

TCVN

TIÊU CHUẨN QUỐC GIA

**TCVN 7789-1 : 2007
ISO/IEC 11179-1 : 2004**

Xuất bản lần 1

**CÔNG NGHỆ THÔNG TIN - SỔ ĐĂNG KÝ
SIÊU DỮ LIỆU(MDR)
PHẦN 1: KHUNG CƠ CẤU**

Information technology — Metadata Registries (MDR) — Part 1: Framework

HÀ NỘI - 2007

Mục lục	Trang
Lời nói đầu	3
1 Phạm vi áp dụng.....	5
2 Tài liệu viện dẫn	5
3 Thuật ngữ và định nghĩa.....	5
3.1 Định nghĩa các cấu trúc mô hình hoá.....	6
3.2 Các thuật ngữ chung được sử dụng trong tiêu chuẩn này	6
3.3 Danh sách thuật ngữ theo bảng chữ cái được sử dụng trong siêu mô hình	11
3.4 Các thuật ngữ đặc tả được sử dụng trong tiêu chuẩn này	19
4 Từ viết tắt và từ cấu tạo bằng chữ đầu của nhóm từ	19
5 Lý thuyết về thuật ngữ học	20
6 Siêu dữ liệu	20
6.1 Khái niệm.....	20
6.2 Mô hình cơ sở về phần tử dữ liệu	21
6.3 Phần tử dữ liệu trong quản lý và trao đổi dữ liệu.....	23
6.4 Mô hình cơ sở của miền giá trị.....	24
6.5 Lược đồ phân loại cơ sở.....	28
7 Sổ đăng ký siêu dữ liệu	28
7.1 Khái quát.....	28
7.2 Mô hình khái quát cho một MDR theo TCVN 7789.....	29
7.3 Cơ sở đăng ký	30
8 Khái quát về bộ tiêu chuẩn TCVN 7789.....	32
8.1 Giới thiệu các phần	32
8.2 Các nguyên tắc cơ sở áp dụng TCVN 7789.....	35
9 Sự phù hợp.....	35

Lời nói đầu

TCVN 7789-1 : 2007 hoàn toàn tương đương với **ISO/IEC 11179-1 : 2004**

TCVN 7789-1 : 2007 do Ban Kỹ thuật Tiêu chuẩn TCVN/TC 154 "*Quá trình, các yếu tố dữ liệu và tài liệu trong thương mại, công nghiệp và hành chính*" biên soạn, Tổng cục Tiêu chuẩn Đo lường Chất lượng đề nghị, Bộ Khoa học và Công nghệ công bố.

Công nghệ thông tin - Sổ đăng ký siêu dữ liệu (MDR)

Phần 1: Khung cơ cấu

Information technology — Metadata Registries (MDR) - Part 1: Framework

1 Phạm vi áp dụng

Tiêu chuẩn này quy định kiểu và chất lượng siêu dữ liệu (metadata) cần thiết để mô tả dữ liệu và quy định việc quản trị và quản lý siêu dữ liệu đó trong một sổ đăng ký siêu dữ liệu (MDR). Tiêu chuẩn này áp dụng cho việc hình thành các cách trình bày dữ liệu, các khái niệm, ý nghĩa và mối quan hệ giữa chúng để chia sẻ giữa con người và máy móc, độc lập với tổ chức tạo ra dữ liệu. Tiêu chuẩn này không áp dụng cho cách biểu diễn mức vật lý của dữ liệu như các bit, byte tại mức máy.

Trong tiêu chuẩn này, siêu dữ liệu là các mô tả dữ liệu, tiêu chuẩn này không bao gồm việc xử lý về siêu dữ liệu. Tiêu chuẩn này cung cấp một phương tiện thông hiểu và liên kết với các phần riêng của tập tiêu chuẩn và là cơ sở để hiểu khái niệm về siêu dữ liệu và sổ đăng ký siêu dữ liệu.

2 Tài liệu viện dẫn

ISO 704:2000, Terminology work — Principles và methods (*Công tác xây dựng thuật ngữ - Nguyên tắc và phương pháp*)

ISO 1087-1:2000, Terminology work — Vocabulary — Part 1: Theory và application (*Công tác xây dựng thuật ngữ - Từ vựng - Phần 1: Lý thuyết và ứng dụng*)

TCVN 7789 (ISO/IEC 11179) (tất cả các phần), *Information technology — Metadata Registries (MDR)* (*Công nghệ thông tin – Sổ đăng ký siêu dữ liệu*)

3 Thuật ngữ và định nghĩa

Trong tiêu chuẩn này áp dụng các thuật ngữ và định nghĩa sau đây:

TCVN 7789-1 : 2007

3.1 Định nghĩa các cấu trúc mô hình hoá

Xác định các cấu trúc mô hình hoá được sử dụng trong tiêu chuẩn này.

3.1.1

Thuộc tính (attribute)

Đặc điểm của một **đối tượng** hoặc **thực thể**.

3.1.2

Lớp (class)

Mô tả một tập các **đối tượng** dùng chung **thuộc tính**, thao tác, phương pháp, **quan hệ** và ngữ nghĩa.

[ISO/IEC 19501-1:2001, 2.5.2.9].

3.1.3

Thẻ định danh (identifier) (trong Sổ đăng ký siêu dữ liệu)

Chuỗi các ký tự có khả năng xác định một cách duy nhất những gì được kết hợp với nó trong một ngữ cảnh cụ thể.

CHÚ THÍCH - Nên sử dụng một tên là một định danh vì tên rõ ràng về mặt ngôn ngữ học.

3.1.4

Quan hệ (relationship)

Kết nối giữa các phần tử của mô hình.

[ISO/IEC 19501-1:2001, 2.5.2.36].

3.2 Các thuật ngữ chung được sử dụng trong tiêu chuẩn này

Xác định các thuật ngữ sử dụng chung tuân theo các nhu cầu cụ thể của tiêu chuẩn này, không đặc tả việc cấu trúc mô hình hoá như trong mục 3.1

3.2.1

Thuộc tính cơ sở (basic attribute)

Thuộc tính của một **mục siêu dữ liệu** cần thiết trong việc đặc tả nó.

3.2.2

Đặc điểm (characteristic)

Sự trừu tượng hoá của một đặc tính của một **đối tượng** hoặc của một tập các đối tượng.

CHÚ THÍCH - Các đặc điểm được sử dụng để mô tả các khái niệm.

[ISO 1087-1:2000, 3.2.4].

3.2.3

Khái niệm (concept)

Đơn vị tri thức được tạo ra bởi một kết hợp duy nhất của các **đặc điểm**.

[ISO 1087-1:2000, 3.2.1].

3.2.4

Hệ thống khái niệm (concept system)

Tập các khái niệm được cấu trúc theo các mối quan hệ giữa chúng.

[ISO 1087-1:2000, 3.2.11]

3.2.5

Mô hình dữ liệu khái niệm (data model concept)

Mô hình khái niệm (model concept)

Mô hình dữ liệu biểu diễn một cách nhìn trừu tượng về thế giới thực.

CHÚ THÍCH - một mô hình khái niệm thể hiện cách hiểu của con người về một hệ thống.

3.2.6

Dữ liệu (data)

Biểu diễn có thể dịch lại của thông tin theo một dạng thức được hình thức hóa phù hợp với truyền đạt, thông dịch hoặc xử lý.

TCVN 7789-1 : 2007

CHÚ THÍCH - Dữ liệu có thể được xử lý bởi con người hoặc các phương tiện tự động.

[ISO 2382-1:1993, 01.01.02].

3.2.7

Mô hình dữ liệu (data model)

Biểu diễn bằng từ và/hoặc hình của **dữ liệu**, đặc tả các đặc tính, cấu trúc và các mối quan hệ giữa chúng.

3.2.8

Định nghĩa (definition)

Biểu diễn của một **khái niệm** bằng một khái niệm mô tả sự khác nhau giữa khái niệm này và các khái niệm liên quan.

[ISO 1087-1:2000, 3.3.1].

3.2.9

Designation (ký hiệu quy ước)

Biểu diễn của một khái niệm bởi dấu hiệu diễn đạt nó.

[ISO 1087-1:2000, 3.4.1].

3.2.10

Thực thể (entity)

Sự vật cụ thể hoặc trừu tượng đang tồn tại, đã tồn tại, hoặc có thể tồn tại gồm cả các mối liên kết giữa các sự vật này.

Ví dụ: Một người, một đối tượng, một sự kiện, một quan điểm, một quá trình, v..v.

CHÚ THÍCH - Một thực thể vẫn tồn tại dù dữ liệu về nó có hay không có.

[ISO/IEC 2382-17:1999, 17.02.05].

3.2.11

Đặc điểm chủ yếu (essential characteristic)

Đặc điểm mà không thể thiếu để hiểu một **khái niệm**.

[ISO 1087-1:2000, 3.2.6].

3.2.12

Mở rộng (extension) (danh từ)

<thuật ngữ>

Toàn bộ các **đối tượng** mà một khái niệm tương ứng với chúng.

[ISO 1087-1:2000, 3.2.8].

CHÚ THÍCH - Thuật ngữ này khác với thuật ngữ trong TCVN 7789-3.

3.2.13

Khái niệm chung (general concept)

Khái niệm tương ứng với hai hay nhiều **đối tượng** được tạo thành một nhóm bởi các đặc tính chung.

CHÚ THÍCH - Các ví dụ của các **khái niệm chung** là 'hành tinh', 'tháp'.

[ISO 1087-1:2000, 3.2.3]

3.2.14

Khái niệm riêng (individual concept)

Khái niệm tương ứng với chỉ một **đối tượng**.

CHÚ THÍCH - Các ví dụ của các khái niệm riêng là: 'sao thổ', 'tháp Eiffel'.

[ISO 1087-1:2000, 3.2.2].

3.2.15

Tập trung (intension) (danh từ)

<thuật ngữ>

Tập các **đặc điểm** tạo nên **khái niệm**

TCVN 7789-1 : 2007

[ISO 1087-1:2000, 3.2.9].

3.2.16

Siêu dữ liệu (metadata)

Dữ liệu để định nghĩa và mô tả **dữ liệu** khác.

3.2.17

Mục siêu dữ liệu (metadata item)

Thể hiện cụ thể của một đối tượng siêu dữ liệu.

3.2.18

Đối tượng siêu dữ liệu (metadata object)

Kiểu đối tượng được định nghĩa bởi một siêu mô hình.

3.2.19

Sổ đăng ký siêu dữ liệu (metadata registry)

MDR

Hệ thống thông tin để đăng ký **siêu dữ liệu**.

3.2.20

Siêu mô hình (metamodel)

Mô hình dữ liệu để đặc tả một hoặc nhiều mô hình dữ liệu khác.

3.2.21

Tên (name)

Ký hiệu quy ước của một **đối tượng** được diễn đạt bằng ngôn ngữ.

3.2.22

Đối tượng (object)

Bất kỳ sự vật nào có thể nhận biết hoặc hiểu được.

CHÚ THÍCH - Các đối tượng có thể là vật chất (như là; động cơ, trang tính, viên kim cương), phi vật chất (như là; tỷ lệ quy ước, kế hoạch dự án), hoặc do tưởng tượng (như là: con kỳ lân).

[ISO 1087-1:2000, 3.1.1].

3.2.23

Mục đăng ký (registry item)

Mục siêu dữ liệu được ghi lại trong một **sổ đăng ký siêu dữ liệu**.

3.2.24

Siêu mô hình đăng ký (registry metamodel)

Siêu mô hình đặc tả một **sổ đăng ký siêu dữ liệu**.

3.2.25

Hệ thống thuật ngữ (terminological system)

Hệ thống khái niệm với các **kí hiệu quy ước** cho mỗi **khái niệm**.

3.3 Danh sách thuật ngữ theo bảng chữ cái được sử dụng trong siêu mô hình

Điều này đưa ra các định nghĩa về các thuật ngữ được sử dụng trong tiêu chuẩn, cũng là tên của các đối tượng siêu dữ liệu trong siêu mô hình được đặc tả trong TCVN 7789-3.

3.3.1

Mục được quản trị (administered item)

Mục đăng ký về thông tin quản trị được ghi lại trong một **bản ghi quản trị**.

3.3.2

Bản ghi quản trị (administration record)

Tập thông tin quản trị cho một **mục được quản trị**.

3.3.3

Trạng thái quản trị (administrative status)

Ký hiệu quy ước ghi nhận trạng thái trong quá trình quản trị của một **tổ chức đăng ký có thẩm quyền** để quản lý các yêu cầu đăng ký.

CHÚ THÍCH - Các giá trị và các ý nghĩa tương ứng của “trạng thái quản trị” được xác định bởi mỗi **tổ chức đăng ký có thẩm quyền**.

C.f. “**trạng thái đăng ký**”.

3.3.4

Lược đồ phân loại (classification scheme)

Thông tin mô tả về việc sắp xếp hoặc phân chia các **đối tượng** thành các nhóm dựa trên các **đặc điểm** chung của các đối tượng đó.

3.3.5

Mục lược đồ phân loại (classification scheme item)

CSI

Mục nội dung trong một **lược đồ phân loại**.

CHÚ THÍCH – Điều này có thể là một nút trong một nguyên tắc phân loại hoặc bản thể học, một thuật ngữ trong từ điển chuyên đề, v..v.

3.3.6

Miền khái niệm (conceptual domain)

CD

Tập các ý nghĩa giá trị hợp lệ

CHÚ THÍCH - Ý nghĩa của giá trị trong một **miền khái niệm** có thể được đánh số hoặc được biểu diễn qua một mô tả.

3.3.7**Ngữ cảnh** (context)

Hoàn cảnh, mục đích và viễn cảnh trong đó một đối tượng được định nghĩa và sử dụng.

CHÚ THÍCH - Thuật ngữ này có nghĩa khác với ý nghĩa trong 7789-3.

3.3.8**Phần tử dữ liệu** (data element)**DE**

Đơn vị dữ liệu mà việc **định nghĩa**, định danh, biểu diễn và **các giá trị cho phép** của nó được đặc tả bởi phương tiện của một tập các **thuộc tính**.

3.3.9**Khái niệm phần tử dữ liệu** (data element concept)**DEC**

Khái niệm có thể được biểu diễn dưới dạng một **phần tử dữ liệu** được mô tả độc lập với bất kỳ cách biểu diễn cụ thể nào.

3.3.10**Định danh dữ liệu** (data identifier)**DI**

Thẻ định danh duy nhất cho một **mục được quản trị** trong một **tổ chức đăng ký có thẩm quyền**.

3.3.11**Kiểu dữ liệu** (datatype)

Tập các giá trị phân biệt được đặc trưng bởi các đặc tính và các thao tác trên các giá trị đó.

[ISO/IEC 11404:1996, 4.11].

3.3.12

Thứ nguyên (dimensionality)

Biểu thức đo lường không có đơn vị.

CHÚ THÍCH - Một lượng là một giá trị được gán với một đơn vị đo. 32° Fa-ren-hét, 0° Celsius, \$100 USD và 10 gam giấy là các số lượng. Sự tương đương giữa hai đơn vị đo được xác định bởi sự tồn tại của một đại lượng tương quan 1-1 giữa các giá trị được đo theo một đơn vị đo và các giá trị được đo theo đơn vị đo khác, độc lập với ngữ cảnh và khi các thao tác đặc trưng là như nhau. Các đơn vị đo tương đương theo ý nghĩa này đều là thứ nguyên. Sự tương đương này được xác định ở đây tạo thành một quan hệ tương đương trên tập tất cả các đơn vị đo. Mỗi lớp tương đương tương ứng với thứ nguyên. Các đơn vị đo "nhiệt độ theo độ Fahrenheit" và "nhiệt độ theo độ Celsius" có cùng thứ nguyên, bởi vì đối với mỗi giá trị được đo theo độ Fahrenheit có một giá trị được đo theo độ Celsius với cùng lượng và ngược lại. Các thao tác giống nhau có thể được thực hiện theo các lượng đơn vị đo. Các tương quan 1-1 được bảo toàn lượng là các phương trình đã biết là $C^{\circ} = (5/9)(F^{\circ} - 32)$ và $F^{\circ} = (9/5)(C^{\circ}) + 32$.

3.3.13

Miền khái niệm liệt kê (enumerated concept domain)

Miền khái niệm được đặc tả bởi danh sách của toàn bộ các **ý nghĩa của giá trị** của nó.

3.3.14

Miền giá trị liệt kê (enumerated value domain)

Miền giá trị được đặc tả bởi danh sách của toàn bộ các **giá trị cho phép** của nó.

3.3.15

Ký hiệu quy ước mã quốc tế (international code designator)

ICD

Định danh của một lược đồ định danh tổ chức.

CHÚ THÍCH - Trên cơ sở ISO/IEC 6523-1:1998, 3.8.

3.3.16

Thẻ định danh mục (item identifier)

Thẻ định danh cho một mục.

3.3.17

Thẻ định danh của tổ chức đăng ký có thẩm quyền mục (item registration authority identifier)

Thẻ định danh của tổ chức đăng ký có thẩm quyền mục đó.

3.3.18

Miền khái niệm phi liệt kê (non-enumerated conceptual domain)

Miền khái niệm không được đặc tả bởi một danh sách của tất cả ý nghĩa giá trị hợp lệ.

3.3.19

Mô tả miền khái niệm phi liệt kê (non-enumerated conceptual domain description)

Mô tả hoặc đặc tả một quy tắc, tham chiếu, hoặc một dải đối với một tập toàn bộ các **ý nghĩa của giá trị** cho miền khái niệm đó.

3.3.20

Miền giá trị phi liệt kê (non-enumerated value domain)

Miền giá trị được đặc tả bởi một mô tả hơn là một danh sách của toàn bộ các **giá trị cho phép**.

3.3.21

Mô tả miền giá trị phi liệt kê (non-enumerated value domain description)

Mô tả hoặc đặc tả của một quy tắc, tham chiếu hoặc dải đối với một tập các tất cả các **giá trị cho phép** đối với **miền giá trị** đó.

3.3.22

Lớp đối tượng (object class)

Tập các quan điểm, khái niệm trừu tượng, hoặc sự vật trong thế giới thực được định danh với ranh giới, ý nghĩa rõ ràng, các đặc tính và hành vi của nó có cùng quy tắc.

3.3.23

Tổ chức (organization)

TCVN 7789-1 : 2007

Khung cơ cấu duy nhất của tổ chức có thẩm quyền trong đó một hoặc nhiều người hành động hoặc được quy ước hành động hướng vào mục đích nào đó.

[ISO/IEC 6523-1:1998, 3.1].

3.3.24

Thẻ định danh tổ chức (organization identifier)

Thẻ định danh được ấn định cho một tổ chức trong lược đồ định danh tổ chức và duy nhất trong lược đồ đó.

[ISO/IEC 6523-1:1998, 3.10].

3.3.25

Bộ phận của tổ chức (organization part)

Bộ phận, dịch vụ hoặc **thực thể** khác trong một **tổ chức** cần được định danh để trao đổi thông tin.

[ISO/IEC 6523-1:1998, 3.2].

3.3.26

Thẻ định danh bộ phận của tổ chức (organization part identifier)

OPI

Thẻ định danh được cấp cho **bộ phận tổ chức** cụ thể.

[ISO/IEC 6523-1:1998, 3.11].

3.3.27

Nguồn định danh bộ phận tổ chức (organization part identifier source)

Nguồn để **định danh bộ phận tổ chức**.

CHÚ THÍCH - Trên cơ sở ISO/IEC 6523-1:1998, 3.12.

3.3.28

Giá trị cho phép (permissible value)

Biểu thức của một **ý nghĩa giá trị** cho phép trong một **miền giá trị** cụ thể.

3.3.29

Đặc tính (property)

Đặc điểm chung cho tất cả thành viên của một **lớp đối tượng**

3.3.30

Người giữ sổ đăng ký (registrar)

Người đại diện của một **tổ chức đăng ký có thẩm quyền đăng ký**.

3.3.31

Đăng ký (registration)

Mối quan hệ giữa một **mục được quản trị** và **tổ chức đăng ký có thẩm quyền đăng ký**.

3.3.32

Tổ chức có thẩm quyền đăng ký (registration authority)

RA

Tổ chức có trách nhiệm duy trì một **sổ đăng ký**.

3.3.33

Thẻ định danh tổ chức có thẩm quyền (registration authority identifier)

Thẻ định danh được ấn định cho một **tổ chức có thẩm quyền đăng ký**.

3.3.34

Trạng thái đăng ký (registration status)

Ký hiệu quy ước về trạng thái trong chu kỳ đăng ký của một **mục được quản trị**.

3.3.35

Lớp biểu diễn (representation class)

TCVN 7789-1 : 2007

Phân loại các kiểu biểu diễn.

3.3.36

Đơn vị đo (measurement unit)

Các đơn vị thực mà trong đó các giá trị tương ứng được đo.

CHÚ THÍCH - **thứ nguyên** của **miền khái niệm** tương ứng phải phù hợp với **đơn vị đo** cụ thể.

3.3.37

Giá trị (value)

Giá trị của **dữ liệu**.

3.3.38

Miền giá trị (value domain)

VD

Tập các giá trị cho phép.

3.3.39

Ý nghĩa của giá trị (value meaning)

Ý nghĩa hoặc nội dung ngữ nghĩa của một **giá trị**.

CHÚ THÍCH - Cho trước một **giá trị cho phép**, thì việc biểu diễn **ý nghĩa giá trị** của nó phải độc lập (và không bắt buộc) với cách biểu diễn **giá trị** tương ứng của nó.

3.3.42

Mô tả ý nghĩa giá trị (value meaning description)

Mô tả một **ý nghĩa giá trị**.

3.3.43

Phiên bản (version)

Định danh phiên bản duy nhất của mục được quản trị.

3.4 Các thuật ngữ đặc tả được sử dụng trong tiêu chuẩn này

Điều này định nghĩa các thuật ngữ có cách sử dụng quy định trong tiêu chuẩn này và không được sử dụng trong các phần của tiêu chuẩn khác.

3.4.1

Cấu trúc dữ liệu (data construct)

Đối tượng mà **mục tiêu dữ liệu** mô tả.

CHÚ THÍCH - Các phần tử dữ liệu riêng, miền giá trị, các khái niệm phần tử dữ liệu, các miền khái niệm, các lớp các đối tượng và các đặc tính là các cấu trúc dữ liệu.

3.4.2

Lượng (Quantity)

Giá trị tương ứng với một đơn vị đo lường.

4 Từ viết tắt và từ cấu tạo bằng chữ đầu của nhóm từ

CHÚ THÍCH - Một số từ viết tắt và hoặc từ được cấu tạo từ chữ cái đầu của cụm từ trong điều này trình bày định nghĩa các thuật ngữ trong điều 3.

CD – Miền khái niệm (Conceptual Domain)

DE – Phần tử dữ liệu (Data Element)

DEC – Khái niệm phần tử dữ liệu (Data Element Concept)

DI – Thẻ định danh dữ liệu (Data Identifier)

EDI – Trao đổi dữ liệu điện tử (Electronic Data Interchange)

IEC – Ủy ban Kỹ thuật điện quốc tế (International Electrotechnical Commission)

ICD – Hệ ký hiệu quy ước mã quốc tế (International Code Designator)

ISO – Tổ chức Tiêu chuẩn hóa quốc tế (International Organization for Standardization)

JTC1 – Ban kỹ thuật liên kết giữa ISO và IEC về CNTT (Joint Technical Committee 1)

MDR – Sổ đăng ký siêu dữ liệu (Metadata Registries)

TCVN 7789-1 : 2007

OPI – Thẻ định danh bộ phận của tổ chức (Organization Part Identifier)

RA – Cơ quan đăng ký (Registration Authority)

SC32 – Tiểu BKT ISO/IEC JTC1/SC32 của ISO (ISO/IEC JTC1/Sub-committee 32)

5 Lý thuyết về thuật ngữ học

Các khái niệm của thuyết thuật ngữ học được sử dụng trong TCVN 7789 phải phù hợp với ISO 704 và ISO 1087-1. Một mô tả ngắn gọn về các lý thuyết cần thiết như sau.

Trong lý thuyết về thuật ngữ, một đối tượng là những gì có thể nhận biết hoặc hiểu được. Các khái niệm là các cấu trúc trí óc, đơn vị của ý nghĩ, hoặc đơn vị của tri thức được tạo bởi một sự kết hợp duy nhất của các đặc điểm. Các khái niệm được tổ chức hoặc được hợp thành nhóm bởi các phần tử chung, được gọi là các đặc điểm. Các đặc điểm chủ yếu không thể thiếu để hiểu một khái niệm. Các đặc điểm khác là thứ yếu. Nội dung tổng quát của các đặc điểm cấu thành nên một khái niệm được gọi là sự đúc kết của nó. Tập các đối tượng của một khái niệm gọi là phần mở rộng của nó.

Trong ngôn ngữ tự nhiên, các khái niệm được thể hiện thông qua các định nghĩa quy định một bản chất bên trong và phần mở rộng duy nhất.

Một hệ thống ký hiệu (thuật ngữ, tên gọi hoặc ký hiệu) trình bày một khái niệm.

Khái niệm chung có hai hoặc nhiều đối tượng tương ứng với nó. Khái niệm riêng có một đối tượng tương ứng với nó. Như vậy, **khái niệm chung** có hai hoặc nhiều đối tượng trong phần mở rộng và một khái niệm riêng có một đối tượng trong phần mở rộng của nó.

Một hệ thống khái niệm là tập các khái niệm được cấu trúc theo các mối quan hệ giữa chúng. Hệ thống thuật ngữ là một hệ thống khái niệm cùng với các hệ thống ký hiệu cho một khái niệm.

6 Siêu dữ liệu

6.1 Khái niệm

Trong bộ tiêu chuẩn TCVN 7789, siêu dữ liệu được định nghĩa là dữ liệu để định nghĩa và mô tả dữ liệu khác. Điều đó có nghĩa siêu dữ liệu là dữ liệu và dữ liệu trở thành siêu dữ liệu khi chúng được sử dụng theo cách này. Điều đó xảy ra trong các hoàn cảnh cụ thể, với mục đích cụ thể và cùng với viễn cảnh nào đó, vì không phải dữ liệu nào cũng luôn là siêu dữ liệu. Tập các hoàn cảnh, mục đích, hoặc viễn cảnh đối với dữ liệu mà trong đó được sử dụng như siêu dữ liệu được gọi là ngữ cảnh. Bởi vậy, siêu dữ liệu là dữ liệu về dữ liệu trong ngữ cảnh nào đó.

Vì siêu dữ liệu là dữ liệu nên siêu dữ liệu có thể được lưu trữ trong một cơ sở dữ liệu và được tổ chức thông qua việc sử dụng mô hình. Một số mô hình có ứng dụng rất cụ thể và một số mô hình khác khái quát hơn. Mô hình được trình bày và được mô tả trong TCVN 7789-3 (Siêu mô hình đăng ký và các thuộc tính cơ sở) mang tính khái quát. Mô hình là một sự biểu diễn về hiểu biết của con người về siêu dữ liệu cần thiết để mô tả các cấu trúc dữ liệu, bao gồm các mối quan hệ tồn tại giữa siêu dữ liệu đó và không cần thiết bao gồm cách thức trình bày siêu dữ liệu đó trong một ứng dụng của một MDR. Mô hình loại này được gọi mô hình khái niệm. Các mô hình khái niệm là phương tiện để con người đọc và hiểu.

Các mô hình để mô tả siêu dữ liệu thường được gọi là các siêu mô hình. Mô hình khái niệm được trình bày trong TCVN 7789-3 là một siêu mô hình theo ý nghĩa này.

6.2 Mô hình cơ sở về phần tử dữ liệu

Hình 1 minh họa các quan điểm gắn với các mệnh đề trong mục này. Bản thân hình này không phải là một chuẩn mực mà chỉ được sử dụng để minh họa các quan điểm cơ sở.

Trong tiêu chuẩn này, một phần tử dữ liệu bao gồm hai phần:

Khái niệm phần tử dữ liệu – DEC là **khái niệm** có thể trình bày dưới dạng một **phần tử dữ liệu**, được mô tả một cách độc lập với bất kỳ cách biểu diễn cụ thể nào.

Biểu diễn – bao gồm một miền giá trị, kiểu dữ liệu, các đơn vị đo (nếu cần) và lớp biểu diễn (tùy chọn).

Từ một viễn cảnh mô hình hoá dữ liệu và đối với mục đích của TCVN 7789, một khái niệm phần tử dữ liệu có thể bao gồm hai phần:

- **Lớp đối tượng** là một tập các quan niệm, quan điểm trừu tượng hoặc sự việc trong thế giới thực mà có thể được định danh với các ranh giới và ý nghĩa rõ ràng và các đặc tính và hành vi của nó có cùng các quy tắc.
- **Đặc tính** là một đặc điểm chung cho tất cả thành viên của một lớp đối tượng.

Các lớp đối tượng là các sự vật mà chúng ta mong muốn tập hợp và lưu trữ dữ liệu về chúng. Chúng cũng là các khái niệm và chúng tương ứng với các cú pháp được sử dụng trong mô hình hướng đối tượng và các thực thể trong các mô hình quan hệ thực thể. Ví dụ; xe hơi, người, người trong gia đình, người làm công và người đặt hàng. Các đặc tính là điều mà con người sử dụng để phân biệt hoặc mô tả các đối tượng. Chúng là các đặc điểm, không nhất thiết là các đặc điểm thiết yếu của lớp đối tượng và tạo thành tập các đặc điểm tạo nên khái niệm. Chúng cũng là các khái niệm và liên kết tương ứng với các cú pháp được sử dụng trong các thuộc tính (không cần các kiểu dữ liệu liên kết) trong các mô hình quan hệ thực thể hoặc hướng đối tượng. Các ví dụ về đặc tính là màu sắc, kiểu dáng, giới tính, tuổi, thu nhập, địa chỉ hoặc giá.

TCVN 7789-1 : 2007

Một lớp đối tượng có thể là một **khái niệm chung**. Điều đó xảy ra khi tập các đối tượng tương ứng với lớp đối tượng có hai hoặc nhiều thành phần. Các ví dụ trong đoạn trước thuộc về kiểu này. Dữ liệu kiểu bản ghi được mô tả theo cách này. Mặt khác, một lớp đối tượng cũng có thể là một khái niệm riêng. Điều này xảy ra khi tập các đối tượng tương ứng với lớp đối tượng có một thành phần. Ví dụ là các khái niệm tương ứng với các đối tượng đơn, "những người sống ở Mỹ" hoặc "tập các cơ sở dịch vụ thiết lập ở Australia". Dữ liệu tập hợp được mô tả theo cách này. Các ví dụ về các đặc tính là thu nhập trung bình hoặc tổng thu nhập.

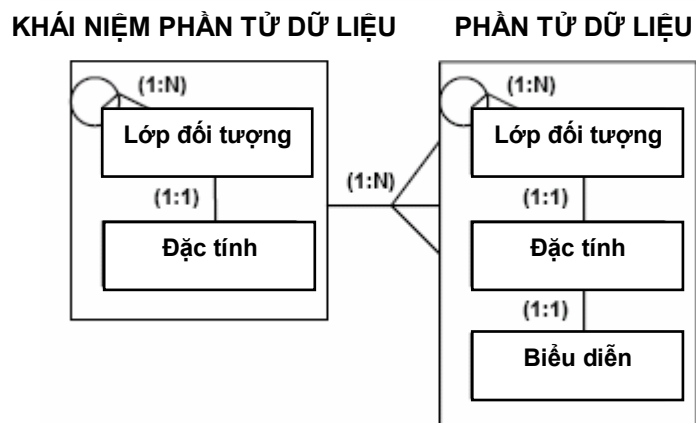
Điều quan trọng là phân biệt lớp đối tượng thực hoặc đặc tính với tên của nó. Đây là sự phân biệt giữa khái niệm và ký hiệu quy ước của chúng. Lớp các đối tượng và đặc tính là các khái niệm; tên của chúng là hệ thống ký hiệu. Sự phức tạp tăng lên do con người truyền đạt các khái niệm qua ngôn từ (hệ thống ký hiệu) và rất dễ nhầm lẫn một khái niệm với hệ thống ký hiệu được sử dụng trình bày khái niệm đó. Ví dụ, phần lớn mọi người sẽ đọc từ thu nhập mà chắc chắn họ hiểu nó một cách rõ ràng. Nhưng, ký hiệu thu nhập không thể hiện cùng khái niệm đối với tất cả người đọc và quan trọng hơn mỗi trường hợp của thu nhập có thể không chỉ rõ cùng một khái niệm.

Không phải tất cả các quan điểm đều được diễn tả đơn giản theo ngôn ngữ tự nhiên. Ví dụ, "Phụ nữ tuổi từ 15 đến 45 có ít nhất một lần sinh nhật trong vòng 12 tháng gần đây nhất" là một lớp đối tượng hợp lệ không dễ để đặt tên theo tiếng Việt. Một số quan điểm có thể được thể hiện dễ dàng hơn trong một ngôn ngữ so với ngôn ngữ khác. Từ *Götterdämmerung* trong tiếng Đức không có từ tương đương đơn giản trong Tiếng Anh.

Một phần tử dữ liệu được tạo ra khi một sự biểu diễn được gắn với một khái niệm phần tử dữ liệu. Cách biểu diễn này mô tả hình thức của dữ liệu đó, bao gồm miền giá trị, kiểu dữ liệu, lớp biểu diễn (tùy chọn), và nếu cần, một đơn vị đo. **Miền giá trị** là tập các giá trị cho phép đối với các phần tử dữ liệu. Ví dụ, phần tử dữ liệu về thu nhập hàng năm của người trong cùng gia đình có thể là tập các số nguyên không âm (đơn vị là US\$) như một tập các giá trị hợp lệ. Đây là miền giá trị của nó.

Một khái niệm phần tử dữ liệu có thể được gắn với các miền giá trị khác nhau khi cần thiết để tạo ra các phần tử dữ liệu tương tự theo khái niệm. Có nhiều cách để trình bày các sự kiện tương tự trong thế giới thực, nhưng khái niệm chỉ các sự kiện đó là giống nhau. Lấy DEC về nơi sinh của một người là một ví dụ. ISO 3166-1 – Mã nước bao gồm bảy cách biểu diễn khác nhau đối với các nước trên thế giới. Mỗi mã trong bảy cách biểu diễn này bao gồm một tập các giá trị có thể được sử dụng trong miền giá trị tương ứng với DEC đó. Mỗi một trong bảy cách tương ứng là một phần tử dữ liệu. Đối với mỗi cách biểu diễn của dữ liệu, các giá trị được phép, kiểu dữ liệu, lớp biểu diễn và các đơn vị đo có thể được thay đổi.

Xem ISO/IEC 20943-1:2003, *Công nghệ thông tin — Thủ tục hoàn thành tính nhất quán nội dung sổ đăng ký siêu dữ liệu— Phần 1: Phần tử dữ liệu về chi tiết việc đăng ký và quản lý các mô tả của phần tử dữ liệu*.



Hình này chỉ để tham khảo, không quy định.

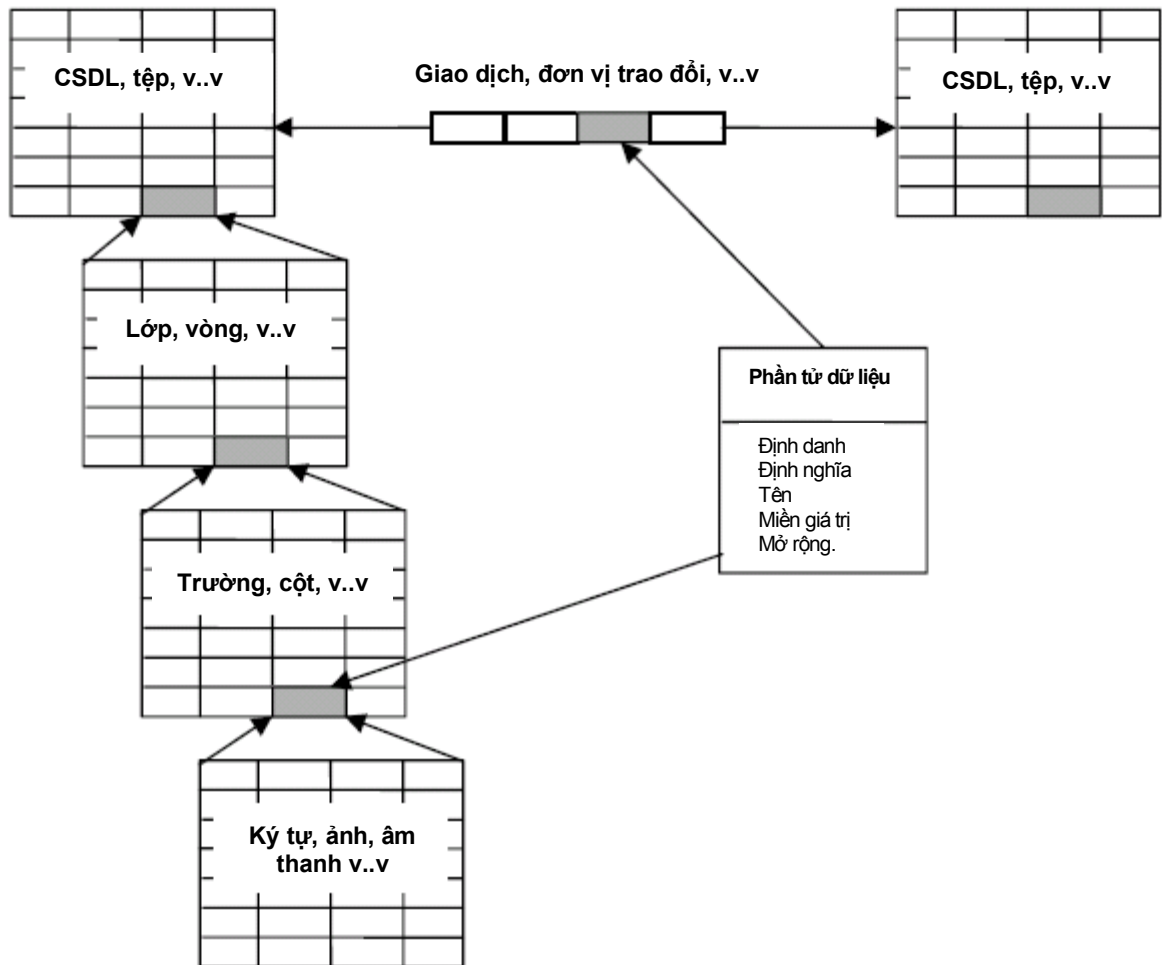
Hình 1 — Mô hình cơ sở cho phần tử dữ liệu

6.3 Phần tử dữ liệu trong quản lý và trao đổi dữ liệu

Hình 2 cung cấp một hình đơn giản để minh họa các trường hợp mà các phần tử dữ liệu thuộc về. Các phần tử dữ liệu xuất hiện trong các cơ sở dữ liệu, các tệp và các tập giao dịch. Các phần tử dữ liệu là các khối đơn vị cơ sở của dữ liệu mà một tổ chức quản lý, vì vậy, chúng phải là một phần của việc thiết kế các cơ sở dữ liệu và các tệp trong tổ chức đó và toàn bộ các tập giao dịch mà tổ chức đó xây dựng để truyền đạt dữ liệu cho các tổ chức khác.

Trong tổ chức đó, các cơ sở dữ liệu hoặc các tệp bao gồm các bản ghi, các đoạn, vòng, v.v., chứa các phần tử dữ liệu. Các phần tử dữ liệu đó tự chúng bao gồm các loại dữ liệu khác nhau như; ký tự, ảnh, âm thanh, v.v.

Khi tổ chức cần truyền tải dữ liệu sang một tổ chức khác, các phần tử dữ liệu các khối đơn vị cơ sở cấu thành nên các tập giao dịch. Các giao dịch xuất hiện chủ yếu giữa các cơ sở dữ liệu hoặc các tệp, trừ các cấu trúc (như là; các bản ghi hoặc vòng) của các tệp và các cơ sở dữ liệu đó mà không có trong cùng các tổ chức. Vì vậy, khối đơn vị chung để truyền tải thông tin (dữ liệu cộng với việc thông hiểu) là phần tử dữ liệu đó.



Hình này chỉ để tham khảo, không quy định.

Hình 2 — Các phần tử dữ liệu và các khái niệm dữ liệu khác

6.4 Mô hình cơ sở của miền giá trị

Hình 3 minh họa các quan niệm trong phạm vi của mục này. Bản thân hình này không chuẩn mực, nó chỉ được sử dụng để minh họa các quan niệm cơ sở.

Một miền giá trị là một tập các giá trị cho phép. Một giá trị cho phép là một kết hợp của một số giá trị và ý nghĩa của chúng. Ý nghĩa được kết hợp đó được gọi là ý nghĩa của giá trị. Một miền giá trị là tập các giá trị hợp lệ cho một hoặc nhiều các phần tử dữ liệu. Nó được sử dụng để giám sát sự hợp lệ của dữ liệu trong các hệ thống thông tin và trong trao đổi dữ liệu. Nó cũng là một phần không thể thiếu của siêu dữ liệu cần để mô tả một phần tử dữ liệu. Cụ thể, một miền giá trị là một hướng dẫn nội dung, hình thức và cấu trúc của dữ liệu được trình bày bởi một phần tử dữ liệu.

Miền giá trị gồm có hai kiểu nhỏ (không loại trừ):

- **Miền giá trị liệt kê** – Một miền giá trị được quy định như một danh sách của các giá trị cho phép (các giá trị và ý nghĩa của chúng)
- **Miền giá trị phi liệt kê** – Một miền giá trị được quy định bởi một mô tả.

Một miền giá trị liệt kê bao gồm một danh sách của toàn bộ các giá trị của nó và các ý nghĩa tương ứng của nó. Mỗi cặp giá trị và ý nghĩa được gọi là một giá trị cho phép. Ý nghĩa đối với mỗi giá trị được gọi ý nghĩa của giá trị.

Một miền giá trị phi liệt kê được quy định bởi một mô tả. Mô tả miền giá trị phi liệt kê mô tả một cách chính xác các giá trị cho phép thuộc vào và miền giá trị nó không thuộc vào. Ví dụ: một mô tả là một cụm từ "tất cả các số thực lớn hơn 0 và nhỏ hơn 1".

Mỗi miền giá trị là một thành phần của phần mở rộng của một khái niệm, được gọi là miền khái niệm. Một miền khái niệm là một tập các ý nghĩa của các giá trị. Bản chất bên trong của một miền khái niệm là ý nghĩa của các giá trị của nó. Nhiều miền giá trị có thể trong phần mở rộng của cùng miền khái niệm, trừ một miền giá trị tương ứng với một miền khái niệm. Các miền khái niệm có thể có các mối quan hệ các miền khái niệm khác, vì vậy, có thể tạo ra một hệ thống khái niệm của các miền khái niệm. Miền giá trị có thể có các mối quan hệ với miền giá trị khác tạo nên một khuôn khổ để bắt lấy cấu trúc của tập các miền giá trị liên quan và các khái niệm tương ứng của chúng.

Các miền khái niệm cũng có hai loại nhỏ (không loại trừ):

- **Miền khái niệm liệt kê** – Một miền khái niệm được quy định như một danh sách của ý nghĩa của các giá trị;
- **Miền khái niệm phi liệt kê** – Một miền khái niệm được quy định bởi một mô tả.

Ý nghĩa của các giá trị đối với một miền khái niệm liệt kê được liệt kê rõ ràng. Kiểu miền khái niệm tương ứng với kiểu tập hợp đối với miền giá trị. Ý nghĩa của các giá trị đối với một miền khái niệm phi liệt kê được thể hiện bằng cách sử dụng một quy tắc, được gọi là mô tả miền khái niệm phi liệt kê. Vì vậy, ý nghĩa của các giá trị không được liệt kê rõ ràng mà hàm ẩn. Quy tắc này mô tả ý nghĩa của các giá trị cho phép trong một phi-miền giá trị liệt kê. Kiểu miền khái niệm này tương ứng với kiểu không tập hợp đối với miền giá trị. Xem ISO/IEC TR 20943-3, Information technology — Procedures for achieving metadata registry content consistency — Part 3: Value domain for detailed examples (*Công nghệ thông tin — Các thủ tục đạt nội dung nhất quán sổ đăng ký siêu dữ liệu — Phần 3: Miền giá trị cho các minh họa chi tiết*).

Một đơn vị đo thỉnh thoảng được yêu cầu để mô tả dữ liệu. Nếu nhiệt độ được ghi trong cơ sở dữ liệu, thì thang nhiệt độ (như là; Fa-ren-hét hoặc độ C) cần thiết để hiểu ý nghĩa của các giá trị đó. Một ví dụ khác: là khối lượng các viên đá được tìm thấy ở Sao Hỏa, được đo theo gram. Tuy nhiên, các

TCVN 7789-1 : 2007

Đơn vị đo không chỉ được dùng cho các đại lượng vật lý, mà còn cho các đại lượng như tiền tệ (ví dụ; US dollars, Lire, British pounds) và các thước đo kinh tế xã hội khác cũng là các đơn vị đo.

Một số các đơn vị đo là tương đương với nhau theo ý nghĩa sau đây: Bất kỳ lượng nào theo đơn vị đo này cũng có thể biến đổi sang cùng một lượng theo một đơn vị đo khác. Toàn bộ các đơn vị đo tương đương được gọi là có cùng thứ nguyên. Ví dụ, Tất cả tiền tệ có cùng thứ nguyên. Đo tốc độ, như dặm/h hoặc m/s có cùng thứ nguyên. Hai đơn vị đo thường bị hiểu nhầm là có cùng thứ nguyên là pounds (trong trọng lượng) và grams. Pounds là đo lực và grams là đo khối lượng.

Một đơn vị đo tương ứng với một miền giá trị và thứ nguyên tương ứng với miền khái niệm.

Một vài miền giá trị chia các giá trị tương tự khi chuyển từ một miền tới một miền khác. Hoặc chính các giá trị là tương tự hoặc ý nghĩa của các giá trị là giống nhau. Khi sự tương tự này xảy ra, miền giá trị có thể nằm trong phần mở rộng của một miền khái niệm. Các ví dụ sau đây minh họa điều này và các ý nghĩa khác trong điều này:

Ví dụ: 1 – Miền giá trị phi liệt kê tương đương

Tên miền khái niệm: Xác suất

Định nghĩa miền khái niệm: các số thực lớn hơn 0 nhỏ hơn 1.

Tên miền giá trị (1): Xác suất – 2 chữ số có nghĩa

Mô tả miền giá trị: Toàn bộ các số thực lớn hơn 0 nhỏ hơn 1 được biểu diễn với cấp chính xác 2 chữ số.

Đơn vị cấp chính xác đo: 2 chữ số bên phải dấu phẩy thập phân

Tên miền giá trị (2): Xác suất – 5 chữ số có nghĩa

Mô tả miền giá trị: Toàn tập các số thực lớn hơn 0 nhỏ hơn 1 được biểu diễn với cấp chính xác 5 chữ số

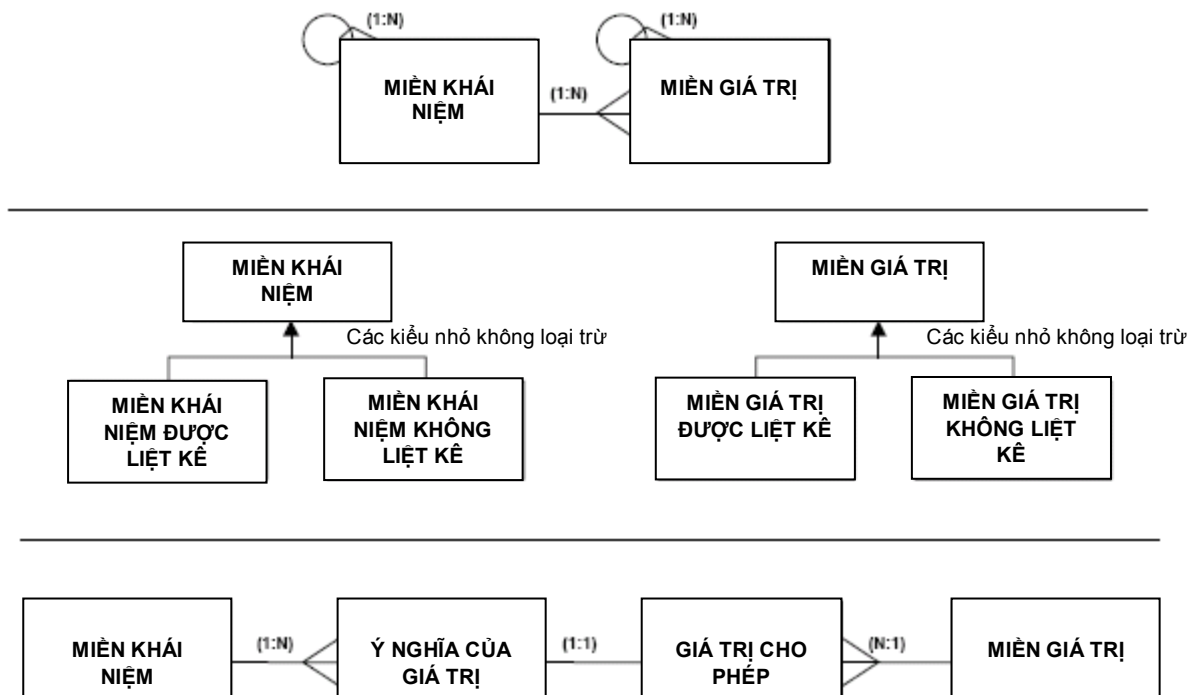
Đơn vị cấp chính xác đo: 5 chữ số bên phải dấu phẩy thập phân.

Ví dụ: 2 –Miền giá trị liệt kê tương đương

<p>Tên miền khái niệm: Các quốc gia trên thế giới Định nghĩa miền khái niệm: Danh sách các quốc gia trên thế giới hiện tại được biểu diễn theo tên hoặc mã.</p> <p>-----</p> <p>Tên miền giá trị (1): Mã quốc gia – chữ anpha 2 ký tự Các giá trị cho phép: <AF, Thực thể địa lý chính trị chính được biết là "Cộng hòa dân chủ Afghanistan"> <AL, Thực thể địa lý chính trị chính được biết là "Cộng hòa xã hội nhân dân Albania"> ... <ZW, Thực thể địa lý chính trị chính được biết là "Cộng hòa Zimbabwe"></p> <p>-----</p> <p>Tên miền giá trị (2): Mã quốc gia – chữ anpha 3 ký tự Các giá trị cho phép: <AFG, Thực thể địa lý chính trị chính được biết là "Cộng hòa dân chủ Afghanistan"> <ALB, Thực thể địa lý chính trị chính được biết là "Cộng hòa xã hội nhân dân Albania"> ... <ZWE, Thực thể địa lý chính trị chính được biết là "Cộng hòa Zimbabwe"></p>

Mọi miền giá trị biểu diễn hai loại khái niệm: khái niệm phần tử dữ liệu (gián tiếp) và miền khái niệm. Miền giá trị là cách biểu diễn đối với phần tử dữ liệu đó và, vì vậy, cũng trình bày gián tiếp khái niệm phần tử dữ liệu. Tuy nhiên, miền giá trị trực tiếp tương ứng với một miền khái niệm, nên trình bày khái niệm này độc lập với bất kỳ phần tử dữ liệu nào.

Xem ISO/IEC TR 20943-3, Information technology — Procedures for achieving metadata registry content consistency — Part 3: Value domain for detailed examples (*Công nghệ thông tin — Các thủ tục đạt nội dung nhất quán sổ đăng ký siêu dữ liệu — Phần 3: Miền giá trị cho các minh họa chi tiết*).



Hình này chỉ để tham khảo, không quy định.

Hình 3 — mô hình cơ sở đối với miền giá trị

6.5 Lược đồ phân loại cơ sở

Đối với mục đích của TCVN 7789, một lược đồ phân loại là một hệ thống khái niệm nhằm phân loại các đối tượng. Nó được tổ chức theo một số cấu trúc quy định, giới hạn nội dung bởi phạm vi và được mô phỏng cho việc ấn định các đối tượng cho các khái niệm được định nghĩa trong nó. Các khái niệm được ký hiệu cho một đối tượng và quá trình này được gọi là phân loại. Mỗi quan hệ liên kết các khái niệm trong hệ thống ký hiệu sẽ liên kết các đối tượng mà các khái niệm liên quan phân loại. Nói chung, bất kỳ hệ thống khái niệm nào cũng là một lược đồ phân loại nếu nó được sử dụng để phân loại đối tượng.

Phạm vi nội dung của lược đồ phân loại hạn chế phạm vi vấn đề của chủ thể được bao hàm bởi lược đồ phân loại đó. Phạm vi của lược đồ phân loại là khái niệm rộng nhất chứa trong hệ thống khái niệm của lược đồ đó. Theo lý thuyết, nó xác định đối tượng nào có thể được phân loại trong lược đồ đó, hoặc không.

Các hệ thống khái niệm và các lược đồ phân loại cụ thể có thể được cấu trúc theo nhiều cách. Cấu trúc xác định các kiểu mối quan hệ có thể tồn tại giữa các khái niệm và mỗi lược đồ phân loại có thể được sử dụng cho mục đích liên kết các khái niệm với đối tượng. Trong một lược đồ phân loại cụ thể, các khái niệm được liên kết cùng với các khái niệm khác liên quan với khái niệm được liên kết trong lược đồ đó cung cấp một khung cơ cấu khái niệm để hiểu ý nghĩa của đối tượng đó. Khung cơ cấu này được hạn chế bởi phạm vi của lược đồ phân loại.

Hệ thống khái niệm có thể được trình bày bởi một hệ thống thuật ngữ. Các hệ ký hiệu được sử dụng để trình bày một trong các khái niệm và được sử dụng như các từ khóa được liên kết đến các đối tượng để tìm kiếm, lập mối liên hệ, hoặc các mục đích khác.

Một kiểu hệ thống khái niệm đặc biệt là một hệ thống quan hệ. Ở đây, các khái niệm là kiểu quan hệ. Một kiểu quan hệ có N luận cứ và nó được gọi một kiểu quan hệ n -ary. Câu "một tập N đối tượng được phân loại bởi một kiểu quan hệ n -ary" có nghĩa là N đối tượng có một mối quan hệ giữa chúng và kiểu quan hệ đã cho.

7 Sổ đăng ký siêu dữ liệu

7.1 Khái quát

Siêu dữ liệu cũng là dữ liệu, vì vậy, siêu dữ liệu có thể được lưu trữ trong một cơ sở dữ liệu. Một cơ sở dữ liệu siêu dữ liệu để hỗ trợ chức năng đăng ký là một sổ đăng ký siêu dữ liệu (MDR). Một mô hình khái niệm MDR để mô tả dữ liệu được đưa ra trong TCVN 7789-3. Các yêu cầu và thủ tục cho các khía cạnh đăng ký trong TCVN 7789 được mô tả trong TCVN 7789-6. Đối với các sổ đăng ký siêu dữ liệu thực tế, có thể có các yêu cầu và thủ tục đăng ký bổ sung, nằm ngoài phạm vi của TCVN 7789. Các quy tắc và hướng dẫn đối với việc cung cấp các định nghĩa hoàn hảo và phát triển các quy ước đặt

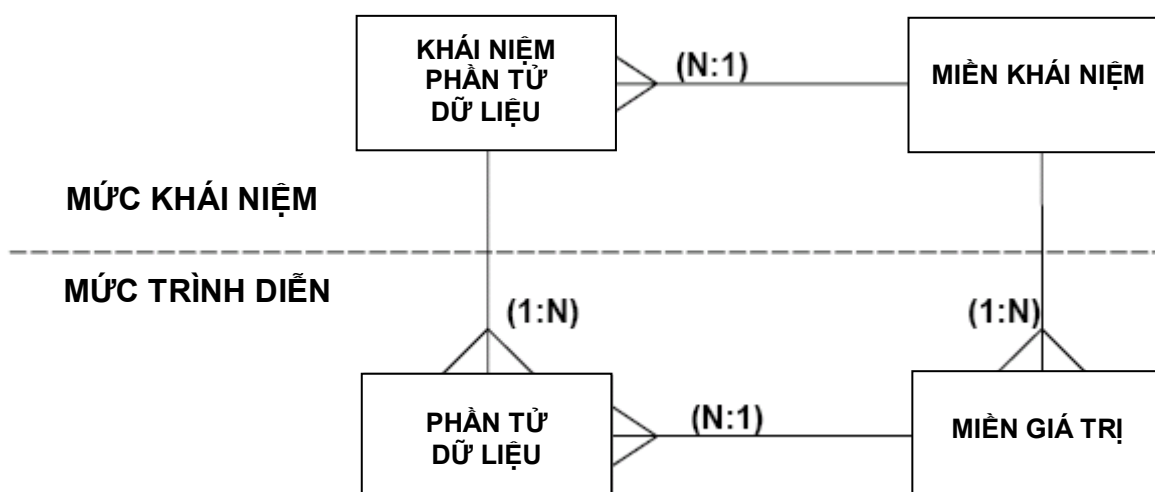
tên được mô tả tương ứng trong TCVN 7789- 4 và TCVN 7789 -5, tương ứng trong các tài liệu. Vai trò của việc phân loại được mô tả trong TCVN 7789-2. Các khuyến cáo và thực tiễn đăng ký các phần tử dữ liệu được mô tả trong ISO/IEC TR 20943-1. Các khuyến cáo và thực tiễn đăng ký miền giá trị được mô tả trong ISO/IEC TR 20943-3.

Một MDR gồm siêu dữ liệu mô tả các cấu trúc dữ liệu. Các thuộc tính nhằm mô tả một cấu trúc dữ liệu cụ thể (như là; các phần tử dữ liệu), được hiểu một cách thống nhất như một đối tượng siêu dữ liệu. Khi các thuộc tính được minh họa cụ thể cùng với mô tả một cấu trúc dữ liệu cụ thể thì chúng được hiểu như một mục siêu dữ liệu. Việc đăng ký mục siêu dữ liệu (nghĩa là; nhập siêu dữ liệu vào trong MDR) sẽ tạo ra một mục đăng ký. Nếu mục đăng ký đó cũng là đối tượng để quản trị (như trong trường hợp của một phần tử dữ liệu) thì nó được gọi là mục được quản trị.

CHÚ THÍCH - Nói chung, việc đăng ký một mục siêu dữ liệu mô tả một cấu trúc dữ liệu được biết như việc đăng ký cấu trúc dữ liệu đó. Trên thực tế, cấu trúc dữ liệu đó không được lưu trữ trong MDR, mà chỉ các mô tả của nó được lưu trữ. Điều này tương tự đối với các sổ đăng ký được chính phủ sử dụng để quản lý các phương tiện mô tô. Mô tả cho mỗi phương tiện mô tô được nhập vào trong sổ đăng ký, còn bản thân phương tiện thì không được nhập vào. Tuy nhiên, chúng ta đều nói rằng đã đăng ký các phương tiện mô tô, chứ không nói là đã đăng ký các mô tả đó.

7.2 Mô hình khái quát cho một MDR theo TCVN 7789

Mô hình khái niệm đối với một MDR theo TCVN 7789 bao gồm hai phần chính: mức khái niệm và mức biểu diễn (hoặc mức cú pháp). Mức khái niệm chứa các lớp khái niệm phần tử dữ liệu và miền khái niệm. Cả hai lớp đều biểu diễn các khái niệm. Mức trình diễn bao gồm các lớp phần tử dữ liệu và miền giá trị. Cả hai lớp đều biểu diễn các hàm chứa các giá trị dữ liệu. Điều 6 chứa các mô tả của mỗi lớp được trình bày lại trong Hình 4.



Hình 4 – Mô hình khái quát về sổ đăng ký siêu dữ liệu trong TCVN 7789 (ISO 11179)

Hình 4 trình bày bằng hình vẽ sự việc cơ sở về bốn lớp:

TCVN 7789-1 : 2007

- Một phần tử dữ liệu là một liên kết của khái niệm và biểu diễn phần tử dữ liệu (chủ yếu là miền giá trị);
- Nhiều phần tử dữ liệu có thể chia sẻ cùng một khái niệm phần tử dữ liệu, điều đó có nghĩa là một DEC có thể được trình bày lại theo nhiều cách khác nhau;
- Nhiều phần tử dữ liệu có thể chia sẻ cùng một sự biểu diễn, điều đó có nghĩa là một miền giá trị có thể được tái sử dụng trong các phần tử dữ liệu khác;
- Miền giá trị không bắt buộc phải liên quan với một phần tử dữ liệu và có thể được quản lý một cách độc lập;
- Miền giá trị mà việc chia sẻ toàn bộ ý nghĩa giá trị của các giá trị cho phép của chúng là tương đương về mặt khái niệm, vì vậy chúng chia sẻ cùng miền khái niệm
- Miền giá trị mà việc chia sẻ một số ý nghĩa giá trị của các giá trị cho phép của chúng có liên quan về mặt khái niệm, vì vậy chúng chia sẻ cùng một miền khái niệm trong hệ thống khái niệm chứa các miền khái niệm. Đó là các miền khái niệm chứa các miền khái niệm tương ứng;
- Nhiều miền giá trị có thể chia sẻ cùng một miền khái niệm;
- Khái niệm phần tử dữ liệu quan hệ với một miền khái niệm đơn, vì vậy mọi phần tử dữ liệu cùng chia sẻ một khái niệm phần tử dữ liệu, sẽ chia sẻ các cách biểu diễn có liên quan về bkhái niệm.

Nhiều sự việc khác không được minh họa trong Hình 4, nhưng một số được mô tả trong điều 6. hai sự việc được mô tả trong Hình 4 là:

- Các mối quan hệ giữa các khái niệm phần tử dữ liệu có thể được duy trì trong một MDR, điều đó hàm ý rằng một hệ thống khái niệm của các khái niệm phần tử dữ liệu có thể được duy trì;
- Các mối quan hệ giữa các miền khái niệm có thể được duy trì trong một MDR, điều đó hàm ý rằng một hệ thống khái niệm của các miền khái niệm có thể được duy trì.

Một số ấn phẩm cơ sở của việc đăng ký và quản trị siêu dữ liệu trong một MDR được mô tả ở phần sau trong Điều này.

7.3 Cơ sở đăng ký

Các vấn đề cơ sở của việc đăng ký và quản trị được quy định trong TCVN 7789-6 phân biệt một MDR với một cơ sở dữ liệu của siêu dữ liệu. Phương tiện để hoàn thành các chức năng này là một phần lớn của việc thiết kế siêu mô hình được quy định trong TCVN 7789-3.

Việc đăng ký tập các quy tắc, thao tác và thủ tục để áp dụng cho MDR. Đăng ký mô tả chi tiết như áp dụng trong TCVN 7789 được đề cập ở TCVN 7789-6. Ba kết quả quan trọng nhất của việc đăng ký là khả năng kiểm tra chất lượng của siêu dữ liệu, truy tìm nguồn gốc (nguồn siêu dữ liệu) và việc ấn định một định danh cho mỗi đối tượng được mô tả trong một MDR. Việc đăng ký cũng đòi hỏi một tập các thủ tục để quản lý một sổ đăng ký, đệ trình siêu dữ liệu để đăng ký các đối tượng và duy trì trách nhiệm quản lý đối tượng đối với siêu dữ liệu đã được đệ trình. Đối với sự triển khai thực tiễn của một sổ đăng ký siêu dữ liệu, có thể có các yêu cầu bổ sung, nằm ngoài phạm vi của TCVN 7789.

Mỗi mục được quản trị được duy trì theo một kiểu thống nhất và được mô tả. Các định danh, đo lường chất lượng, các tổ chức có trách nhiệm, tên và định nghĩa là toàn bộ các phần của siêu dữ liệu khái quát cần phải quản trị. Việc đăng ký là quá trình tạo ra hoặc duy trì siêu dữ liệu quản trị và siêu dữ liệu được chi tiết khác.

Chất lượng của siêu dữ liệu được giám sát thông qua việc sử dụng một trạng thái đăng ký. Các trạng thái này ghi lại mức chất lượng. Mỗi mức được quy định trong TCVN 7789-6. Mọi mục được quản trị được ấn định một trạng thái đăng ký và trạng thái này có thể thay đổi theo thời gian. Ngoài ra, chất lượng của siêu dữ liệu có nhiều khía cạnh. Đó là, có nhiều mục đích để kiểm tra chất lượng của siêu dữ liệu. Các mục đích chính là:

- Giám sát sự gắn kết với các quy tắc cho việc cung cấp siêu dữ liệu cho mỗi thuộc tính;
- Giám sát sự gắn kết của các quy ước cho việc hình thành các định nghĩa, việc tạo ra tên gọi và thực hiện phân loại;
- Xác định mục được quản trị vẫn còn liên quan hay ko;
- Xác định sự tương tự của các mục được quản trị liên quan và làm hài hòa các điểm khác nhau của chúng;
- Xác định liệu có thể có được siêu dữ liệu có chất lượng cao hơn đối với một số mục được quản trị;

Các quy tắc tạo ra và ấn định các định danh được mô tả trong TCVN 7789-6. Mỗi mục được quản trị trong một MDR được ấn định một định danh duy nhất.

Tổ chức đăng ký có thẩm quyền là tổ chức có trách nhiệm để đưa ra các thủ tục, quản trị và duy trì một MDR. Tổ chức đệ trình có trách nhiệm yêu cầu một mục siêu dữ liệu mới được đăng ký trong sổ đăng ký. Người quản lý có trách nhiệm đối với nội dung quản lý đối tượng của mỗi mục được đăng ký. Mỗi vai trò trong các vai trò này được mô tả trong TCVN 7789-6.

8 Khái quát về bộ tiêu chuẩn TCVN 7789

8.1 Giới thiệu các phần

Mục này giới thiệu từng phần của tập tiêu chuẩn TCVN 7789. Nó tóm lược các điểm chính và đề cập sự quan trọng của mỗi phần.

8.1.1 Phần 1

TCVN 7789-1, Khung cơ cấu, giới thiệu và đề cập các quan điểm cơ sở của các phần tử dữ liệu, miền giá trị, các khái niệm phần tử dữ liệu, các miền khái niệm và các lược đồ phân loại thiết yếu cho việc hiểu tập tiêu chuẩn này và cung cấp để liên kết với từng phần của tập tiêu chuẩn TCVN 7789.

8.1.2 Phần 2

TCVN 7789-2, Phân loại, cung cấp một mô hình khái niệm để quản lý các hệ thống khái niệm được sử dụng như các lược đồ phân loại. Các khái niệm từ các lược đồ này được liên kết với các mục được quản trị thông qua quá trình phân loại. Người quản lý thư viện, người làm về thuật ngữ chuyên môn, nhà ngôn ngữ học và các nhà khoa học máy tính đang hoàn thiện quá trình phân loại này, vì vậy không được mô tả ở đây. Nội dung ngữ nghĩa bổ sung được xuất phát từ việc phân loại là quan điểm quan trọng.

Việc liên kết một đối tượng với một hoặc nhiều khái niệm từ một hoặc nhiều lược đồ phân loại cung cấp

- Thông hiểu thêm về đối tượng đó;
- Thông tin so sánh với các đối tượng tương tự;
- Việc thông hiểu một đối tượng trong ngữ cảnh của một lĩnh vực vấn đề của chủ thể (được xác định bởi phạm vi của một lược đồ phân loại);
- Khả năng xác định sự khác nhau không đáng kể về ý nghĩa giữa các đối tượng tương tự.

Vì vậy, việc quản lý các lược đồ phân loại là một phần quan trọng của tối đa hóa tiềm năng thông tin trong một MDR. TCVN 7789-2 cung cấp khung cơ cấu cho điều này.

8.1.3 Phần 3

TCVN 7789-3, Siêu mô hình đăng ký và các thuộc tính cơ sở, quy định một mô hình khái niệm đối với một MDR. Nó hạn chế đối với một tập các thuộc tính cơ sở cho các phần tử dữ liệu, các khái niệm phần tử dữ liệu, miền giá trị, các miền khái niệm, các lược đồ phân loại và các lớp liên quan khác. Các

thuộc tính cơ sở được quy định đối với các phần tử dữ liệu trong TCVN 7789-3:1994 được đưa ra ở phiên bản này.

Siêu mô hình đăng ký được thể hiện theo ngôn ngữ lập mô hình thống nhất (Unified Modeling Language). Nó được chia thành sáu khoảng để dễ đọc. Mọi điều khoản được trình bày lại trong mô hình được mô tả dưới dạng văn bản. Một vài điều khoản được trình bày lại trong hộp chú thích trong lược đồ được mô tả trong văn bản.

Tiêu chuẩn này bao gồm một từ điển của toàn bộ các cấu trúc mô hình (các lớp, các thuộc tính và các mối quan hệ) được sử dụng trong mô hình này. Tập hợp các thuộc tính này được biết như "các thuộc tính cơ sở". Toàn bộ các thuộc tính được chứa trong các phần tiêu chuẩn 2, 4, 5 và 6 được chứa trong siêu mô hình đăng ký.

Siêu mô hình đăng ký không phải là một mô tả đầy đủ của tất cả siêu dữ liệu trong một tổ chức mong muốn ghi lại. Vì vậy, mô hình này được thiết kế để có thể mở rộng nếu cần. Tuy nhiên, Việc mở rộng theo bản chất tự nhiên của chúng, không là một phần của tiêu chuẩn này.

Mục này đưa ra các mô tả phù hợp. Sự tuân thủ được mô tả cho cả phù hợp nghiêm ngặt (toàn bộ các điều khoản đáp ứng) hoặc phù hợp (toàn bộ các điều khoản đáp ứng, trừ các điều khoản bổ sung nếu tồn tại).

8.1.4 Phần 4

TCVN 7789-4 (ISO/IEC 11179-4: Hệ thống định nghĩa dữ liệu), cung cấp hướng dẫn về cách thức xây dựng các định nghĩa dữ liệu một cách rõ ràng. Một số các quy tắc và hướng dẫn được trình bày trong TCVN 7789-4 để quy định chính xác cách thức một định nghĩa dữ liệu nên được tạo ra. Nói một cách chính xác, định nghĩa được tạo ra hoàn hảo là một trong các yêu cầu quyết định nhất đối với việc thông hiểu dữ liệu được chia sẻ; định nghĩa được tạo ra hoàn hảo là yêu cầu bắt buộc đối với việc trao đổi thông tin. Chỉ khi mọi người sử dụng có sự thông hiểu dữ liệu chung và chính xác mới có thể trao đổi dữ liệu đó không có trục trặc.

Tính hữu dụng của các định nghĩa là một khía cạnh trong chất lượng siêu dữ liệu. Các quy tắc và hướng dẫn sau được cung cấp trong phần 4 sẽ hỗ trợ thiết lập sự hữu ích này.

8.1.5 Phần 5

TCVN 7789-5, Nguyên tắc đặt tên và định danh, cung cấp hướng dẫn cho ký hiệu quy ước của các mục được quản trị. Ký hiệu quy ước là một thuật ngữ rộng để đặt tên hoặc định danh một cấu trúc dữ liệu cụ thể.

Các tên gọi được áp dụng cho các cấu trúc dữ liệu thông qua việc sử dụng một quy ước đặt tên. Quy ước đặt tên là các thuật toán để tạo ra các tên trong một ngữ cảnh cụ thể. Đó là ngữ nghĩa, cú pháp và

TCVN 7789-1 : 2007

các quy tắc từ vựng được sử dụng để tạo ra một quy ước đặt tên. Các tên gọi là một phương tiện đơn giản để cung cấp một số ngữ nghĩa về các cấu trúc dữ liệu, Tuy nhiên, các ngữ nghĩa này không đầy đủ. Các quy tắc cú pháp và các quy tắc về từ vựng nhằm vào các yếu tố cấu thành (như là; các ký tự cho phép), dạng thức và các xem xét khác.

Các cấu trúc dữ liệu có thể được gán nhiều tên gọi và một tên gọi có thể được định danh khi được ưu tiên. Thông thường, mỗi tên được ấn định để sử dụng trong ngữ cảnh mà nó được tạo ra.

Các thẻ định danh và hệ ký hiệu có ý nghĩa đối với việc tham chiếu lại các cấu trúc dữ liệu đối với việc quản lý và trao đổi siêu dữ liệu. Một RA ký hiệu một định danh duy nhất cho mỗi cấu trúc dữ liệu, duy nhất trong nội dung sổ đăng ký. Vì vậy, các định danh được ký hiệu cho các cấu trúc dữ liệu để sử dụng một cách rõ ràng các cấu trúc dữ liệu.

8.1.6 Phần 6

TCVN 7789-6, Việc đăng ký, cung cấp hướng dẫn sử dụng một ứng dụng đăng ký có thể đăng ký một cấu trúc dữ liệu với một RA và việc ấn định các định danh duy nhất cho mỗi cấu trúc dữ liệu. Duy trì các mục được quản trị đã được đăng ký cũng như được quy định trong tài liệu này. Việc đăng ký chủ yếu nhằm vào việc định danh, chất lượng và truy lại nguồn gốc của siêu dữ liệu trong một MDR.

Một thẻ định danh của mục được quản trị được hình thành bởi kết hợp của định danh RA, định danh duy nhất này được ấn định cho mục được quản trị một RA và phiên bản đó. Mỗi sổ đăng ký được duy trì bởi một RA mà các cấu trúc dữ liệu phụ thuộc về logic và chức năng. Ví dụ, các cấu trúc dữ liệu liên quan đến vấn đề hóa học gần như nên được đăng ký với tổ chức sản xuất hóa chất có thẩm quyền đăng ký.

Việc đăng ký là phức tạp hơn một chỉ dẫn đơn giản cho dù một mục siêu dữ liệu có được đăng ký hay không. Mặc dù nó gợi ý nên một đòi hỏi là chỉ có siêu dữ liệu "hoàn hảo" mới có thể được đăng ký, điều này là không thực tế. Vì vậy, việc cải thiện chất lượng của các mục được quản trị việc chia thành các mức khác nhau được gọi là trạng thái quản trị. Ngoài ra, Có các mức trạng thái đối với việc quản trị giữa mỗi mức chất lượng này. Nói tóm lại, các mức trạng thái này được gọi là trạng thái quản trị. Chúng chỉ ra một giai đoạn trong chu kỳ đăng ký để đạt được hiện tại đối với một mục được quản trị.

Nguồn gốc của siêu dữ liệu và chuỗi trách nhiệm được quản lý trong một MDR. Nhiệm vụ và vai trò của tổ chức đăng ký có thẩm quyền, người quản lý dữ liệu của tổ chức, tổ chức có trách nhiệm và tổ chức đệ trình được mô tả. Một khung cơ cấu cho một quá trình đăng ký được sử dụng trong một MDR được cung cấp.

Việc đăng ký vừa là quá trình vừa là mục tiêu. Việc ấn định một định danh, trạng thái chất lượng, trạng thái chu kỳ và mô tả nguồn gốc là các mục tiêu. Các quy tắc để các mục tiêu được hoàn thành là quá trình đó.

8.2 Các nguyên tắc cơ sở áp dụng TCVN 7789

Mỗi phần của TCVN 7789 trợ giúp trong một khía cạnh khác nhau của việc tạo, tổ chức và đăng ký siêu dữ liệu; mỗi phần nên được sử dụng chung với các phần khác. TCVN 7789-1 thiết lập các mối quan hệ giữa các phần và đưa ra các hướng dẫn sử dụng chúng một cách tổng thể. TCVN 7789-3 quy định các mục siêu dữ liệu một ứng dụng đăng ký phải cung cấp cho mỗi đối tượng được đăng ký. Các đặc điểm chi tiết của mỗi thuộc tính cơ sở là được cho trước. Do sự quan trọng của chúng trong việc quản trị siêu dữ liệu mô tả các cấu trúc dữ liệu, có ba thuộc tính (tên, định nghĩa và định danh) đưa ra các xử lý đặc biệt và mở rộng trong hai tài liệu. TCVN 7789-4 phải tiếp theo khi xây dựng cấu trúc các định nghĩa dữ liệu. Định danh và đặt tên phải tiếp theo các nguyên tắc nêu ra trong TCVN 7789-5. TCVN 7789-2 quy định một tập các thuộc tính sử dụng trong việc đăng ký và quản trị các lược đồ phân loại và các thành phần của chúng. Các mục siêu dữ liệu được đăng ký như các mục đăng ký và được quản trị như các mục được quản trị trong một MDR. TCVN 7789-6 cung cấp hướng dẫn về các thủ tục này.

9 Sự phù hợp

Không có chuẩn mực về sự phù hợp cụ thể nào đối với tiêu chuẩn này. Nó là một khung cơ cấu để nói các phần khác nhau của tập tiêu chuẩn TCVN 7789 với nhau. Như vậy, sự phù hợp không là một vấn đề của tiêu chuẩn này. Mỗi một phần của tiêu chuẩn TCVN 7789 có mục sự phù hợp của chính nó.
