BỘ THÔNG TIN VÀ TRUYỀN THÔNG

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

**VIỆN KHOA HỌC KỸ THUẬT BƯU ĐIỆN**

**THUYẾT MINH**

**DỰ THẢO TIÊU CHUẨN QUỐC GIA**

**VỀ CÔNG NGHỆ THÔNG TIN – KIẾN TRÚC THAM CHIẾU DỮ LIỆU LỚN**

**PHẦN 1: KHUNG VÀ QUY TRÌNH ỨNG DỤNG**

**Ký hiệu: TCVN ISO/IEC TR 20547-1:2022**

**HÀ NỘI - 2022**

# MỤC LỤC

[MỤC LỤC 1](#_Toc127633105)

[DANH MỤC HÌNH VẼ 1](#_Toc127633106)

[DANH MỤC BẢNG BIỂU 2](#_Toc127633107)

[DANH MỤC CÁC CHỮ VIẾT TẮT 3](#_Toc127633108)

[1. Giới thiệu dự thảo TCVN 4](#_Toc127633109)

[1.1. Tên dự thảo theo đề cương 4](#_Toc127633110)

[1.2. Mục tiêu 4](#_Toc127633111)

[2. Đặt vấn đề 4](#_Toc127633112)

[3. Tình hình chuẩn hoá Kiến trúc tham chiếu dữ liệu lớn – Khung và quy trình ứng dụng trên thế giới 6](#_Toc127633113)

[3.1. Các tổ chức tiêu chuẩn hoá quốc tế 7](#_Toc127633116)

[3.1.1. Tổ chức ISO/IEC 7](#_Toc127633117)

[3.1.2. Liên minh viễn thông quốc tế ITU 8](#_Toc127633118)

[3.1.3. Viện Tiêu chuẩn và Công nghệ Quốc gia Mỹ (NIST) 9](#_Toc127633119)

[4. Tình hình chuẩn hoá Kiến trúc tham chiếu dữ liệu lớn – Khung và quy trình ứng dụng tại Việt Nam 10](#_Toc127633120)

[4.1. TCVN 13238:2020 (ISO/IEC 20546:2019) 10](#_Toc127633121)

[4.2. TCVN 13239-2:2020 (ISO/IEC TR 20547-2:2018) 10](#_Toc127633122)

[4.3. Dự thảo TCVN ISO/IEC 20547-3:2022: Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 3: Kiến trúc tham chiếu 10](#_Toc127633123)

[4.4. TCVN 13239-5:2020 (ISO/IEC TR 20547-5:2018) 11](#_Toc127633124)

[4.5. Nhận xét chung 18](#_Toc127633125)

[5. Xây dựng bộ tiêu chuẩn quốc gia về Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn – Phần 1: Khung và Quy trình ứng dụng (dựa trên ISO/IEC TR 20547-1:2020) 20](#_Toc127633126)

[5.1. Lựa chọn tài liệu tham khảo 20](#_Toc127633129)

[5.2. Cách thức xây dựng 21](#_Toc127633130)

[5.3. Về hình thức trình bày 21](#_Toc127633131)

[5.4. Tên dự thảo tiêu chuẩn 21](#_Toc127633132)

[5.5. Nội dung dự thảo tiêu chuẩn 22](#_Toc127633133)

[TÀI LIỆU THAM KHẢO 25](#_Toc127633134)

# DANH MỤC HÌNH VẼ

[Hình 1: Các mối quan hệ giữa các phần của bộ tiêu chuẩn ISO/IEC 20547 18](#_Toc127633135)

# DANH MỤC BẢNG BIỂU

[Bảng 1 - Bảng tổng hợp tình hình tiêu chuẩn hoá Công nghệ thông tin – Kiến trúc tham chiếu dữ liệu lớn 13](#_Toc119025779)

[Bảng 2 - Bảng đối chiếu nội dung TCVN và tài liệu tham khảo 22](#_Toc119025780)

# DANH MỤC CÁC CHỮ VIẾT TẮT

|  |  |
| --- | --- |
| DLL | Dữ liệu lớn |
| KH&CN | Khoa học và Công nghệ |
| IEC | Uỷ ban kỹ thuật điện quốc tế |
| ISO | Tổ chức tiêu chuẩn hoá quốc tế |
| ITU | Liên minh viên thông quốc tế |
| NIST | Viện Tiêu chuẩn và Công nghệ Quốc gia Mỹ |
| TTg | Thủ tướng |
| TCVN | Tiêu chuẩn quốc gia |
| TT&TT | Thông tin và Truyền thông |

**THUYẾT MINH**

**DỰ THẢO TIÊU CHUẨN KỸ QUỐC GIA**

**VỀ CÔNG NGHỆ THÔNG TIN – KIẾN TRÚC THAM CHIẾU DỮ LIỆU LỚN**

**PHẦN 1: KHUNG VÀ QUY TRÌNH ỨNG DỤNG**

*Information technology - Big data reference architecture –*

*Part 1: Framework and application process*

# Giới thiệu dự thảo TCVN

## Tên dự thảo theo đề cương

Tiêu chuẩn quốc gia về công nghệ thông tin – Khung tham chiếu dữ liệu lớn – Phần 1: Khung và quy trình ứng dụng.

Ký hiệu: TCVN ISO/IEC TR 20547-1:2022.

## Mục tiêu

Phục vụ công tác chuẩn hóa về kiến trúc tham chiếu dữ liệu lớn.

# Đặt vấn đề

Thủ tướng Chính phủ đã ban hành Quyết định số 749/QĐ-TTg ngày 03/6/2020 về việc phê duyệt “Chương trình Chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030”, chương trình Chuyển đổi số quốc gia nhằm mục tiêu kép là vừa phát triển Chính phủ số, kinh tế số, xã hội số, vừa hình thành các doanh nghiệp công nghệ số Việt Nam có năng lực đi ra toàn cầu. Theo đó, tiềm năng dữ liệu để chuyển đổi số quốc gia là rất lớn, cụ thể:

* Cơ sở dữ liệu quốc gia tạo nền tảng phát triển Chính phủ điện tử;
* Từng bước mở dữ liệu của các cơ quan nhà nước để cung cấp dịch vụ công kịp thời, một lần khai báo, trọn vòng đời phục vụ người dân và phát triển kinh tế - xã hội;
* Hình thành nền tảng dữ liệu cho các ngành kinh tế trọng điểm dựa trên dữ liệu của các cơ quan nhà nước;
* Mở dữ liệu cho các tổ chức, doanh nghiệp dựa trên dữ liệu phục vụ người dân, doanh nghiệp.

Trong quyết định này, nhận thấy tầm quan trọng của dữ liệu trong chương trình chuyển đổi số quốc gia, Thủ tướng Chính phủ giao cho các bộ chuyên ngành, đặc biệt là Bộ Thông tin và Truyền thông và Bộ Khoa học và Công nghệ thực hiện xây dựng chiến lược, chính sách quản trị dữ liệu, kế hoạch phát triển dữ liệu, các cơ sở dữ liệu bảo đảm khả năng kết nối, chia sẻ; thiết lập khung danh tính số quốc gia; nghiên cứu, rà soát, sửa đổi, bổ sung hoặc đề xuất sửa đổi, bổ sung các tiêu chuẩn, quy chuẩn kỹ thuật về ứng dụng và phát triển công nghệ số, các mối quan hệ mới trong quá trình chuyển đổi số.

Vì vậy, việc xây dựng hệ thống văn bản quy phạm pháp luật và hành lang pháp lý liên quan đến dữ liệu, trong đó bao gồm công tác chuẩn hoá khung tham chiếu dữ liệu lớn (DLL) nhằm thúc đẩy chương trình chuyển đổi số quốc gia là cần thiết.

Kiến trúc tham chiếu là một mô hình kiến trúc tổng thể, cung cấp nguồn thông tin có căn cứ về một lĩnh vực, chủ đề cụ thể nhằm mục đích hướng dẫn và hạn chế việc tạo ra nhiều kiến trúc và giải pháp. Kiến trúc tham chiếu thường đóng vai trò là nền tảng tham chiếu cho các giải pháp kiến trúc và cũng có thể được sử dụng cho mục đích so sánh và định hướng.

Kiến trúc tham chiếu DLL cần đảm bảo các mục tiêu sau:

* Cung cấp một ngôn ngữ chung cho các bên liên quan;
* Khuyến khích việc tuân thủ các tiêu chuẩn, thông số kỹ thuật và khuôn mẫu chung;
* Cung cấp các phương pháp nhất quán để triển khai hệ thống, giải quyết các vấn đề tương tự;
* Minh họa và nâng cao nhận thức về các thành phần, quy trình và hệ thống DLL khác nhau;
* Cung cấp tài liệu tham khảo kỹ thuật cho các đơn vị, tổ chức và người dùng hiểu, thảo luận, phân loại và so sánh các giải pháp DLL khác nhau;
* Tạo điều kiện thuận lợi cho việc phân tích các tiêu chuẩn về khả năng tương tác, tính di động, khả năng tái sử dụng và mở rộng hệ thống.

Kiến trúc tham chiếu DLL không đại diện cho kiến trúc hệ thống của một hệ thống dữ liệu lớn cụ thể nào, mà là công cụ để mô tả, thảo luận và phát triển kiến trúc hệ thống cụ thể bằng cách sử dụng các khung tham chiếu. Mô hình không bị ràng buộc với bất kỳ sản phẩm, dịch vụ hoặc thực thi tham chiếu của một nhà cung cấp cụ thể nào, cũng như không xác định các giải pháp mang tính quy định, ngăn cản sự đổi mới, sáng tạo.

Hiện nay, Bộ TT&TT và Bộ KH&CN đã ban hành và xây dựng nhiều quy chuẩn, tiêu chuẩn liên quan đến quy định cho khung tham chiếu DLL dự trên bộ tài liệu ISO/IEC 20546 và ISO/IEC 20547, cụ thể:

* TCVN 13238:2020 (ISO/IEC 20546:2019): Công nghệ thông tin – Dữ liệu lớn – Tổng quan và từ vựng. Tiêu chuẩn này hoàn toàn tương đương ISO/IEC 20546:2019;
* TCVN 13239-2:2020 (ISO/IEC TR 20547-2:2018): Công nghệ thông tin – Kiến trúc tham chiếu dữ liệu lớn – Phần 2: Các trường hợp sử dụng và yêu cầu dẫn xuất. Tiêu chuẩn này hoàn toàn tương đương ISO/IEC TR 20547-2:2018;
* Dự thảo TCVN: Công nghệ thông tin – Kiến trúc tham chiếu dữ liệu lớn – Phần 3: Kiến trúc tham chiếu. Dự thảo tiêu chuẩn này hoàn toàn tương đương ISO/IEC 20547-3:2020; và
* TCVN 13239-5:2020 (ISO/IEC TR 20547-2:2018): Công nghệ thông tin – Kiến trúc tham chiếu dữ liệu lớn – Phần 5: Lộ trình tiêu chuẩn. Tiêu chuẩn này hoàn toàn.

Tuy nhiên, bộ tiêu chuẩn quốc gia trên là chưa đầy đủ so với các tiêu chí cần phải đảm bảo cho khung kiến trúc tham chiếu dữ liệu lớn. Hiện nay, bộ tài liệu ISO/IEC ISO/IEC 20547 đưa ra các quy định đầy đủ cho khung tham chiếu dữ liệu lớn bao gồm 5 phần:

* ISO/IEC TR 20547-1, Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 1: Khung và quy trình úng dụng;
* ISO/EC TR 20547-2,  Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 2: Các trường hợp sử dụng và các yêu cầu dẫn xuất;
* ISO/IEC 20547-3, Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 3: Kiến trúc tham chiếu;
* ISO/IEC 20547-4, Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 4: Bảo mật và quyền riêng tư; và
* ISO/IEC TR 20547-5, Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 5: Lộ trình cho các tiêu chuẩn.

Do đó, việc xây dựng các dự thảo tiêu chuẩn:

* **Dự thảo TCVN: Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 1: Khung và quy trình ứng dụng dựa trên ISO / IEC TR 20547-1;** và
* Dự thảo TCVN: Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 4: Bảo mật và quyền riêng tư dựa trên ISO/IEC 20547-4.

là cần thiết nhằm hoàn thiện bộ tiêu chuẩn quốc gia đầy đủ cho kiến trúc tham chiếu dữ liệu phục vụ công tác chuẩn hoá chuyên ngành công nghệ thông tin.

# Tình hình chuẩn hoá Kiến trúc tham chiếu dữ liệu lớn – Khung và quy trình ứng dụng trên thế giới

Trong những năm gần đây, các tổ chức tiêu chuẩn quốc tế cũng như nhiều quốc gia, khu vực đang nỗ lực xúc tiến nghiên cứu và thực hiện tiêu chuẩn hoá về các nội dung liên quan đến DLL, trong đó có tiêu chuẩn về kiến trúc tham chiếu DLL, điển hình là ISO/IEC, ITU, NIST.



## Các tổ chức tiêu chuẩn hoá quốc tế

### Tổ chức ISO/IEC

Bộ tiêu chuẩn ISO/IEC 20547 bổ sung cho tiêu chuẩn nền tảng về thuật ngữ DLL ([ISO/IEC 20546](https://www.iso.org/standard/68305.html)) và cung cấp một kiến trúc tham chiếu DLL toàn diện. Khung kiến trúc tham chiếu DLL giải quyết các yêu cầu, kiến trúc, bảo mật và quyền riêng tư, các trường hợp sử dụng. Ngoài ra, bộ tiêu chuẩn này nhằm cung cấp cho người dùng phương pháp tiếp cận được tiêu chuẩn hoá để phát triển và triển khai các kiến trúc DLL. Bộ tiêu chuẩn thiết lập một khuôn khổ mà các tổ chức có thể dựa vào đó để mô tả một cách hiệu quả và nhất quán kiến ​​trúc của chính họ và các ứng dụng của nó, đảm bảo các công nghệ DLL được sử dụng một cách an toàn và hiệu quả.

Bộ tiêu chuẩn ISO/IEC 20547 bao gồm năm phần:

* ISO/IEC 20546:2019: Công nghệ thông tin - Dữ liệu lớn - Tổng quan và từ vựng: Tiêu chuẩn này quy định tập các thuật ngữ và định nghĩa cần thiết để thúc đẩy giao tiếp và thông hiểu trong lĩnh vực này, đồng thời đưa ra một nền tảng thuật ngữ cho các tiêu chuẩn liên quan DLL. Tiêu chuẩn này đưa ra tổng quan khái niệm về lĩnh vực DLL, mối quan hệ với các lĩnh vực kỹ thuật và các nỗ lực trong hoạt động tiêu chuẩn khác, và các khái niệm được gán cho DLL.
* **ISO / IEC TR 20547-1: Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 1: Khung và quy trình ứng dụng.** Tiêu chuẩn này đưa ra các yêu cầu cho khung kiến trúc DLL và quá trình giúp người dùng tiêu chuẩn có thể áp dụng vào một miền vấn đề cụ thể.
* ISO / IEC TR 20547-2:  Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 2: Các trường hợp sử dụng và yêu cầu dẫn xuất. Đưa ra các ví dụ về các trường hợp sử dụng dữ liệu lớn với các lĩnh vực ứng dụng và các xem xét kỹ thuật được tạo từ các trường hợp sử dụng.
* ISO / IEC 20547-3, Công nghệ thông tin - Kiến trúc tham chiếu DLL - Phần 3: Kiến trúc tham chiếu: Tiêu chuẩn này đưa ra kiến trúc tham chiếu theo quan điểm người dùng và chức năng. Các góc nhìn đó có thể được kiến trúc sư dữ liệu lớn sử dụng để mô tả một hệ thống cụ thể.
* ISO / IEC 20547-4: Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 4: Bảo mật và sự riêng tư. Mô tả khía cạnh bảo mật và quyền riêng tư của DLL. Tiêu chuẩn này đưa ra các yêu cầu bảo mật và và quyền riêng tư cho kiến trúc dữ liệu lớn (BDRA) bao gồm các vai trò, các hoạt động, các thành phần chức năng DLL, đồng thời cung cấp các hướng dẫn về các hoạt động bảo mật và quyền riêng tư đối với DLL.
* ISO / IEC TR 20547-5: Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 5: Lộ trình tiêu chuẩn. Tiêu chuẩn này mô tả các tiêu chuẩn liên quan đến DLL, cả đang tồn tại và đang xây dựng, cùng với các ưu tiên xây dựng tiêu chuẩn DLL trong tương lai dựa trên phân tích sự thiếu hụt tiêu chuẩn.

**Nhận xét:**

Mỗi phần trên được xây dựng dựa trên từ vựng và khái niệm chung được mô tả trong ISO/IEC 20546. Nhìn chung, kiến trúc tham chiếu cung cấp nguồn thông tin có thẩm quyền về một lĩnh vực cụ thể để hướng dẫn và hạn chế việc khởi tạo nhiều kiến trúc và giải pháp. Kiến trúc tham chiếu thường đóng vai trò là nền tảng tham chiếu cho các kiến trúc giải pháp và cũng có thể được sử dụng cho mục đích so sánh và điều chỉnh.

Tài liệu ISO/IEC TR 20547-1:2020: cung cấp cho người dùng cái nhìn tổng quan về khung kiến trúc tham chiếu và quy trình áp dụng trong việc phát triển kiến trúc.

### Liên minh viễn thông quốc tế ITU

Liên minh viễn thông quốc tế, với bộ phận chuẩn hóa viễn thông (ITU-T) là tổ chức chuyên môn của Liên hợp quốc thành lập nhằm thúc đẩy các hoạt động chuẩn hóa trong lĩnh vực viễn thông. Với vai trò này, xét ở khía cạnh nào đó thì hoạt động của ITU-T tương tự các tổ chức tiêu chuẩn quốc tế. Trong bối cảnh phát triển của DLL trong những năm gần đây, ITU-T cũng đã xúc tiến hoạt động nghiên cứu, xây dựng các khuyến nghị trong lĩnh vực này.

Hiện nay, ITU-T đã thực hiện phát triển nhiều tài liệu khuyến nghị sử dụng chung cho lĩnh vực DLL. Các khuyến nghị được tổ chức này đang phát triển hoặc đã được xuất bản bao gồm:

* Y.3600: Dữ liệu lớn – Các yêu cầu và khả năng dựa trên điện toán đám mây
* **Y.3601: Dữ liệu lớn – Khung và các yêu cầu cho trao đổi dữ liệu**
* Y.3602: Dữ liệu lớn – Các yêu cầu chức năng đối với dữ liệu gốc
* Y.3603: Dữ liệu lớn – Các yêu cầu và mô hình khái niệm của siêu dữ liệu cho danh mục dữ liệu
* Y.3604: Dữ liệu lớn – Tổng quan và các yêu cầu đối với dữ liệu gốc
* Y.3605: Dữ liệu lớn – Kiến trúc tham chiếu
* Y.3606: Dữ liệu lớn – Cơ chế kiểm soát sâu gói tin dữ liệu lớn trong mạng
* Y.3650: Dữ liệu lớn – Khung của mạng định hướng dữ liệu lớn.
* Y.3651: Mạng định hướng dữ liệu lớn – Quản lý và lập kế hoạch lưu lượng mạng di động
* Y.3652: Mạng định hướng dữ liệu lớn – Các yêu cầu
* Y.3653: Mạng định hướng dữ liệu lớn – Kiến trúc chức năng
* Y.3654: Mạng định hướng dữ liệu lớn – Cơ chế học máy

Nhận xét:

Theo ITU-T, kiến trúc tham chiếu là mô hình chung của một loại hình nghiệp vụ, bộ giải pháp, hệ thống và trình điều khiển. Hầu hết các tổ chức bắt đầu từ các mô hình đã triển khai hiệu quả trước đó, mô tả một tổ chức, tổng thể, dữ liệu, ứng dụng và công nghệ điển hình phải hoạt động cùng nhau. Kiến ​​trúc tham chiếu cung cấp một mô hình để phát triển các tiêu chuẩn hệ thống, thông tin dịch vụ và công nghệ phù hợp với toàn ngành và có thể được sử dụng lại bởi nhiều tổ chức điều hành và quản lý, nhà cung cấp và nhà tích hợp trên toàn miền.

Nhằm cung cấp cách tiếp cận nhất quán để mô tả kiến trúc và triển khai hệ thống DLL, ITU-T đã ban hành khuyến nghị về khung và các yêu cầu cho trao đổi dữ liệu (ITU-T Y.3605) trong hệ thống DLL. Khuyến nghị này quy định Khung và Các yêu cầu cho việc trao đổi dữ liệu trong hệ thống DLL. Đưa ra các quy định về khái niệm, định dạng, hoạt động và các yêu cầu chức năng chung dựa trên hệ thống DLL và dung lượng được định nghĩa trong ITU-T Y.3600. Các yêu cầu về chức năng được cung cấp trong khuyến nghị này được dẫn xuất từ các trường hợp người dùng.

Kiến trúc tham chiếu dữ liệu lớn (ITU-T Y.3605), tài liệu khuyến nghị này xác định kiến trúc tham chiếu DLL có thể đóng vai trò là điểm tham chiếu cơ bản cho việc chuẩn hoá DLL và cung cấp một cái nhìn tổng quan về các khái niệm và nguyên tắc cơ bản của DLL. Khuyến nghị cung cấp mô tả về các khái niệm kiến trúc tham chiếu; góc nhìn người dùng, góc nhìn chức năng và các khía cạnh xuyên suốt; đồng thời đề cập đến khung phân lớp, các thành phần chức năng trong khung và các mô tả chức năng chi tiết cho DLL.

ITU-T Y.3601 đưa ra khung và các yêu cầu cho việc trao đổi dữ liệu trong hệ sinh thái dữ liệu lớn.

### Viện Tiêu chuẩn và Công nghệ Quốc gia Mỹ (NIST)

Để thúc đẩy tiến bộ trong dữ liệu lớn, nhóm công tác dữ liệu lớn của NIST (NBD-PWG) đã được ra mắt với sự tham gia rộng rãi của các ngành, học viện và chính phủ trên toàn quốc. Nhóm được hình thành với mục tiêu phát triển sự đồng thuận về các khái niệm cơ bản quan trọng liên quan đến DLL tạo ra một khung mẫu (framework) không phụ thuộc vào nhà cung cấp, công nghệ và cơ sở hạ tầng cho phép các bên liên quan đến DLL xác định và sử dụng các công cụ phân tích tốt nhất cho các yêu cầu xử lý và ảo hoá trên hệ thống cluster và nền tảng điện toán phù hợp nhất. Và các kết quả này được bao gồm trong bộ tài liệu về “Khung khả năng tương tác dữ liệu lớn của NIST”. Bộ tài liệu bao gồm 7 tập, mỗi tập đề cập đến một chủ đề chính, cụ thể bao gồm:

* NIST.SP.1500-1: Định nghĩa;
* NIST.SP.1500-2: Phân loại;
* NIST.SP.1500-3: Các trường hợp sử dụng và yêu cầu chung;
* NIST.SP.1500-4: Bảo mật và quyền riêng tư;
* NIST.SP.1500-5: Khảo sát kiến trúc sách trắng;
* NIST.SP.1500-6: Kiến trúc tham chiếu;
* NIST.SP.1500-7: Lộ trình tiêu chuẩn.

**Nhận xét:**

NIST không đưa ra tổng quan về khung kiến trúc tham chiếu DLL và quy trình áp dụng trong việc phát triển kiến trúc. Tuy nhiên, tài liệu NIST.SP không đưa ra các yêu cầu cho khung kiến trúc dữ liệu lớn và quá trình giúp người dùng tiêu chuẩn có thể áp dụng vào một miền vấn đề cụ thể.

# Tình hình chuẩn hoá Kiến trúc tham chiếu dữ liệu lớn – Khung và quy trình ứng dụng tại Việt Nam

## TCVN 13238:2020 (ISO/IEC 20546:2019)

TCVN 13238:2020 (ISO/IEC 20546:2019): Tiêu chuẩn Công nghệ thông tin – Dữ liệu lớn – Tổng quan và từ vựng. Tiêu chuẩn này quy định tập các thuật ngữ và định nghĩa cần thiết để thúc đẩy giao tiếp và thông hiểu trong lĩnh vực này, đồng thời đưa ra một nền tảng thuật ngữ cho các tiêu chuẩn liên quan dữ liệu lớn. Tiêu chuẩn này đưa ra tổng quan khái niệm về lĩnh vực dữ liệu lớn, mối quan hệ với các lĩnh vực kỹ thuật và các nỗ lực trong hoạt động tiêu chuẩn khác, và các khái niệm được gán cho dữ liệu lớn.

## TCVN 13239-2:2020 (ISO/IEC TR 20547-2:2018)

TCVN 13239-2:2020 (ISO/IEC TR 20547-2:2018): Công nghệ thông tin – Kiến trúc tham chiếu dữ liệu lớn – Phần 2: Các trường hợp sử dụng và yêu cầu dẫn xuất. Tiêu chuẩn này đưa ra các ví dụ về các trường hợp sử dụng dữ liệu lớn với các lĩnh vực ứng dụng và các xem xét kỹ thuật được tạo từ các trường hợp sử dụng đóng góp vào.

## Dự thảo TCVN ISO/IEC 20547-3:2022: Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 3: Kiến trúc tham chiếu

Dự thảo TCVN ISO/IEC 20547-3:2022: Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 3: Kiến trúc tham chiếu hoàn toàn tương đương ISO/IEC 20547-3:2020. Dự thảo tiêu chuẩn này quy định kiến trúc tham chiếu dữ liệu lớn. Kiến trúc tham chiếu bao gồm các khái niệm và góc nhìn kiến trúc.

Kiến trúc tham chiếu trong tài liệu này xác định hai góc nhìn kiến trúc:

* Góc nhìn người dùng: định nghĩa các vai trò/vai trò phụ, mối quan hệ và các hoạt động của chúng trong hệ sinh thái dữ liệu lớn;
* Góc nhìn chức năng: định nghĩa các lớp kiến trúc và các lớp của các thành phần chức năng bên trong, thực thi các hoạt động của các vai trò/vai trò phụ trong góc nhìn người dùng.

Mục đích của kiến trúc tham chiếu dữ liệu lớn:

* Cung cấp một ngôn ngữ chung cho các bên liên quan;
* Khuyến khích việc tuân thủ các tiêu chuẩn, thông số kỹ thuật và các khuôn mẫu chung;
* Cung cấp một phương thức công nghệ nhất quán để giải quyết các vấn đề tương tự;
* Tạo điều kiện cho việc tìm hiểu sự phức tạp trong vận hành dữ liệu lớn;
* Mô tả và hiểu rõ các thành phần, quy trình và các hệ thống dữ liệu lớn khác nhau, trong bối cảnh một mô hình khái niệm dữ liệu lớn tổng thể;
* Cung cấp một tài liệu kỹ thuật tham khảo cho các chính phủ, các cơ quan, đơn vị và những người dùng khác để nắm bắt, thảo luận, phân loại và so sánh các giải pháp dữ liệu lớn;
* Tạo điều kiện cho việc phân tích khả năng tương thích, tính di động, khả năng tái sử dụng và mở rộng của các tiêu chuẩn đề xuất.

## TCVN 13239-5:2020 (ISO/IEC TR 20547-5:2018)

TCVN 13239-5:2020 (ISO/IEC TR 20547-5:2018): Tiêu chuẩn Công nghệ thông tin – Kiến trúc tham chiếu dữ liệu lớn – Phần 5: Lộ trình tiêu chuẩn. Tiêu chuẩn này mô tả các tiêu chuẩn liên quan đến dữ liệu lớn, cả đang tồn tại và đang xây dựng, cùng với các ưu tiên xây dựng tiêu chuẩn dữ liệu lớn trong tương lai dựa trên phân tích sự thiếu hụt tiêu chuẩn.

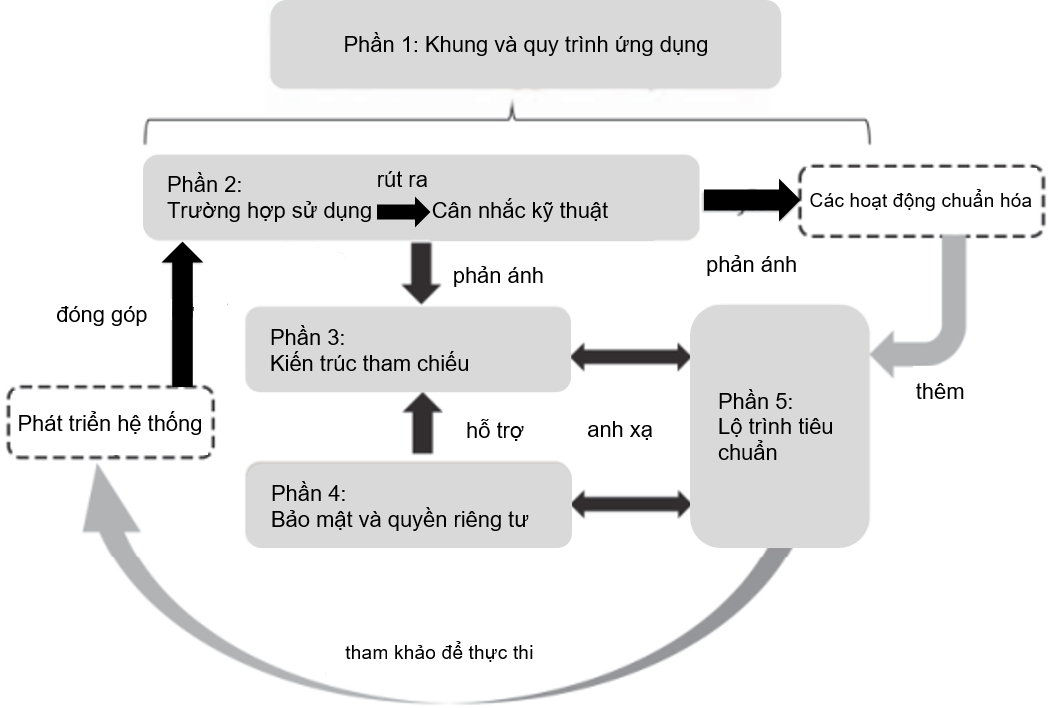
Bảng 1 – Bảng tổng hợp tình hình tiêu chuẩn hoá Công nghệ thông tin – Kiến trúc tham chiếu dữ liệu lớn

|  | **ISO/IEC** | **ITU** | **NIST** | **QCVN, TCVN** |
| --- | --- | --- | --- | --- |
| Từ vựng, thuật ngữ và định nghĩa | ISO/IEC 20546:2019  *Công nghệ thông tin – Dữ liệu lớn: Tổng quan và từ vựng* |  | NIST.SP.1500-1  *Khung khả năng tương tác dữ liệu lớn của NIST- Tập 1: Các định nghĩa* | TCVN 13238:2020 (ISO/IEC 20546:2019)  *Tiêu chuẩn quốc gia - Công nghệ thông tin – Dữ liệu lớn – Tổng quan và từ vựng* |
| Tiêu chuẩn này quy định tập các thuật ngữ và định nghĩa cần thiết để thúc đẩy giao tiếp và thông hiểu trong lĩnh vực này, đồng thời đưa ra một nền tảng thuật ngữ cho các tiêu chuẩn liên quan dữ liệu lớn.  Tiêu chuẩn này đưa ra tổng quan khái niệm về lĩnh vực dữ liệu lớn, mối quan hệ với các lĩnh vực kỹ thuật và các nỗ lực trong hoạt động tiêu chuẩn khác, và các khái niệm được gán cho dữ liệu lớn. |  |  | Tiêu chuẩn này quy định tập các thuật ngữ và định nghĩa cần thiết để thúc đẩy giao tiếp và thông hiểu trong lĩnh vực này, đồng thời đưa ra một nền tảng thuật ngữ cho các tiêu chuẩn liên quan dữ liệu lớn.  Tiêu chuẩn này đưa ra tổng quan khái niệm về lĩnh vực dữ liệu lớn, mối quan hệ với các lĩnh vực kỹ thuật và các nỗ lực trong hoạt động tiêu chuẩn khác, và các khái niệm được gán cho dữ liệu lớn. |
| **Khung và quy trình áp dụng** | **ISO/IEC TR 20547-1:2020**  *Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 1: Khung và quy trình ứng dụng* | Y.3601  *Dữ liệu lớn – Khung và các yêu cầu cho trao đổi dữ liệu* |  | Chưa có |
| Tiêu chuẩn này đưa ra các yêu cầu cho khung kiến trúc dữ liệu lớn và quá trình giúp người dùng tiêu chuẩn có thể áp dụng vào một miền vấn đề cụ thể. | Khuyến nghị này đưa ra khung và các yêu cầu cho việc trao đổi dữ liệu trong hệ sinh thái dữ liệu lớn. |  |  |
| Các trường hợp sử dụng | ISO/IEC TR 20547-2:2018  *Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 2: Các trường hợp sử dụng và yêu cầu dẫn xuất* |  | NIST.SP.1500-3  *Khung khả năng tương tác dữ liệu lớn của NIST- Tập 3: Các trường hợp sử dụng và yêu cầu chung* | TCVN 13239-2:2020 (ISO/IEC TR 20547-2:2018)  *Tiêu chuẩn quốc gia - Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 2: Các trường hợp sử dụng và yêu cầu dẫn xuất* |
| Tiêu chuẩn này đưa ra các ví dụ về các trường hợp sử dụng dữ liệu lớn với các lĩnh vực ứng dụng và các xem xét kỹ thuật được tạo từ các trường hợp sử dụng đóng góp vào. |  | Tài liệu này đưa ra một danh sách các trường hợp sử dụng và yêu cầu cụ thể cho việc áp dụng triển khai dữ liệu lớn. | Tiêu chuẩn này đưa ra các ví dụ về các trường hợp sử dụng dữ liệu lớn với các lĩnh vực ứng dụng và các xem xét kỹ thuật được tạo từ các trường hợp sử dụng đóng góp vào. |
| Kiến trúc tham chiếu | ISO/IEC 20547-3  *Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 3: Kiến trúc tham chiếu* | ITU-T Y.3605  *Dữ liệu lớn – Kiến trúc tham chiếu* | NIST.SP.1500-6  *Khung khả năng tương tác dữ liệu lớn của NIST-Tập 6: Kiến trúc tham chiếu* | Dự thảo TCVN  *Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 3: Kiến trúc tham chiếu* |
| Tiêu chuẩn này quy định kiến trúc tham chiếu dữ liệu lớn. Kiến trúc tham chiếu bao gồm các khái niệm và góc nhìn kiến trúc.  Kiến trúc tham chiếu trong tài liệu này xác định hai góc nhìn kiến trúc:  - Góc nhìn người dùng: định nghĩa các vai trò/vai trò phụ, mối quan hệ và các hoạt động của chúng trong hệ sinh thái dữ liệu lớn;  - Góc nhìn chức năng: định nghĩa các lớp kiến trúc và các lớp của các thành phần chức năng bên trong, thực thi các hoạt động của các vai trò/vai trò phụ trong góc nhìn người dùng.  Mục đích của kiến trúc tham chiếu dữ liệu lớn:  - Cung cấp một ngôn ngữ chung cho các bên liên quan;  - Khuyến khích việc tuân thủ các tiêu chuẩn, thông số kỹ thuật và các khuôn mẫu chung;  - Cung cấp một phương thức công nghệ nhất quán để giải quyết các vấn đề tương tự;  - Tạo điều kiện cho việc tìm hiểu sự phức tạp trong vận hành dữ liệu lớn;  - Mô tả và hiểu rõ các thành phần, quy trình và các hệ thống dữ liệu lớn khác nhau, trong bối cảnh một mô hình khái niệm dữ liệu lớn tổng thể;  - Cung cấp một tài liệu kỹ thuật tham khảo cho các chính phủ, các cơ quan, đơn vị và những người dùng khác để nắm bắt, thảo luận, phân loại và so sánh các giải pháp dữ liệu lớn;  - Tạo điều kiện cho việc phân tích khả năng tương thích, tính di động, khả năng tái sử dụng và mở rộng của các tiêu chuẩn đề xuất. | Khuyến nghị này mô tả về các khái niệm kiến trúc tham chiếu; các góc nhìn và các khía cạnh xuyên suốt.  Tập trung chi tiết vào 2 góc nhìn là góc nhìn người dùng và góc nhìn chức năng.  \* Góc nhìn người dùng: hệ sinh thái dữ liệu lớn với các khái niệm:  - Vai trò/vai trò phụ: đề cập đến 3 vai trò dữ liệu lớn (nhà cung cấp dữ liệu lớn, nhà cung cấp dịch vụ dữ liệu lớn, khách hàng dịch vụ dữ liệu lớn); và 02 vai trò phụ của vai trò nhà cung cấp dữ liệu.  - Hoạt động: các hoạt động của các vai trò dữ liệu lớn.  - Khía cạnh xuyên suốt: đề cập đến 3 khía cạnh bao gồm: dữ liệu dư thừa, chất lượng dữ liệu, hiệu năng.  \* Góc nhìn chức năng: đề cập đến các khái niệm:  - Các lớp chức năng: bốn lớp, gồm có lớp truy cập, lớp ứng dụng, lớp xử lý, lớp nguồn dữ liệu.  - Các chức năng nhiều lớp, gồm có: hệ thống hỗ trợ phân tích, quản lý vòng đời dữ liệu, bảo mật. | Tài liệu này mô tả dữ liệu lớn từ góc độ kiến trúc, trình bày mô hình khái niệm Kiến trúc tham chiếu dữ liệu lớn của NIST (NBDRA) và đề cập đến các thành phần và cấu trúc của NBDRA.  Kiến trúc tham chiếu dữ liệu lớn đại diện cho một hệ thống dữ liệu lớn bao gồm năm thành phần chính đại diện cho các vai trò khác nhau tồn tại trong mọi hệ thống dữ liệu lớn và hai khía cạnh bao phủ các thành phần của hệ thống dữ liệu lớn. Cụ thể:  - 5 vai trò dữ liệu lớn: : Điều phối hệ thống; Nhà cung cấp dữ liệu; Nhà cung cấp ứng dụng dữ liệu lớn; Nhà cung cấp framework dữ liệu lớn; Người tiêu dùng dữ liệu.  - 2 khía cạnh bao phủ xuyên suốt các vai trò: Quản lý, bảo mật và quyền riêng tư. | Tiêu chuẩn này quy định kiến trúc tham chiếu dữ liệu lớn. Kiến trúc tham chiếu bao gồm các khái niệm và góc nhìn kiến trúc.  Kiến trúc tham chiếu trong tài liệu này xác định hai góc nhìn kiến trúc:  - Góc nhìn người dùng: định nghĩa các vai trò/vai trò phụ, mối quan hệ và các hoạt động của chúng trong hệ sinh thái dữ liệu lớn;  - Góc nhìn chức năng: định nghĩa các lớp kiến trúc và các lớp của các thành phần chức năng bên trong, thực thi các hoạt động của các vai trò/vai trò phụ trong góc nhìn người dùng.  Mục đích của kiến trúc tham chiếu dữ liệu lớn:  - Cung cấp một ngôn ngữ chung cho các bên liên quan;  - Khuyến khích việc tuân thủ các tiêu chuẩn, thông số kỹ thuật và các khuôn mẫu chung;  - Cung cấp một phương thức công nghệ nhất quán để giải quyết các vấn đề tương tự;  - Tạo điều kiện cho việc tìm hiểu sự phức tạp trong vận hành dữ liệu lớn;  - Mô tả và hiểu rõ các thành phần, quy trình và các hệ thống dữ liệu lớn khác nhau, trong bối cảnh một mô hình khái niệm dữ liệu lớn tổng thể;  - Cung cấp một tài liệu kỹ thuật tham khảo cho các chính phủ, các cơ quan, đơn vị và những người dùng khác để nắm bắt, thảo luận, phân loại và so sánh các giải pháp dữ liệu lớn;  - Tạo điều kiện cho việc phân tích khả năng tương thích, tính di động, khả năng tái sử dụng và mở rộng của các tiêu chuẩn đề xuất. |
| **Bảo mật và quyền riêng tư** | **ISO / IEC 20547-4: 2020**  *Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 4: Bảo mật và sự riêng tư* |  | NIST.SP.1500-4  *Khung khả năng tương tác dữ liệu lớn của NIST- Tập 4: Bảo mật và quyền riêng tư* | Chưa có |
| Tiêu chuẩn này đưa ra các yêu cầu bảo mật và và quyền riêng tư cho kiến trúc dữ liệu lớn (BDRA) bao gồm các vai trò, các hoạt động, các thành phần chức năng dữ liệu lớn, đồng thời cung cấp các hướng dẫn về các hoạt động bảo mật và quyền riêng tư đối với dữ liệu lớn. |  | Tài liệu này đưa ra các yêu cầu bảo mật và quyền rêng tư cho dữ liệu lớn.  Đưa ra các ví dụ về trường hợp sử dụng bảo mật và quyền riêng tư và ánh xạ các trường hợp dử dụng đến NBDRA. |  |
| Lộ trình tiêu chuẩn | ISO / IEC TR 20547-5  *Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 5: Lộ trình tiêu chuẩn)* |  | NIST.SP.1500-7  *Khung khả năng tương tác dữ liệu lớn của NIST-Tập 7: Lộ trình tiêu chuẩn* | TCVN 13239-5:2020 (ISO/IEC TR 20547-5:2018)  *Tiêu chuẩn quốc gia - Công nghệ thông tin – Kiến trúc tham chiếu dữ liệu lớn – Phần 5: Lộ trình tiêu chuẩn* |
| Tiêu chuẩn này mô tả các tiêu chuẩn liên quan đến dữ liệu lớn, cả đang tồn tại và đang xây dựng, cùng với các ưu tiên xây dựng tiêu chuẩn dữ liệu lớn trong tương lai dựa trên phân tích sự thiếu hụt tiêu chuẩn. |  | Tài liệu này đưa ra khái niệm hệ sinh thái dữ liệu lớn. Danh sách các tổ chức tiêu chuẩn đang xây dựng tiêu chuẩn về dữ liệu lớn và ưu tiên khi xây dựng hệ thống tiêu chuẩn dữ liệu lớn. | Tiêu chuẩn này mô tả các tiêu chuẩn liên quan đến dữ liệu lớn, cả đang tồn tại và đang xây dựng, cùng với các ưu tiên xây dựng tiêu chuẩn dữ liệu lớn trong tương lai dựa trên phân tích sự thiếu hụt tiêu chuẩn. |

## Nhận xét chung

Trong môi trường hệ thống dữ liệu lớn, tiêu chuẩn này cung cấp cấu trúc BDRA thông qua phạm vi áp dụng của từng phần trong bộ tiêu chuẩn ISO/IEC 20547, mối quan hệ lô-gíc giữa các phần và quy trình ứng dụng BDRA. Phạm vi áp dụng của mỗi phần trong bộ tiêu chuẩn ISO/IEC 20547 cụ thể như sau:

* ISO/IEC 20547-1: Khung và quy trình ứng dụng mô tả khung BDRA và quy trình giúp người dùng tiêu chuẩn có thể áp dụng vào một miền vấn đề cụ thể.
* ISO/IEC 20547-2 (TCVN 13239-2:2020): Các trường hợp sử dụng và các yêu cầu dẫn xuất cung cấp các ví dụ về các trường hợp sử dụng dữ liệu lớn với các miền ứng dụng và các xem xét kỹ thuật được rút ra từ các trường hợp sử dụng.
* ISO/IEC 20547-3: Kiến trúc tham chiếu nêu rõ kiến trúc tham chiếu dữ liệu lớn. Kiến trúc tham chiếu bao gồm các khái niệm và các góc nhìn kiến trúc (góc nhìn người dùng và góc nhìn chức năng).
* ISO/IEC 20547-4: Tính bảo mật và quyền riêng tư nêu rõ khía cạnh bảo mật và riêng tư áp dụng cho BDRA bao gồm các vai trò, hoạt động và các thành phần chức năng dữ liệu lớn; đồng thời cung cấp hướng dẫn về các hoạt động bảo mật và quyền riêng tư đối với dữ liệu lớn.
* ISO/IEC 20547-5 (TCVN 13239-5:2020): Lộ trình các tiêu chuẩn mô tả các tiêu chuẩn liên quan đến dữ liệu lớn, cả đang tồn tại và đang xây dựng, cùng với các ưu tiên xây dựng tiêu chuẩn dữ liệu lớn trong tương lai dựa trên phân tích sự thiếu hụt tiêu chuẩn.



Hình 1: Các mối quan hệ giữa các phần của bộ tiêu chuẩn ISO/IEC 20547

Hình 1 thể hiện các mối quan hệ và chu kỳ lặp lại của từng phần trong bộ ISO/IEC 20547. Dựa trên các đóng góp từ các doanh nghiệp, các tổ chức và chuyên gia trong lĩnh vự nghiên cứu, ISO/IEC 20547-2 thu thập các trường hợp sử dụng và rút ra các cân nhắc kỹ thuật. ISO/IEC 20547-3:2022 xác định kiến trúc tham chiếu cho dữ liệu lớn bằng cách phản ánh các cân nhắc kỹ thuật này. Cụ thể, ISO/IEC 20547-4 quy định khía cạnh bảo mật và riêng tư để hỗ trợ dữ liệu lớn. ISO/IEC 20547-5 cung cấp các danh sách các tiêu chuẩn có thể áp dụng theo quan điểm kiến trúc tham chiếu. Bộ ISO/IEC 20547 thể hiện góc nhìn theo thời gian về các hệ thống và kiến trúc dữ liệu lớn. Khi các thực thi dữ liệu lớn được tạo ra và phát triển dựa trên ISO/IEC TR 20547-1, ISO/IEC 20547-2, ISO/IEC 20547-3 và ISO/IEC 20547-4, chúng sẽ tham chiếu và sử dụng các tiêu chuẩn được ghi trong ISO/IEC 20547-5. Các hệ thống mới có này thể được ghi lại trong ISO/IEC 20547-2 như là các trường hợp sử dụng mới và sẽ từ đó dẫn tới các cân nhắc kỹ thuật mới. Các cân nhắc kỹ thuật được giởi thiệu bởi các trường hợp sử dụng này có thể dẫn tới các hoạt động tiêu chuẩn hóa mới, từ đó dẫn đến các tiêu chuẩn mới được ghi lại trong ISO/IEC 20547-5.

Khi áp dụng khung tham chiếu trong một trường hợp sử dụng cụ thể, người dùng cần phải hiểu rõ môi trường tổng thể triển khai hệ thống dữ liệu lớn, bên liên quan đến hệ thống này và các mối quan tâm của bên liên quan. Trong ISO/IEC 20547-1, điều 8.2 đến 8.4 mô tả từng thành phần chính của khung tham chiếu, bao gồm: các bên liên quan, các mối quan tâm và các góc nhìn.

# Xây dựng bộ tiêu chuẩn quốc gia về Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn – Phần 1: Khung và Quy trình ứng dụng (dựa trên ISO/IEC TR 20547-1:2020)



## Lựa chọn tài liệu tham khảo

Hiện nay, chiến lược chuyển đổi số đã được ban hành và xu thế chuyển đổi số đang diễn ra rất nhanh chóng theo xu hướng chung trên toàn thế giới. Do vậy, việc sớm ban hành các tiêu chuẩn trong lĩnh vực DLL và đặc biệt là kiến trúc tham chiếu DLL là vô cùng cần thiết, nó sẽ cung cấp cho các tổ chức, đơn vị hoạt động trong lĩnh vực công nghệ, các kiến trúc sư hệ thống một khuôn mẫu chung, thống nhất và đồng bộ trong việc xây dựng giải pháp cho các hệ thống DLL tại Việt Nam.

Để thực hiện việc xây dựng và ban hành các tiêu chuẩn về kiến trúc tham chiếu DLL tại Việt Nam thì việc lựa chọn tài liệu tiêu chuẩn, làm cơ sở tham chiếu là hết sức quan trọng. Tài liệu được chọn không những cần đáp ứng được các tiêu chuẩn chung về kiến trúc tham chiếu DLL, mà còn phải phù hợp với các đặc thù của công nghệ, kỹ thuật và lộ trình chuẩn hóa tại Việt Nam.

Căn cứ vào tình hình xây dựng tiêu chuẩn kiến trúc tham chiếu DLL tại Việt Nam và các tài liệu tiêu chuẩn kiến trúc tham chiếu DLL của các tổ chức uy tín trên thế giới, nhóm nghiên cứu đã tổng hợp, phân tích và đánh giá như sau:

* Tại Việt Nam, Bộ Khoa học và Công nghệ đã ban hành các tiêu chuẩn về kiến trúc tham chiếu DLL hoàn toàn tương đương các tiêu chuẩn ISO/IEC TR 20547-2:2018 và ISO/IEC TR 20547-5:2018. Bộ Thông tin và Truyền thông đang xây dựng dự thảo tiêu chuẩn về Khung kiến trúc tham chiếu DLL, phần Kiến trúc tham chiếu hoàn toàn tương đương ISO/IEC 20547-4:2020. Tuy nhiên, bộ tiêu chuẩn và dự thảo tiêu chuẩn này chưa phải đầy đủ về khung kiến trúc tham chiếu DLL, chưa có các quy định cho khung và quy trình ứng dụng trong khung kiến trúc tham chiếu DLL.
* Tiêu chuẩn kiến trúc tham chiếu DLL của ISO: được xây dựng bởi tiểu ban SC42 của Ủy ban kỹ thuật chung ISO/IEC JTC1 chuyên về việc xây dựng tiêu chuẩn trong lĩnh vực trí tuệ nhân tạo và DLL của Tổ chức tiêu chuẩn hóa quốc tế (ISO), ban hành năm 2020. Tài liệu này được xây dựng trên cơ sở kế thừa, vận dụng một loạt các tiêu chuẩn cùng lĩnh vực đã được ban hành trước đó của ISO như: ISO 8000-2, Data quality – Part 2: Vocabulary; ISO/TS 8000-60, Data quality – Part 60: Data quality management: Overview; ISO 8000-61, Data quality – Part 61: Data quality management: process reference model; ISO/IEC/IEEE 42010, Systems and software engineering — Architecture description; ISO/IEC 20546, Information technology — Big data — Overview and vocabulary,… Do vậy, tài liệu đã đảm bảo được tính đồng bộ và nhất quán trong lộ trình xây dựng và phát triển tiêu chuẩn. Về nội dung, tài liệu ISO/IEC TR 20547-1 mô tả rất đầy đủ và toàn diện về các yêu cầu cho khung kiến trúc DLL và quá trình giúp người dùng tiêu chuẩn có thể áp dụng vào một miền vấn đề cụ thể. Do vậy, nhóm nghiên cứu đánh giá đây là tài liệu phù hợp nhất để làm cơ sở tham chiếu, xây dựng tiêu chuẩn kiến trúc tham chiếu DLL tại Việt Nam.
* Tiêu chuẩn kiến trúc tham chiếu DLL của ITU: được xây dựng bởi bộ phận Tiêu chuẩn hóa viễn thông (ITU-T), ban hành năm 2020. Cũng giống như tài liệu ISO/IEC 20547-4:2020, tài liệu ITU-T Y.3601 cũng được xây dựng trên cơ sở kế thừa, vận dụng các tiêu chuẩn đã được ban hành trước đó của ITU như: Recommendation ITU-T Y.3600 (2015), Big data – cloud computing based requirements and capabilities; Recommendation ITU-T Y.3601 (2018), Big data – framework and requirements for data exchange; Recommendation ITU-T Y.3603 (2019), Big data – requirements and conceptual model of metadata for data catalogue; Recommendation ITU-T Y.3604 (2020), Big data – overview and requirements for data preservation. Về nội dung, tài liệu ITU-T Y.3601 cũng đưa ra khung và các yêu cầu cho việc trao đổi dữ liệu trong hệ sinh thái dữ liệu lớn như tài liệu ISO/IEC TR 20547-1:2020. Tuy nhiên, mức độ chi tiết của các nội dung không được cụ thể, mạch lạc và đầy đủ như tài liệu của ISO. Do vậy, nhóm nghiên cứu đánh giá tài liệu này cũng ở mức tương đối, mức độ phù hợp kém hơn so với tài liệu tiêu chuẩn ISO/IEC TR 20547-1:2020 của ISO.

Do đó, nhóm thực hiện lựa chọn tài liệu tham khảo xây dựng dự thảo TCVN ISO/IEC TR 20547-1:2022 “Công nghệ thông tin - Kiến trúc tham chiếu dữ liệu lớn - Phần 1: Khung và quy trình ứng dụng” là ISO/IEC TR 20547-1:2020.

## Cách thức xây dựng

Cách thức xây dựng dự thảo tiêu chuẩn tuân thủ các quy định tại Thông tư số 13/2019/TT-BTTTT ngày 22 tháng 11 năm 2019 của Bộ Thông tin và Truyền thông quy định hoạt động xây dựng quy chuẩn kỹ thuật quốc gia, tiêu chuẩn quốc gia, tiêu chuẩn cơ sở thuộc lĩnh vực quản lý của Bộ Thông tin và Truyền thông, bao gồm các nội dung:

* Tổ chức nghiên cứu, xây dựng dự thảo tiêu chuẩn;
* Tổ chức các hội nghị, hội thảo, lấy ý kiến của chuyên gia và các tổ chức, cá nhân có liên quan;
* Lấy ý kiến góp ý của các cơ quan, tổ chức, cá nhân có liên quan và lấy ý kiến trên cổng thông tin điện tử của Chính phủ, của Bộ Thông tin và Truyền thông;
* Tổ chức thẩm tra và thực hiện các thủ tục ban hành quy chuẩn.

## Về hình thức trình bày

Dự thảo TCVN được trình bày theo đúng hướng dẫn về việc trình bày và thể hiện nội dung quy chuẩn quy định tại Phụ lục số V ban hành kèm theo Thông tư số 13/2019/TT-BTTTT ngày 22 tháng 11 năm 2019 của Bộ Thông tin và Truyền thông.

## Tên dự thảo tiêu chuẩn

Theo đề cương khoa học công nghệ đã đăng ký, nhiệm vụ nhằm mục đích phục vụ công tác chuẩn hóa về kiến trúc tham chiếu dữ liệu lớn, bên cạnh đó, tên dự thảo tiêu chuẩn quốc gia phải ngắn gọn mà đủ ý, nhóm thực hiện nhiệm vụ đề xuất tên dự thảo TCVN là “Công nghệ thông tin – Kiến trúc tham chiếu dữ liệu lớn – Phần 1: Khung và quy trình ứng dụng”.

## Nội dung dự thảo tiêu chuẩn

Bố cục dự thảo TCVN được xây dựng trên cơ sở tham khảo bố cục các TCVN về công nghệ thông tin được Bộ Thông tin và Truyền thông ban hành.

Nội dung của dự thảo tiêu chuẩn được biên soạn hoàn toàn tương đương với tài liệu ISO/IEC TR 20547-1:2020 và được bố cục lại, bao gồm các mục:

Nội dung đối chiếu dự thảo quy chuẩn kỹ thuật với tài liệu tham khảo được thể hiện trong bảng dưới đây:

1. Phạm vi áp dụng

2. Tài liệu viện dẫn

3. Thuật ngữ và định nghĩa

4. Chữ viết tắt

5. Tổng quan về tiêu chuẩn

6. Chuẩn hóa dữ liệu lớn: động lực và mục tiêu

7. Các cơ sở khái niệm

7.1. Yêu cầu chung

7.2. Các khái niệm kiến trúc tham chiếu

7.3. Cấu trúc kiến trúc tham chiếu

8. Các yếu tố của kiến trúc tham chiếu dữ liệu lớn

8.1. Tổng quan

8.2. Các bên liên quan

8.3. Các mối quan tâm

8.4. Các quan điểm

8.4.1. Quan điểm người dùng

8.4.2. Quan điểm chức năng

9. Quá trình áp dụng kiến trúc tham chiếu dữ liệu lớn

9.1. Tổng quan

9.2. Xác định các bên liên quan và các mối quan tâm

9.3. Ánh xạ các bên liên quan và các mối quan tâm tới các vai trò và các vai trò phụ

9.4. Xây dựng các mô tả hoạt động chi tiếp và ánh xạ tới các mối quan tâm

9.5. Xác định các thành phần chức năng để thực hiện các hoạt động

9.6. Các hoạt động Crosswalk/các thành phần chức năng trở lại các mối quan tâm

Thư mục tài liệu tham khảo

Bảng 2 - Bảng đối chiếu nội dung TCVN và tài liệu tham khảo

| **Dự thảo TCVN** | **Tài liệu tham khảo** | **Hình thức xây dựng** |
| --- | --- | --- |
| 1. Phạm vi áp dụng | ISO/IEC TR 20547-1:2020, điều 1 | Chấp thuận nguyên vẹn |
| 2. Tài liệu viện dẫn | ISO/IEC TR 20547-1:2020, điều 2 | Chấp thuận nguyên vẹn |
| 3. Thuật ngữ và định nghĩa | ISO/IEC TR 20547-1:2020, điều 4 | Chấp thuận nguyên vẹn |
| 4. Chữ viết tắt | ISO/IEC TR 20547-1:2020, điều 5 | Chấp thuận nguyên vẹn |
| 5. Tổng quan về tiêu chuẩn | ISO/IEC TR 20547-1:2020, điều 5 | Chấp thuận nguyên vẹn |
| 6. Chuẩn hóa dữ liệu lớn: động lực và mục tiêu | ISO/IEC TR 20547-1:2020, điều 6 | Chấp thuận nguyên vẹn |
| 7. Các cơ sở khái niệm | ISO/IEC TR 20547-1:2020, điều 7 | Chấp thuận nguyên vẹn |
| 7.1. Yêu cầu chung | ISO/IEC TR 20547-1:2020, khoản 7.1 | Chấp thuận nguyên vẹn |
| 7.2. Các khái niệm kiến trúc tham chiếu | ISO/IEC TR 20547-1:2020, khoản 7.2 | Chấp thuận nguyên vẹn |
| 7.3. Cấu trúc kiến trúc tham chiếu | ISO/IEC TR 20547-1:2020, khoản 7.3 | Chấp thuận nguyên vẹn |
| 8. Các yếu tố của kiến trúc tham chiếu dữ liệu lớn | ISO/IEC TR 20547-1:2020, điều 8 | Chấp thuận nguyên vẹn |
| 8.1. Tổng quan | ISO/IEC TR 20547-1:2020, khoản 8.1 | Chấp thuận nguyên vẹn |
| 8.2. Các bên liên quan | ISO/IEC TR 20547-1:2020, khoản 8.2 | Chấp thuận nguyên vẹn |
| 8.3. Các mối quan tâm | ISO/IEC TR 20547-1:2020, khoản 8.3 | Chấp thuận nguyên vẹn |
| 8.4. Các quan điểm | ISO/IEC TR 20547-1:2020, khoản 8.4 | Chấp thuận nguyên vẹn |
| 8.4.1. Quan điểm người dùng | ISO/IEC TR 20547-1:2020, điểm 8.4.1 | Chấp thuận nguyên vẹn |
| 8.4.2. Quan điểm chức năng | ISO/IEC TR 20547-1:2020, điểm 8.4.2 | Chấp thuận nguyên vẹn |
| 9. Quá trình áp dụng kiến trúc tham chiếu dữ liệu lớn | ISO/IEC TR 20547-1:2020, điều 9 | Chấp thuận nguyên vẹn |
| 9.1. Tổng quan | ISO/IEC TR 20547-1:2020, khoản 9.1 | Chấp thuận nguyên vẹn |
| 9.2. Xác định các bên liên quan và các mối quan tâm | ISO/IEC TR 20547-1:2020, khoản 9.2 | Chấp thuận nguyên vẹn |
| 9.3. Ánh xạ các bên liên quan và các mối quan tâm tới các vai trò và các vai trò phụ | ISO/IEC TR 20547-1:2020, khoản 9.3 | Chấp thuận nguyên vẹn |
| 9.4. Xây dựng các mô tả hoạt động chi tiếp và ánh xạ tới các mối quan tâm | ISO/IEC TR 20547-1:2020, khoản 9.4 | Chấp thuận nguyên vẹn |
| 9.5. Xác định các thành phần chức năng để thực hiện các hoạt động | ISO/IEC TR 20547-1:2020, khoản 9.5 | Chấp thuận nguyên vẹn |
| 9.6. Các hoạt động Crosswalk/các thành phần chức năng trở lại các mối quan tâm | ISO/IEC TR 20547-1:2020, khoản 9.6 | Chấp thuận nguyên vẹn |
| Thư mục tài liệu tham khảo |  | Chấp thuận sửa đổi, bổ sung một số tài liệu tham khảo |

# TÀI LIỆU THAM KHẢO

1. ISO/IEC TR 20547-1:2020 Information technology - Big data reference architecture - Part 1: Framework and application process
2. ISO/IEC TR 20547-1:2020 Information technology - Big data reference architecture - Part 1: Framework and application process
3. ISO/IEC TR 20547-2:2018 Information technology - Big data reference architecture - Part 2: Use cases and derived requirements
4. ISO/IEC 20547-3:2020 Information technology - Big data reference architecture - Part 3: Reference architecture
5. ISO/IEC 20547-4:2020 Information technology - Big data reference architecture - Part 4: Security and privacy
6. ISO/IEC TR 20547-5:2018 Information technology - Big data reference architecture - Part 5: Standards roadmap
7. [Y.3600](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3600) Big data - Cloud computing based requirements and capabilities
8. [Y.3601](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3601) Big data - framework and requirements for data exchange
9. [Y.3602](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3602) Big data - Functional requirements for data provenance
10. [Y.3603](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3603) Big data - Requirements and conceptual model of metadata for data catalogue

[Y.3604](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3604) Big data - Overview and requirements for data preservation

1. [Y.3605](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3605) Big data - Reference architecture
2. [Y.3606](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3606) Big data - Deep packet inspection mechanism for big data in network
3. [Y.3650](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3650) Framework of big-data-driven networking
4. [Y.3651](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3651) Big-data-driven networking - mobile network traffic management and planning
5. [Y.3652](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3652) Big data driven networking - requirements
6. [Y.3653](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3653) Big data driven networking - functional architecture
7. [Y.3654](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3654) Big data driven networking - Machine learning mechanism
8. [Y.3655](https://www.itu.int/rec/T-REC-Y/recommendation.asp?lang=en&parent=T-REC-Y.3655) Big data driven networking - Management and control mechanisms
9. [NIST SP 1500-1](https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-1.pdf) Volume 1: Definitions
10. [NIST SP 1500-2](https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-2.pdf) Volume 2: Taxonomies
11. [NIST SP 1500-3](https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-3.pdf) Volume 3: Use Case & Requirements
12. [NIST SP 1500-4](https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-4.pdf) Volume 4: Security and Privacy
13. [NIST SP 1500-5](https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-5.pdf) Volume 5: Architectures White Paper Survey
14. [NIST SP 1500-6](https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-6.pdf) Volume 6: Reference Architecture
15. [NIST SP 1500-7](https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-7.pdf) Volume 7: Standards Roadmap