản lý dữ liệu của ngành; bảo đảm kết nối, liên thông, đồng bộ và chia sẻ dữ liệu với hệ thống dữ liệu tập trung cấp Bộ;

- Chủ động bố trí nguồn lực, tổ chức đào tạo, bồi dưỡng nâng cao năng lực quản trị, quản lý dữ liệu cho cán bộ tại đơn vị; xây dựng và phát triển văn hóa dữ liệu tại đơn vị theo định hướng của Bộ Giáo dục và Đào tạo;

- Định kỳ hoặc đột xuất báo cáo tình hình triển khai Khung quản trị, quản lý dữ liệu tại đơn vị theo yêu cầu của Bộ; chủ động đề xuất, kiến nghị Bộ Giáo dục và Đào tạo xem xét, xử lý các vấn đề phát sinh trong quá trình thực hiện.

PHỤ LỤC I

(Ban hành kèm theo Khung Quản trị, quản lý dữ liệu tại Quyết định số /QĐ-BGDĐT
ngày tháng 6 năm 2026 của Bộ Giáo dục và Đào tạo)

1. Danh sách thành viên Ban Chỉ đạo về dữ liệu Bộ Giáo dục và Đào tạo

TT	Chức vụ người phụ trách	Đơn vị thuộc Bộ Giáo dục & Đào tạo	Vai trò trong Ban chỉ đạo
1	Bộ trưởng	Bộ Giáo dục và Đào tạo	Trưởng ban chỉ đạo
2	Thứ trưởng Thường trực	Bộ Giáo dục và Đào tạo	Phó Trưởng Ban Thường trực Ban Chỉ đạo
3	Thứ trưởng	Bộ Giáo dục và Đào tạo	Phó Trưởng Ban Chỉ đạo
4	Thứ trưởng	Bộ Giáo dục và Đào tạo	Phó Trưởng Ban Chỉ đạo
5	Cục trưởng	Cục Khoa học, Công nghệ và Thông tin	Thành viên
6	Vụ trưởng	Vụ Tổ chức cán bộ	Thành viên
7	Vụ trưởng	Vụ Kế hoạch - Tài chính	Thành viên
8	Vụ trưởng	Vụ Giáo dục Mầm non	Thành viên
9	Vụ trưởng	Vụ Giáo dục Phổ thông	Thành viên
10	Vụ trưởng	Vụ Giáo dục Đại học	Thành viên
11	Vụ trưởng	Vụ Học sinh, sinh viên	Thành viên
12	Vụ trưởng	Vụ Giáo dục Quốc phòng và An ninh	Thành viên
13	Vụ trưởng	Vụ Pháp chế	Thành viên

14	Chánh Văn phòng	Văn phòng	Thành viên
15	Cục trưởng	Cục Nhà giáo và Cán bộ quản lý giáo dục	Thành viên
16	Cục trưởng	Cục Quản lý chất lượng.	Thành viên
17	Cục trưởng	Cục Hợp tác quốc tế	Thành viên
18	Cục trưởng	Cục Giáo dục nghề nghiệp và Giáo dục thường xuyên	Thành viên
19	Giám đốc	Ban Quản lý các dự án Bộ Giáo dục và Đào tạo.	Thành viên

2. Phân công vai trò quản trị dữ liệu Cấp 1

STT	Vai trò	Đơn vị	Chức vụ người chịu trách nhiệm chính
1	Ban chỉ đạo dữ liệu về Giáo dục và Đào tạo	Tham chiếu Phụ lục 1	Tham chiếu Phụ lục 1
2	Người chịu trách nhiệm quản trị, quản lý dữ liệu	Cục Khoa học, Công nghệ và Thông tin	Cục trưởng
3	Bộ phận tuân thủ	Cục Khoa học, Công nghệ và Thông tin là thành viên thường trực; Các Cục/Vụ chuyên môn	
4	Bộ phận quản trị dữ liệu	Cục Khoa học, Công nghệ và Thông tin	
5	Bộ phận chủ quản hệ thống	Cục Khoa học, Công nghệ và Thông tin	

6	Bộ phận kỹ thuật dữ liệu	Cục Khoa học, Công nghệ và Thông tin	
7	Bộ phận đổi mới sáng tạo	Cục Khoa học, Công nghệ và Thông tin	

3. Phân công vai trò quản trị dữ liệu Cấp 2

Mỗi Sở Giáo dục và Đào tạo các Tỉnh/Thành phố; các Cục Vụ chuyên môn và Viện, trường Đại học, Cao đẳng thuộc Bộ Giáo dục và Đào tạo ban hành theo mẫu và gửi về Cục Khoa học, Công nghệ và Thông tin để tổng hợp, theo dõi.

STT	Vai trò	Tên Sở/Cục/Vụ	Họ và tên người phụ trách	Chức vụ người phụ trách	Số điện thoại	Email	Ngày có hiệu lực
1	Quản lý dữ liệu Cấp đơn vị						
2	Bộ phận quản trị dữ liệu cấp đơn vị						
3	Bộ phận chủ quản hệ thống						
4	Bộ phận đổi mới sáng tạo						

