

DOI:10.22144/ctu.jvn.2020.139

QUẢN LÝ DỮ LIỆU KHÔNG GIAN TRONG CÁC HỆ THỐNG THÔNG TIN NỀN WEB: CÁC VẤN ĐỀ PHÁT SINH VÀ GIẢI PHÁP CHUẨN HÓA

Lê Thành Phiêu^{1*}, Trương Xuân Việt², Ông Thị Mỹ Linh², Hồ Hưng Phát³, Biện Công Nhựt Trường², Võ Ngọc Giàu³, Phạm Thành Lê¹, Vũ Ánh Nguyệt¹, Trần Thị Phượng⁴, Phan Huy Phương⁴ và Nguyễn Hoàng Việt²

¹Phòng Quản trị Thiết bị, Trường Đại học Cần Thơ

²Trung tâm Công nghệ Phần mềm, Trường Đại học Cần Thơ

³Khoa Công nghệ thông tin và Truyền thông, Trường Đại học Cần Thơ

⁴Khoa Công nghệ, Trường Đại học Cần Thơ

*Người chịu trách nhiệm về bài viết: Lê Thành Phiêu (email: ltpheiu@ctu.edu.vn)

Thông tin chung:

Ngày nhận bài: 28/08/2020

Ngày nhận bài sửa: 08/10/2020

Ngày duyệt đăng: 28/12/2020

Title:

Spatial data management on web-based information systems: Issued problems and Standardization solutions

Từ khóa:

Cơ sở dữ liệu không gian, domain-based web service (DWS), hạ tầng dữ liệu không gian, hạ tầng không gian, thực thể học phân tán, web features service (WFS)

Keywords:

Distributed ontology, domain-based web service (DWS), spatial data infrastructure, spatial database, spatial infrastructure, web features service (WFS)

ABSTRACT

The need for monitoring the spatial infrastructure is becoming an essential part of the management systems of large organizations. Although possessing many different specialized information systems, most of them ignore the spatial factors. That leads to a lot of restrictions related to the management of distributed objects as infrastructure. The paper has successfully proposed a process and model for standardizing (spatial and non-spatial) infrastructure data based on the principle of distributed ontology. The proposed model, based on two groups of Web services: Domain-based Web Services (DWS) and Web Features Services (WFS), achieves the goal of standardizing two specific infrastructure data groups, non-spatial and spatial, ready to operate in a distributed environment for rendering Web-based interactive maps.

TÓM TẮT

Trong một hệ thống quản lý cho các tổ chức có quy mô lớn, nhu cầu theo dõi và quản lý hạ tầng không gian đang ngày trở thành nhu cầu thiết yếu. Mặc dù phần lớn các tổ chức này sở hữu rất nhiều hệ thống thông tin chuyên ngành khác nhau nhưng phần lớn đều bỏ qua các thông số không gian. Điều đó dẫn đến rất nhiều hạn chế liên quan đến việc quản trị các đối tượng phân tán như cơ sở hạ tầng. Bài viết đề xuất quy trình và mô hình chuẩn hóa dữ liệu hạ tầng (không gian và phi không gian) dựa trên nguyên lý thực thể học phân tán. Mô hình đề xuất, dựa trên hai nhóm dịch vụ Web: Domain-based Web Services (DWS) và Web Features Services (WFS), đã đạt được mục tiêu chuẩn hóa hai nhóm dữ liệu hạ tầng đặc thù là phi không gian và không gian, sẵn sàng vận hành trên môi trường phân tán để kết xuất các bản đồ tương tác nền Web.

Trích dẫn: Lê Thành Phiêu, Trương Xuân Việt, Ông Thị Mỹ Linh, Hồ Hưng Phát, Biện Công Nhựt Trường, Võ Ngọc Giàu, Phạm Thành Lê, Vũ Ánh Nguyệt, Trần Thị Phượng, Phan Huy Phương và Nguyễn Hoàng Việt, 2020. Quản lý dữ liệu không gian trong các hệ thống thông tin nền web: các vấn đề phát sinh và giải pháp chuẩn hóa. Tạp chí Khoa học Trường Đại học Cần Thơ. 56(6A): 9-21.

1 GIỚI THIỆU

Đối với các tổ chức có quy mô lớn, nhu cầu theo dõi và quản lý cơ sở hạ tầng không gian đang trở nên thiết yếu, do các đối tượng này gắn liền với hầu hết các hoạt động thường nhật. Thông thường, các nhà quản lý chỉ quan tâm đến việc tiếp cận và quản lý thông tin theo cơ chế lập kế hoạch, tức là theo nguyên tắc 3W (*where, when, who*). Tuy nhiên, yếu tố không gian (*where*) thường chỉ được quản lý ở mức mô tả chứ chưa quan tâm đến các thông số tọa độ địa lý, hình dạng, hình trạng, ... Chẳng hạn, “*giảng dạy tại phòng học 101/XH*” hay “*triển khai hội nghị tại Hội trường Rùa*” là cách mô tả về địa điểm của các đối tượng không gian, nhưng vị trí cụ thể các hoạt động “*giảng dạy*” và “*hội nghị*” không được chú trọng, nếu có chỉ tồn tại trên các sơ đồ tĩnh. Hay nói cách khác, nhiều ứng dụng chưa tận dụng sự hỗ trợ của các *hệ thống thông tin địa lý* (GIS – geographic information system) (Kresse and Danko, 2012), trong khi các nền tảng quản trị dữ liệu đang ngày càng hỗ trợ chúng một cách mạnh mẽ. Trong một số trường hợp, đôi khi bản đồ được đưa vào ứng dụng như một tiện ích, tuy nhiên chủ yếu dựa trên các dịch vụ có sẵn chứ chưa phải một công cụ quản trị thật sự, chẳng hạn tính năng hiển thị tọa độ địa lý của tổ chức trên bản đồ Google Map¹. Mặc dù phần lớn các tổ chức này sở hữu rất nhiều hệ thống thông tin chuyên ngành khác nhau nhưng hầu hết đều bỏ qua yếu tố vị trí địa lý, trong khi các tổ chức lớn đều có diện tích rộng và thậm chí được bố trí trên nhiều khu vực cách biệt. Điều đó dẫn đến các hạn chế sau trong quản lý hạ tầng không gian: (1) *thiếu khả năng quản lý và thống kê dữ liệu không gian, một thông số rất cần thiết* và (2) *thiếu khả năng đưa ra các phân tích nâng cao về không gian hoặc kết hợp giữa dữ liệu không gian và phi không gian*.

Quản trị cơ sở hạ tầng và quản lý quy hoạch là một trong các lĩnh vực cần nhất các thuộc tính không gian, trong đó liên quan đến vị trí địa lý của các đối tượng. Trong đó, các giải pháp phổ biến nhất là sử dụng các ứng dụng thuộc nhóm CAD (computer-aided design) hoặc nền tảng GIS truyền thống như MicroStation, MapInfo, ESRI, ... Sự đứt quãng trong tiếp cận quản trị dữ liệu chính là vấn đề nghiêm trọng nhất mà các nhà quản lý đang gặp phải. Chẳng hạn, tại Trường Đại học Cần Thơ, việc quản trị quy hoạch hạ tầng đã được quan tâm thực hiện từ năm 2009, theo Quyết định số 1258/QĐ-UBND ngày 22 tháng 4 năm 2009 của Ủy ban Nhân dân Thành phố

Cần Thơ về việc phê duyệt quy hoạch chi tiết xây dựng tỷ lệ 1/500 Trường Đại học Cần Thơ (Khu II) đến năm 2020. Tuy nhiên, dữ liệu này được quản lý dưới dạng AutoCAD nên công việc quản lý cũng gặp nhiều khó khăn, do các đối tượng chưa được phân loại, không có chức năng truy vấn từ câu lệnh, không cho phép truy cập phân tán, ... Do vậy, cần thiết phải xây dựng một hệ thống thông tin bản đồ cho phép kết xuất dữ liệu theo hình thức WebGIS. Đó là chưa kể rất nhiều dữ liệu quy hoạch khác phát sinh trong quản lý các dự án xây dựng, quản lý hạ tầng, ... cũng đòi hỏi phải phân tích để đưa vào tích hợp. Belussi *et al.* (2007) cũng hệ thống được một số khó khăn trong quản trị dữ liệu không gian trên nền web, trong đó có các vấn đề về sự khác biệt độ phân giải, tốc độ đường truyền và tính *bán cấu trúc* trong biểu diễn dữ liệu nền Web (với các ngôn ngữ HTML, GML, XML, ...).

Xét về khả năng tương thích với các hệ quản trị CSDL quan hệ, các nền tảng GIS truyền thống có ưu thế hơn so với nhóm CAD do có cùng cách thức quản trị dữ liệu dạng bảng dữ liệu (data table), hay cụ thể hơn, đây là dữ liệu có cấu trúc. Một trong số đó đã hỗ trợ cung cấp dữ liệu hướng dịch vụ, chẳng hạn ArcGIS Server² (ESRI), MapServer³, GeoServer⁴, ... được xây dựng cho mục đích này. Bên cạnh đó, các hệ quản trị CSDL đã sớm nhận ra sự cần thiết của việc quản trị dữ liệu không gian và bắt đầu cung cấp khả năng lưu trữ và truy vấn dữ liệu này từ những phiên bản cách nay hàng thập kỷ. Có thể nói, hầu hết các hệ quản trị CSDL phổ biến hiện nay đều đã cung cấp khả năng quản trị dữ liệu không gian, như Microsoft SQL Server, PostgreSQL, MySQL, ...

Bài viết này đề xuất *Quy trình và Mô hình chuẩn hóa dữ liệu hạ tầng* (không gian và phi không gian) dựa trên nguyên lý *thực thể học phân tán* (OMG, 2018). Đây là đóng góp mới và hữu hiệu cho lĩnh vực quản lý *hạ tầng dữ liệu không gian* (SDI – *spatial data infrastructure*) (Hu, 2017) cho các tổ chức có quy mô lớn, hoặc các bài toán quản lý trên các địa bàn rộng như quản lý đô thị, giao thông, ...

2 NGHIÊN CỨU LIÊN QUAN

2.1 Cơ sở dữ liệu không gian (spatial database)

Dữ liệu không gian (spatial data) (Kresse and Danko, 2012) còn có các tên gọi khác là *dữ liệu tham chiếu địa lý* (geographically referenced data) hay *dữ*

¹<http://maps.google.com/>

²<https://enterprise.arcgis.com/en/server/>

³<https://mapserver.org/>

⁴<http://geoserver.org/>

liệu không gian địa lý (geospatial data), được dùng để quản lý các đối tượng không gian với các đặc trưng địa lý (geographic feature) chủ yếu được biểu diễn ở dạng hình học hai chiều như điểm (point), đường (line), đa giác (polygon), đa điểm (multi-point), đa đường (multi-line),... Các đặc trưng này được chuẩn hóa bởi OGC như với tên gọi các Đặc trưng đơn giản (simple features hay simple feature access), và từ đó được hỗ trợ bởi hầu hết các hệ quản trị CSDL quan hệ và các nền tảng phân tích dữ liệu khác, chẳng hạn ngôn ngữ R (Pebesma, 2018). Điểm là loại đối tượng không có kích thước với đặc trưng là vị trí. Điểm có thể dùng để mô tả các đối tượng liên quan đến vị trí như cây xanh, trường học, tòa nhà, ... trong trường hợp không cần quan tâm đến các đối tượng bên trong của các đối tượng này. Đường là loại đối tượng một chiều, có các đặc trưng như vị trí, chiều dài,... Đường được sử dụng để mô tả các đối tượng như sông, suối, đường đi, đường điện, đường ống, ... Vùng là loại đối tượng hai chiều với đặc trưng là vị trí, chiều dài, chiều rộng, diện tích,... Vùng thường được dùng để mô tả các đối tượng như tòa nhà, phòng học, khu vực,...

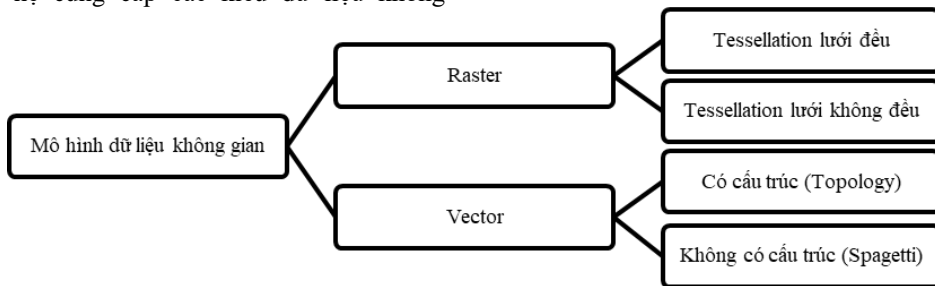
Cơ sở dữ liệu không gian (spatial database) (Kresse and Danko, 2012) là một hệ thống cơ sở dữ liệu quan hệ cung cấp các kiểu dữ liệu không

gian trong mô hình dữ liệu và các ngôn ngữ truy vấn. Chỉ mục được sử dụng trong cơ sở dữ liệu không gian để tối ưu hóa câu lệnh truy vấn và tăng tốc độ truy vấn dữ liệu đồng thời giảm bộ nhớ lưu trữ.

2.2 Các mô hình dữ liệu không gian (spatial data model)

Dữ liệu không gian (spatial data) được tổ chức theo hai mô hình chính là mô hình Raster và mô hình Vector. Các mô hình này gọi là mô hình dữ liệu không gian (Saylor Academy, 2012).

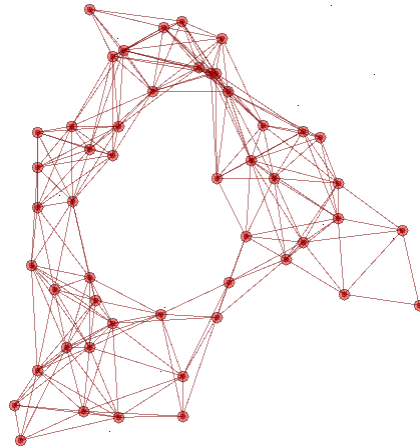
Mô hình dữ liệu Raster biểu diễn các đặc trưng địa lý bằng các điểm ảnh (cell/pixel) được sắp xếp theo hàng và cột, trong đó mỗi điểm ảnh sẽ chứa giá trị đại diện cho dữ liệu. Raster thường được sử dụng để lưu trữ hình ảnh chụp từ không trung, hình ảnh từ vệ tinh, hình ảnh kỹ thuật số, ảnh được chụp hoặc quét,... Đây là một mô hình lưu trữ dữ liệu có cấu trúc đơn giản, phù hợp với phân tích, thống kê không gian; có khả năng biểu diễn các bề mặt liên tục và thực hiện phân tích bề mặt; có khả năng lưu trữ đồng nhất các điểm, đường thẳng, đa giác và bề mặt; có khả năng thực hiện các lớp phủ nhanh chóng với các bộ dữ liệu phức tạp,... Trong phạm vi của nghiên cứu này, mô hình dữ liệu này không được đề cập.



Hình 1: Phân loại các mô hình dữ liệu không gian (Saylor Academy, 2012)

Ở một cách thức khác, mô hình Vector có cách thức biểu diễn các đặc trưng địa lý dựa trên việc ghi nhận các vị trí không gian đặc thù, sau đó vẽ lại các đối tượng dựa trên các giá trị được lưu trữ. Trong mô hình này, các đối tượng được cấu trúc chính từ các thành phần cơ sở như điểm, đoạn thẳng và vùng/đa giác (gồm tập hợp nhiều đoạn thẳng khép kín) trên một hệ thống tọa độ nào đó. Trên mặt phẳng, mỗi điểm được biểu diễn với một cặp tọa độ

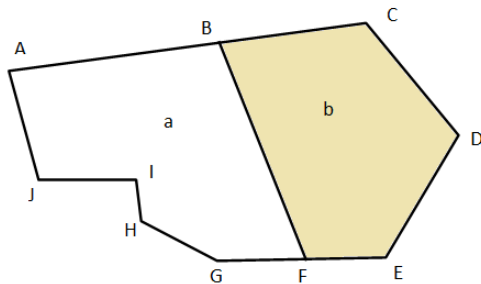
(x, y), mỗi đường được xác định là một chuỗi các điểm liên tiếp, được kết thúc bằng các nút, đoạn thẳng/cạnh được xác định bằng sự nối nhau của hai điểm, vùng được xác định bởi các đường khép kín. Mô hình Vector không tốn bộ nhớ như mô hình Raster nên rất hữu hiệu trong việc lưu trữ. Ở đây, đoạn thẳng và điểm được sử dụng để nhận biết vị trí của thế giới thực. Mô hình dữ liệu Vector được chia thành hai loại: (1) Mô hình phi cấu trúc (spagetti) và (2) Mô hình có cấu trúc (topology).



Hình 2: Ví dụ về mô hình Vector có cấu trúc Topology – Mô hình đồ thị đĩa (diskgraph) cho mạng lưới bãi đèn Đồng bằng Sông Cửu Long (Truong et al., 2011)

Mô hình Vector có cấu trúc (topology) tập trung xây dựng mối quan hệ giữa các thành phần không gian với hai thành phần là *Cung* (Arc) và *Nút* (Node). Mỗi cung được xác định bởi hai nút, các

phần tử giữa hai nút là các điểm điều khiển (Vertex), các điểm này xác định hình dạng của cung. Các cung giao nhau tại một nút. Một cung được kết thúc bằng một nút. Vùng là tập hợp các cung khép kín.



Đặc trưng	Vị trí
Điểm A	(x_A, y_A)
Điểm B	(x_B, y_B)
...	...
Đường AB	$((x_A, y_A), (x_B, y_B))$
Đường AJ	$((x_A, y_A), (x_J, y_J))$
...	...
Vùng a	$((x_A, y_A), (x_B, y_B), (x_F, y_F), (x_G, y_G), (x_H, y_H), (x_I, y_I), (x_J, y_J), (x_A, y_A))$
...	...

Hình 3: Mô tả dữ liệu với mô hình Spagetti

Mô hình phi cấu trúc (Spagetti) mô tả các đối tượng một cách độc lập với các đối tượng khác, các quan hệ hình học logic phải được tính toán dựa trên lệnh. Mỗi đối tượng điểm được xác định bằng một cặp tọa độ (x, y) , mỗi đối tượng đường được xác định bằng một chuỗi các cặp tọa độ (x_i, y_i) , mỗi đối tượng vùng được xác định bằng một chuỗi các cặp tọa độ (x_i, y_i) với điểm đầu và điểm cuối trùng nhau (Hình 3). Đây là một phương pháp lưu trữ dữ liệu đơn giản, việc thêm đối tượng vào cơ sở dữ liệu đã tồn tại được thực hiện dễ dàng, có lợi thế khi biểu diễn một tập các đối tượng không đồng nhất. Điểm hạn chế của phương pháp này là sự dư thừa dữ liệu và việc không ghi nhận đặc trưng kề nhau của hai vùng kề nhau (cạnh chung của hai vùng liền kề được biểu diễn là hai cạnh độc lập).

2.3 Thực thể học (ontology) và Thực thể học phân tán (distributed ontology)

Thực thể học (ontology) (Gruber, 1993) là một đặc tả hình thức, rõ ràng của một nhận thức chung với bốn khái niệm chính: mô tả trừu tượng của hiện tượng (nhận thức), diễn đạt rõ ràng bằng toán học (hình thức), các khái niệm và quan hệ giữa chúng phải được định nghĩa một cách chính xác và rõ ràng, tồn tại một **sự đồng thuận** của những người sử dụng *ontology*. Đó là một mô hình dữ liệu được sử dụng để suy luận về các đối tượng và mối quan hệ của chúng trong một lĩnh vực nào đó. Các thành phần trong một ontology bao gồm các *thực thể* (individual), *lớp* (class), *thuộc tính* (property) và *quan hệ* (relation) (Guarino et al., 2009). Thực thể là thành phần cơ bản của ontology, bao gồm các đối tượng cụ thể như con người, động vật, đồ vật,... hoặc các đối tượng trừu tượng như tài khoản, môn học,...

Vai trò quan trọng nhất của ontology là dùng để chia sẻ những hiểu biết chung về các khái niệm, cấu trúc thông tin giữa con người hoặc giữa các hệ thống phần mềm. Vai trò thứ hai là *ontology* được sử dụng để tái sử dụng tri thức, cho phép tri thức trở nên nhất quán và tường minh, độc lập với ngôn ngữ. Ngoài ra, đây cũng là phương tiện để thực hiện việc mô hình hóa và suy luận trong rất nhiều bài toán khác nhau. Narula *et al.* (2018) đã trình bày các phân tích và so sánh các công cụ sử dụng bản thể học trong việc tích hợp (nối kết) các đối tượng trong web ngữ nghĩa. Các công cụ được xem xét bao gồm ChimaeraKSL, PROMPT, FCA, MOMIS, GLUE, LILY, ASMOV và các khía cạnh được phân tích bao gồm: *mô tả bản thể học* (ontology representation), *thông tin ảnh xạ* (định nghĩa khái niệm, đô thị, cấu trúc,...), *các thành tố và bản chất các ảnh xạ*.

Trong các hệ thống lớn và phức tạp, các thông tin liên quan đến một thực thể không tập trung tại một máy chủ địa lý mà phân tán rải rác trên nhiều hệ thống độc lập (và có mức độ bảo mật nhất định). Khi đó, khái niệm thực thể học cần được hiểu rộng hơn ở phạm vi *Thực thể học phân tán* (distributed ontology). OMG (2018) cũng đã đề xuất đặc tả cho khái niệm *Thực thể học phân tán* và đây được xem là cơ sở lý thuyết quan trọng dùng để áp dụng cho việc quản lý các thuộc tính hạ tầng không gian trong nghiên cứu này. Sự đồng thuận trong hệ thống phân tán dựa trên chế dịch vụ Web (web service), bao gồm cả các nguyên tắc bảo mật.

3 CÁC VẤN ĐỀ PHÁT SINH TRONG QUẢN LÝ DỮ LIỆU KHÔNG GIAN

Quản lý dữ liệu không gian thường bị bỏ qua trong các hệ thống quản lý không phải vì sự thiếu cần thiết của chúng mà chủ yếu là do tính phức tạp cả về tổ chức dữ liệu lẫn các yếu tố kỹ thuật. Một số vấn đề liên quan đến quản trị dữ liệu không gian trên nền web được phân tích trong Belussi *et al.* (2007) cũng được xem xét thêm các khía cạnh khác và khái quát hóa thành ba nhóm vấn đề trọng tâm sau:

3.1 Dữ liệu rời rạc, không thống nhất

Dữ liệu trong quá trình thu thập thường rời rạc, không thống nhất do thu thập từ nhiều nguồn khác nhau. Mỗi bộ phận/phòng ban/tổ chức trong thực tế sử dụng một nguồn dữ liệu phục vụ riêng cho công việc của họ. Điều này dẫn đến những khó khăn rất lớn khi chúng ta muốn hệ thống hóa toàn bộ các dữ liệu vì bản chất của chúng là rời rạc và không thống nhất. Như vậy, một giai đoạn rất quan trọng trong quá trình chuẩn hóa dữ liệu là hệ thống hóa các dữ liệu từ nhiều nguồn dữ liệu rời rạc nhau.

Sự không đồng bộ xuất phát từ các nguyên nhân chính sau đây: (1) có nhiều đơn vị quản lý khác nhau nên cấu trúc dữ liệu không đồng nhất (cả dữ liệu không gian và phi không gian), (2) ở phạm vi tổng thể, do quá trình quản lý trải qua thời gian lâu dài và mỗi giai đoạn có đặc thù và mục tiêu khác nhau, thậm chí được quản lý bởi nhiều dự án khác nhau nên cách thức lưu trữ (Microsoft Excel, Microsoft Word, AutoCAD, cơ sở dữ liệu quan hệ,...) và các thuộc tính dữ liệu là khác nhau và (3) sự thiếu hụt hoặc chồng chéo dữ liệu đòi hỏi việc đầu tư kiểm tra và bổ sung dữ liệu.

Với các dữ liệu về cơ sở hạ tầng của Trường Đại học Cần Thơ đã thu thập được, mỗi loại thông tin được lưu trữ với một định dạng riêng. Ví dụ như dữ liệu về các phòng thí nghiệm, phòng thực hành, nhà học đa năng được lưu trữ với định dạng bảng tính Microsoft Excel, dữ liệu về nhà học được lưu trữ dưới định dạng AutoCAD, dữ liệu về các trại thực nghiệm ngoài trời được mô tả trong tập tin văn bản Microsoft Word,... Tương tự như vậy, rất nhiều dữ liệu được lưu trữ và quản lý trên các hệ quản trị cơ sở dữ liệu quan hệ (phi không gian), chẳng hạn trong các phần mềm quản lý tài sản, phần mềm quản lý phòng học, phần mềm quản lý hoạt động ngoại khóa,...

3.2 Không thống nhất định dạng dữ liệu

Trong phạm vi bài viết, khái niệm **định dạng dữ liệu** được hiểu là cấu trúc dữ liệu dùng để biểu diễn một đối tượng không gian cần quản lý, chẳng hạn cấu trúc dữ liệu cho phòng học, đường giao thông, các địa điểm sinh hoạt ngoại khóa. Như đã phân tích trong Phần 3.1 về vấn đề rời rạc và thiếu nhất quán của dữ liệu, sự không thống nhất về định dạng như một hệ quả tất yếu.

Việc tìm kiếm sự thống nhất về định dạng dữ liệu là một nhu cầu bắt buộc đối với việc quản lý dữ liệu không gian trong các hệ thống thông tin nền Web. Lý do chủ yếu là *điều kiện tối thiểu* để xây dựng một hệ thống bản đồ tương tác là cần một *CSDL quan hệ không gian phân tán* (distributed spatial RDBMS). Nếu không, các hệ thống thông tin bản đồ web không thể truy xuất hay xử lý các nguồn dữ liệu trên. Điều này đặt ra vấn đề cần chuẩn hóa dữ liệu về đúng định dạng mà các loại hệ thống thông tin bản đồ web yêu cầu. Các *điều kiện mở rộng* được phân tích để xây dựng một hệ thống bản đồ tương tác ở **Phần 3.3**.

Đối với dữ liệu không gian như cơ sở hạ tầng, việc không thống nhất định dạng dữ liệu thể hiện ở hai đặc điểm sau đây:

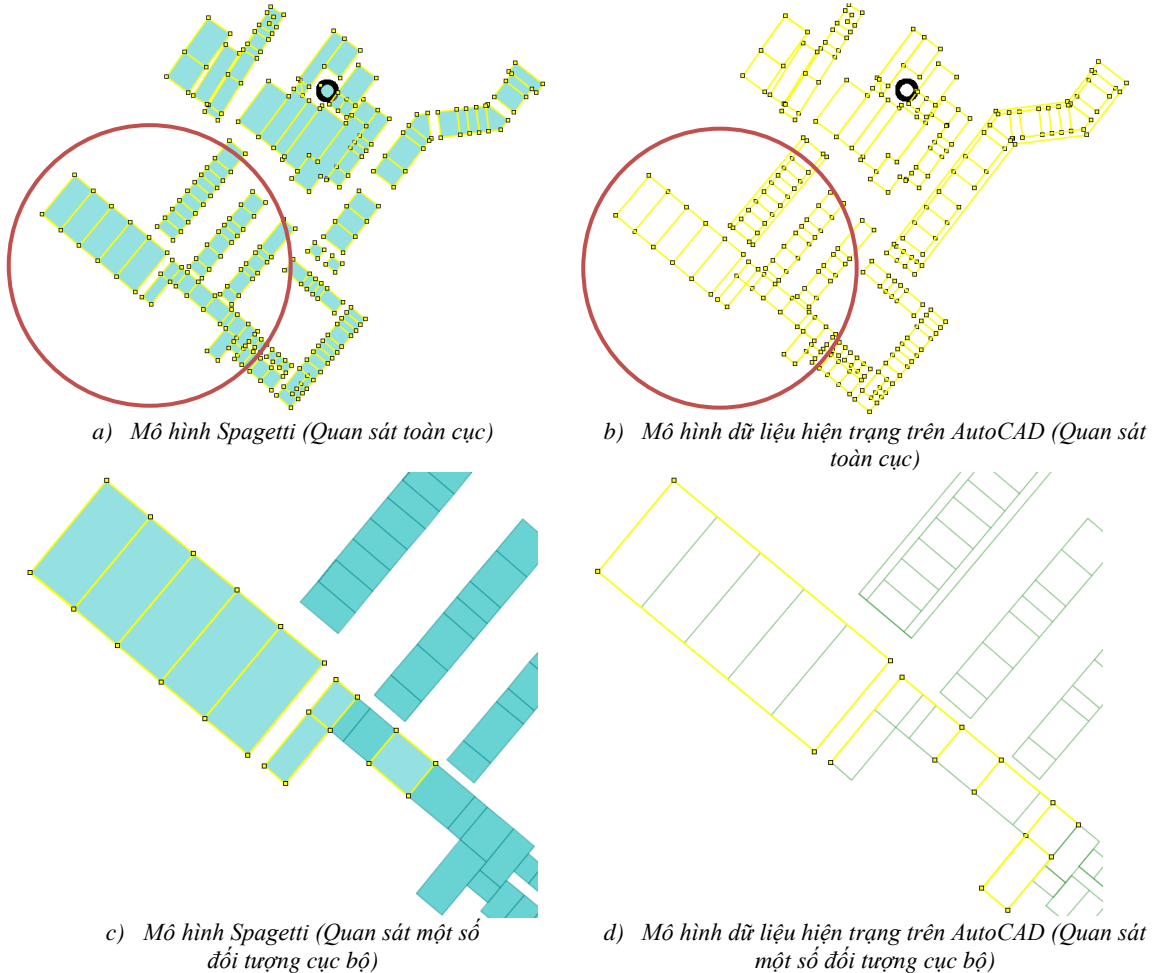
Không thống nhất mô hình dữ liệu phi không gian: Dữ liệu chưa được phân tích một cách khoa học theo quy trình phân tích hệ thống, theo đó chưa có một sơ đồ quan hệ thực thể (chẳng hạn, sơ đồ E-R) hoặc một mô hình tương tự để quản lý dữ liệu. Hay nói các khác, dữ liệu phi không gian chưa tuân thủ các chuẩn dữ liệu căn bản của một hệ thống thông tin (Guarino *et al.*, 2009).

Không thống nhất mô hình dữ liệu không gian: Đặc điểm này thể hiện ở việc các đối tượng không gian cần quản lý chưa được đối tượng hóa, hay thực thể hóa. Khi đó, dữ liệu nhìn ở mặt tổng thể thì có vẻ là hoàn chỉnh về mặt cấu trúc, nhưng khi quan sát ở cấp độ vi mô (từng đối tượng) sẽ gặp nhiều vấn đề phát sinh về quản lý.

Trong Hình 4, nếu quan sát toàn cục, ta khó có thể phân biệt được sự khác nhau trong việc tổ chức dữ liệu giữa Mô hình Spagetti (Hình 4a) và Mô hình

dữ liệu hiện trạng trên AutoCAD (Hình 4b). Nếu quan sát chi tiết hơn trong vùng vòng tròn đỏ của Hình 4a và Hình 4b, tương ứng là Hình 4c và Hình 4d của hai mô hình dữ liệu, ta thấy trong Hình 4c, các phòng học đều được tổ chức thành các POLYGON, trong khi ở Hình 4d, các đối tượng không gian không được rõ ràng, chủ yếu là sự nối kết giữa các POLYGON và LINE không theo một nguyên tắc nhất định nào cả. Hay nói các khác, các đối tượng cụ thể như phòng làm việc, sân chơi,... chưa được chú trọng phân hoạch. Qua đó, không thể giữ nguyên trạng dữ liệu có sẵn trên AutoCAD để chuyển vào các hệ thống quản trị dữ liệu, mà phải thực hiện một quá trình chuẩn hóa công phu (phần lớn là thủ công) để đạt được dữ liệu theo mô hình Spagetti.

Một trường hợp cụ thể về sự thiếu nhất quán trong quản lý dữ liệu trên CAD:



Hình 4: Quan sát cách thức tổ chức dữ liệu Khoa Thủy sản, Đại học Cần Thơ trên Mô hình Spagetti (đã chuẩn hóa) và Mô hình dữ liệu hiện trạng trên AutoCAD (chưa chuẩn hóa)

Qua ví dụ trên, có thể thấy các hạn chế về không thống nhất định dạng dữ liệu được bộc lộ rõ trong bản thiết kế tổng thể cơ sở hạ tầng Khu II Đại học Cần Thơ. Hạn chế này là chấp nhận được đối với các thiết kế xây dựng, trong đó mục đích phục vụ là *chỉ đọc* (nghĩa là kết xuất đầu cuối), nhưng việc này trở thành một trở ngại rất lớn đối với một hệ thống thông tin tương tác, khi mục đích truy vấn, cập nhật, nối kết là các chức năng mặc định. Tuy nhiên, hạn chế lớn nhất của các hệ thống CAD là dữ liệu chủ yếu phục vụ mục đích quan sát, không có cơ chế tương tác giống SQL trong các hệ quản trị cơ sở dữ liệu chuyên dụng. Hơn nữa, CAD cũng gặp một hạn chế lớn hơn là dữ liệu không được quản lý theo bất cứ một cấu trúc chuẩn nào, vào cũng không thấy có một định nghĩa nào về chuẩn cấu trúc cho quản lý dữ liệu CAD. Một khi dữ liệu không có cấu trúc, và tiếp theo không có khả năng phân tán, tương tác thì đây chính là các dữ liệu tĩnh. Nói cách khác, so với kiến trúc 3 tầng (3-tiers) thì dữ liệu CAD chỉ đáp ứng được tầng trình bày (presentation), không đảm bảo được chức năng của tầng cơ sở dữ liệu (database) và tầng xử lý nghiệp vụ (business logic).

Như vậy, do nhiều ưu điểm về các thức tổ chức dữ liệu và sự hỗ trợ của các công nghệ GIS, trong nghiên cứu này chúng tôi hướng đến việc sử dụng *Mô hình phi cấu trúc (spagetti)* trong quản trị không gian các đối tượng cơ sở hạ tầng. Trong đó, các đối tượng sẽ được lưu trữ dưới các kiểu dữ liệu Geometry đặc thù: POINT, LINE, POLYGON, hoặc các kiểu dữ liệu phức: MULTIPOINTS, MULTILINES, MULTIPOLYGON.

3.3 Vấn đề thực thể học phân tán (distributed ontology (OMG, 2018))

Như đã trình bày trong Phần 3.2, việc giải quyết tốt vấn đề định dạng dữ liệu (cả không gian và phi không gian) chỉ đóng vai trò thỏa mãn điều kiện tối thiểu đối với một hệ thống thông tin nền Web. Trong trường hợp này, tất cả dữ liệu không gian và phi không gian cần được tổ chức trong cùng một **hệ quản trị cơ sở dữ liệu duy nhất**, trên thực tế, điều này trở nên bất khả thi đối với các hệ thống thông tin lớn và phức tạp. Cụ thể hơn đối với trường hợp Đại học Cần Thơ, phần lớn dữ liệu phi không gian đã được phân tích và quản lý riêng trên nhiều hệ thống thông tin khác nhau. Do vậy, khi xây dựng một hệ thống bản đồ hạ tầng không gian cần phải xem xét kỹ các cách tiếp cận: (1) chuyển tất cả dữ liệu phi không gian về cùng hệ thống không gian để quản lý tập trung và (2) phải xây dựng được một hệ

thống cho phép nối kết chặt chẽ giữa các hệ thống dữ liệu phi không gian bên ngoài vào hệ thống dữ liệu không gian vừa xây dựng.

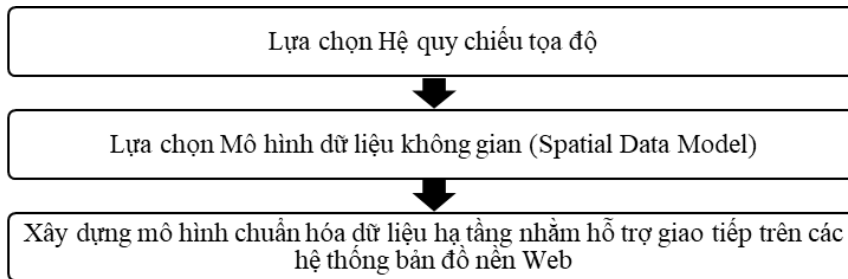
Cách tiếp cận thứ nhất là bất khả thi do liên quan đến vấn đề bảo mật và quyền hạn truy cập các nguồn dữ liệu phi không gian. Bên cạnh đó, dữ liệu phi không gian đang tồn tại ở trạng thái động và được cập nhật thường xuyên ở một hệ thống khác, trong khi hệ thống quản trị cơ sở hạ tầng không có được các chức năng này. Do vậy, việc chuyển dữ liệu về một hệ thống không duy trì được tính ổn định tổng thể.

Do đó, cách tiếp cận thứ hai được lựa chọn. Ở cách tiếp cận này, một sơ đồ thực thể dạng E-R được cài đặt trên một hệ quản trị CSDL là không đủ để giải quyết bài toán, mà phải mở rộng sơ đồ này ra ở một hệ thống lớn hơn, bao gồm nhiều hệ thống thông tin nối kết. Như vậy, nguồn dữ liệu không gian (và một phần phi không gian), sau khi thống nhất về định dạng, cần được xem xét dưới góc độ đối tượng để dễ dàng nối kết với các hệ thống khác. Đây được xem là vấn đề *thực thể học phân tán (distributed ontology)* (OMG, 2018). Đối với một tổ chức lớn và có nhiều hệ thống thông tin phức tạp, việc đối mặt vấn đề thực thể học phân tán là tất yếu, và cách thức giải quyết thông thường dựa trên nguyên lý dịch vụ (client/server), trong đó một Hệ thống thông tin sẽ cung cấp các dịch vụ thông tin liên quan đến một *đối tượng (object)* cho một hệ thống khác dựa trên một yêu cầu cụ thể. Trong bài toán quản lý cơ sở hạ tầng nền web, hệ thống thông tin bản đồ sẽ chỉ tập trung quản lý các thuộc tính không gian của các đối tượng và một số thuộc tính phi không gian đặc thù, phần còn lại sẽ được cung cấp bởi các hệ thống thông tin khác đang vận hành.

Đối với bài toán của Đại học Cần Thơ, nhu cầu theo dõi hoạt động ngoại khóa ngoài trời và trong phòng học, cũng như theo dõi thông tin về cơ sở hạ tầng bên trong một phòng học là một thách thức dạng thực thể học phân tán.

4 ĐỀ XUẤT MÔ HÌNH CHUẨN HÓA DỮ LIỆU KHÔNG GIAN PHỤC VỤ CÁC HỆ THỐNG THÔNG TIN NỀN WEB

Như phân tích ở trên, các dữ liệu không gian thu thập được từ thực tế hầu như chưa được chuẩn hóa, không thể đưa vào sử dụng trong các hệ thống bản đồ web. Trong bài viết này, quy trình chuẩn hóa dữ liệu không gian được đề xuất như sau:



Hình 5: Quy trình chuẩn hóa dữ liệu không gian

4.1 Lựa chọn hệ quy chiếu tọa độ

Hệ tọa độ địa lý (geographic coordinate system – GCS) là một hệ tọa độ cho phép tất cả mọi điểm trên Trái Đất đều có thể xác định được bằng một tập hợp các số có thể kèm ký hiệu. Hệ tọa độ phổ biến hiện dùng là hệ tọa độ cầu tương ứng với tâm Trái Đất với các tọa độ là vĩ độ, kinh độ và cao độ. Hệ quy chiếu là một hệ tọa độ, dựa vào đó vị trí của mọi điểm trên các vật thể và vị trí của các vật thể khác được xác định.

Hệ quy chiếu và hệ tọa độ quốc gia VN-2000 được áp dụng thống nhất trên toàn quốc để xây dựng hệ thống tọa độ các cấp hạng, hệ thống bản đồ địa hình cơ bản, hệ thống bản đồ nền, hệ thống bản đồ địa chính, hệ thống bản đồ hành chính quốc gia và các loại bản đồ chuyên đề khác theo Thông tư số 973/2001/TT-TCĐC về việc *Hướng dẫn áp dụng hệ quy chiếu và hệ tọa độ quốc gia VN-2000* ngày 20 tháng 6 năm 2001. VN-2000 còn có mã hiệu quốc tế là *EPSG:3405*⁵.

*EPSG:3857*⁶ (hay *WGS 84/Pseudo-Mercator*) là hệ thống tọa độ thống nhất trên phạm vi toàn cầu và được sử dụng phổ biến bởi các dịch vụ web như Google Maps và Open Street Map⁷. Do vậy, để nối kết với các bản đồ web từ lấy từ Google Maps hay Open Street Map thì các dữ liệu không gian của hệ thống cục bộ cũng cần được quy đổi về *EPSG:3857*.

4.2 Lựa chọn Mô hình dữ liệu không gian (Spatial Data Model)

Mỗi mô hình dữ liệu không gian (được trình bày trong Phần 2.2) đều có những ưu điểm và hạn chế khi sử dụng. Việc lựa chọn mô hình nào phụ thuộc hoàn toàn vào loại dữ liệu không gian cần lưu trữ, nhu cầu thao tác/xử lý dữ liệu sau khi lưu trữ và kích thước bộ nhớ.

Trở lại với việc xây dựng bản đồ cơ sở hạ tầng nền Web của Trường Đại học Cần Thơ, với nhu cầu

quản trị dữ liệu không gian trong quản trị điều hành, cụ thể là quản lý các đối tượng như tòa nhà, phòng học, đường điện, khu vực sinh hoạt ngoại khóa,... Các đối tượng này, trong thực tế, hầu như khác biệt nhau về hình dạng, kích thước, thuộc tính, nhu cầu sử dụng,... Trong quản lý, sự thay đổi liên quan đến các đối tượng trên xảy ra thường xuyên theo nhiều lý do, nên nhu cầu cập nhật lại thông tin cho các đối tượng này là tương đối lớn. Với các hiện trạng và nhu cầu trước mắt, Mô hình Spagetti được xem là phù hợp nhất để lưu trữ dữ liệu không gian về cơ sở hạ tầng của trường. Mô hình này rất tiện lợi do có cùng đối tượng quản lý với các HTTP phi không gian khác đang tồn tại, hay nói cách khác là có sự đồng nhất về thực thể học với các hệ thống khác.

4.3 Xây dựng mô hình chuẩn hóa dữ liệu hạ tầng (Không gian và Phi không gian)

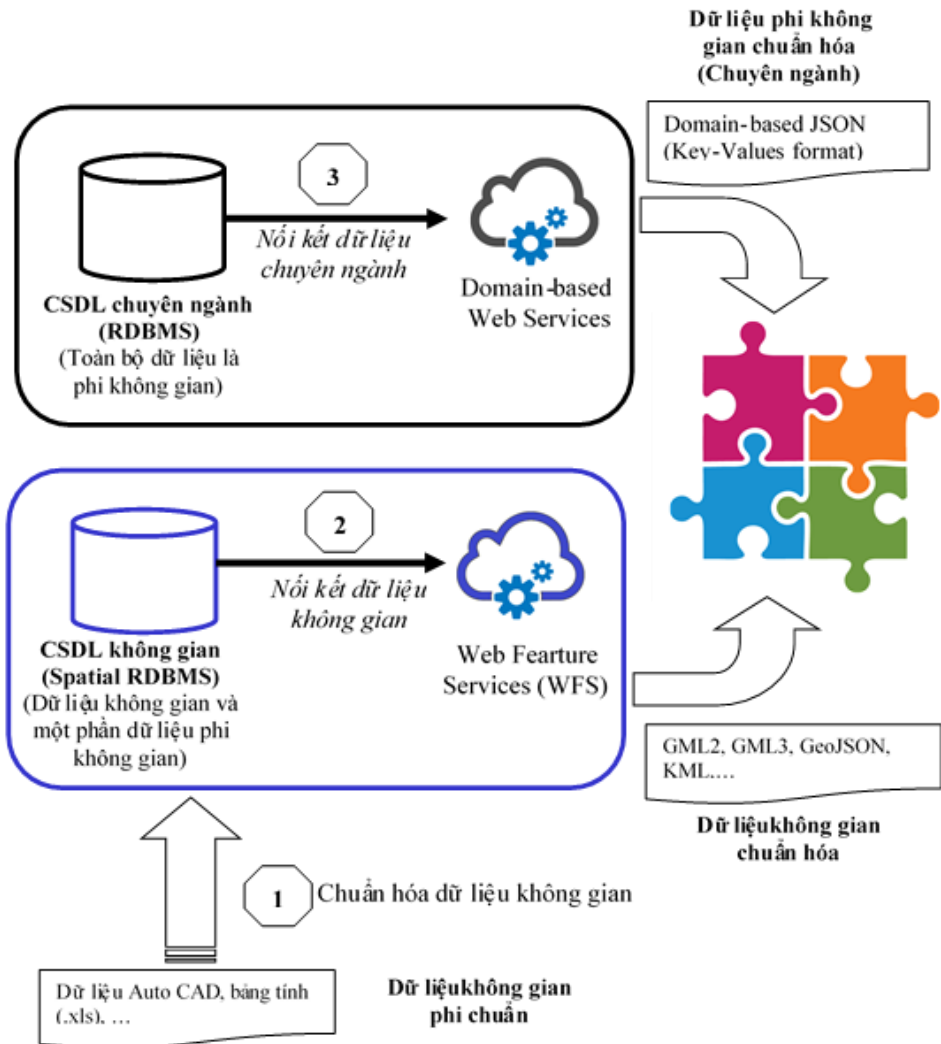
Galarza (2015) giới thiệu *Quy trình chú thích ngữ nghĩa* (Semantic Annotation Process) cho phép giải quyết bài toán nối kết giữa WFS (web feature service) và các dịch vụ RESTful. Cách tiếp cận này về bản chất có thể giải quyết vấn đề thực thể học phân tán được nêu trong Phần 3.3 và có thể xem xét để áp dụng trong mô hình chuẩn hóa mà chúng tôi cần xây dựng.

Vilches-Blázquez and Saavedra (2019) sử dụng thuật ngữ *dữ liệu liên kết* (linked data) để mô tả các thuộc tính dữ liệu liên quan đến các đối tượng không gian trong quản lý các cơ sở hạ tầng quản trị không gian (spatial data infrastructure – SDI), đây cũng là bài toán nối kết giữa WFS và các công nghệ Web ngữ nghĩa. Một mô hình bản đồ Web dựa trên việc tích hợp trực tiếp các thuộc tính phi không gian (đọc từ các dịch vụ web) đã được giới thiệu trong (Nguyễn Văn Kiệt và *ctv.*, 2011), tuy nhiên mô hình này không được mềm dẻo do việc tích hợp được thực hiện trong mã Java.

⁵<https://epsg.io/3405>

⁶<https://epsg.io/3857>

⁷<https://www.openstreetmap.org/>



Hình 6: Mô hình chuẩn hóa dữ liệu hạ tầng

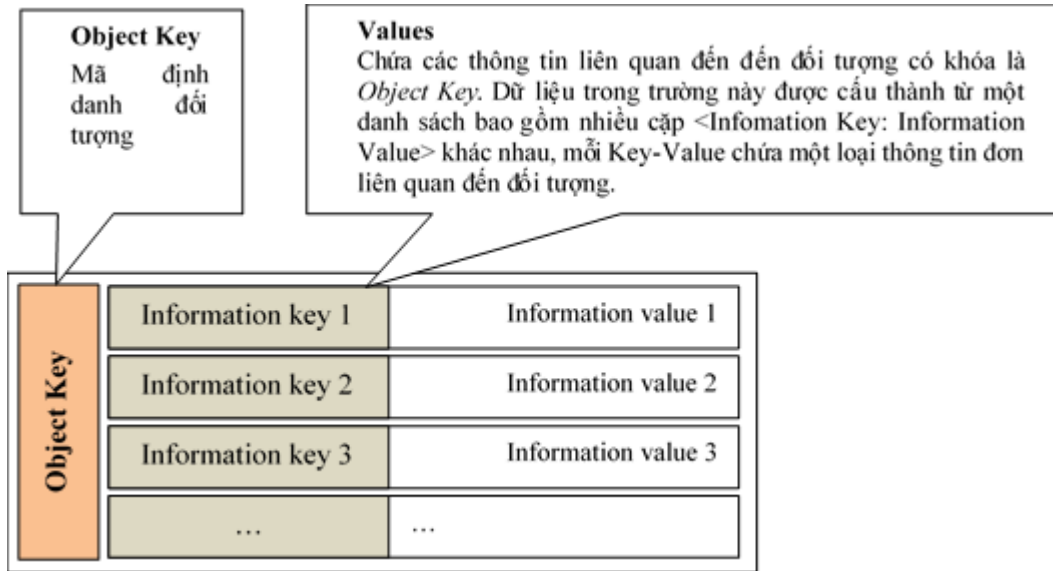
Như đã phân tích trong Phần 4.3.1, các DWS sẽ gắn liền với Hệ thống 1, ở vai trò các API cung cấp dữ liệu phi không gian. Tương tự như vậy, các WFS (phân tích trong Phần 4.3.2) sẽ gắn liền với Hệ thống 2, với các API cung cấp dữ liệu không gian. Hình 6 trình bày *Mô hình chuẩn hóa dữ liệu hạ tầng*, bao gồm hai hệ thống:

Hệ thống Dữ liệu chuyên ngành: Hệ quản trị CSDL phi không gian (chứa chủ yếu các thuộc tính phi không gian của đối tượng) và các dịch vụ *Domain-based Web Service* (DWS).

Hệ thống Dữ liệu không gian: Hệ quản trị CSDL không gian (chứa chủ yếu các thuộc tính không gian và kiểu hiển thị) và các dịch vụ chuẩn của OGC, trọng tâm nghiên cứu trong mô hình này là *Web Feature Service* (WFS).

4.3.1 Domain-based Web Service (DWS)

Domain-based Web Service (DWS) là hệ thống dịch vụ web chuyên ngành được thiết kế phục vụ cho nhu cầu truy xuất các thuộc tính *phi không gian* có sẵn từ các hệ thống thông tin khác nhau. Khác với WFS, các DWS chưa được chuẩn hóa và không chứa dữ liệu không gian. Đây là các dịch vụ được xây dựng để giải quyết vấn đề thực thể học phân tán được nêu trong Phần 3.3. Tham số chính cho các DWS là mã định danh các *đối tượng* (hay *thực thể*) cần truy vấn. Kết quả trả về là một cấu trúc dữ liệu chứa thông tin bao gồm các thuộc tính mở rộng cho các đối tượng không gian có trên bản đồ. Trong đó, kết xuất của các DWS theo định dạng <Khóa đối tượng: Tập giá trị> (<Object Key : Values>) được mô tả như sau:



Hình 7: Định dạng dữ liệu trả về của các DWS

Để hiểu rõ hơn các định dạng dữ liệu của các DWS, hai ví dụ điển hình liên quan đến quản lý các đối tượng **PHÒNG HỌC** và **ĐỊA ĐIỂM HOẠT ĐỘNG NGOÀI TRỜI** trong quản lý bản đồ Khu II Đại học Cần Thơ được đưa ra.

[1] Phòng học:

Cấu trúc dữ liệu trả về:

```
"54":{"TENPHONG":"101/C1",
      "SUCCHUA":"80",           "DIENTICH":"77",
      "CSVC":"Bàn 2 chỗ(40), Loa(1), Máy chiếu(1),
      Quạt(1), Âm thanh(1), Đèn(12)",
      "TRANGTHAI":"NO",
      "TENHOATDONG":"NO"}
```

Trong đó, các thành phần của tập giá trị trả về được phân tích như sau:

Object Key	Values	Giải thích các thuộc tính
"54"	<pre>{ "TENPHONG":"101/C1", "SUCCHUA":"80", "DIENTICH":"77", "CSVC":"Bàn 2 chỗ(40), Loa(1), Máy chiếu(1), Quạt(1), Âm thanh(1), Đèn(12)", "TRANGTHAI":"NO", "TENHOATDONG":"NO" }</pre>	<p>Tên phòng</p> <p>Sức chứa</p> <p>Diện tích</p> <p>Cơ sở vật chất (Bên trong)</p> <p>Trạng thái hoạt động (Sinh viên có mượn phòng để tổ chức hoạt động ngoại khóa hay không)</p> <p>Tên hoạt động ngoại khóa (Nếu có)</p>

[2] Địa điểm hoạt động ngoài trời:

Cấu trúc dữ liệu trả về:

```
"20":{"MADIADIEM":"CN1",
      "TENDIADIEM":"Sân bóng chuyền K. Công nghệ",
      "MAKHUVUC":"CN",
      "TENKHUVUC":"Khoa Công nghệ",
```

```
"TRANGTHAI":"YES",
      "TENHOATDONG":"Giải bóng chuyền khoa Công nghệ"}
```

Trong đó, các thành phần của tập giá trị trả về được phân tích như sau:

Object Key	Values	Giải thích các thuộc tính
"20"	<pre> { "MADIADIEM": "CN1", "TENDIADIEM": "Sân bóng chuyên K. Công nghệ", "MAKHUVUC": "CN", "TENKHUVUC": "Khoa Công nghệ", "TRANGTHAI": "YES", "TENHOATDONG": "Giải bóng chuyên khoa Công nghệ" } </pre>	Mã địa điểm Tên địa điểm Mã khu vực (Khoa, Phòng, ...) Tên khu vực (Khoa, Phòng, ...) Trạng thái hoạt động (Sinh viên có mượn địa điểm để tổ chức hoạt động ngoại khóa hay không) Tên hoạt động ngoại khóa (Nếu có)

Trong hai ví dụ trên, các thuộc tính TRANGTHAI và TENHOATDONG liên quan đến việc sinh viên đăng ký với trường việc mượn địa điểm (có thể trong phòng học hoặc ngoài trời) để phục vụ cho các hoạt động ngoại khóa. Do vậy, hai thuộc tính này tồn tại trong cả hai nhóm DWS. Tuy nhiên, các thuộc tính khác cũng có thể được bổ sung vào kết quả trả về tùy theo yêu cầu riêng, chẳng hạn, trạng thái của PHÒNG HỌC có thể liên quan đến việc việc bố trí lịch giảng dạy, báo cáo đồ án, ... Khi đó, các thông tin liên quan sẽ được bổ sung vào thuộc tính TENHOATDONG, hoặc bổ sung thêm các thuộc tính mới tùy ý của bên kết xuất dịch vụ.

4.3.2 Web Feature Service (WFS)

Web Feature Service (WFS) là cách phân phối các đặc trưng địa lý thông qua một dịch vụ web đến với ứng dụng phía Người dùng (Client) hoặc một Trình duyệt (Browser). Người dùng có thể yêu cầu dữ liệu một cách có chọn lọc để phục vụ cho nhu cầu riêng của mình. Web Feature Service là một giao

thức chuẩn cho việc phân phối dữ liệu vector đến ứng dụng đầu cuối theo các nhu cầu truy vấn khác nhau, kết quả trả về rất đa dạng, trong đó có các chuẩn dữ liệu bản đồ như GML2, GML3, GeoJSON, KML, ... Đây là một trong một số nhóm dịch vụ chuẩn phổ biến được định nghĩa bởi Tổ chức OGC (The Open Geospatial Consortium) (OGC, n.d.), như WMS (web map service), WCS (web coverage service), WMTS (web map tile service) và một số dịch vụ mở rộng như CSW (catalogue service) và WPS (web processing service).

WFS được chúng tôi lựa chọn do cung cấp dữ liệu dạng Vector, qua đó cung cấp khả năng tương tác rất hữu hiệu ở người dùng đầu cuối. Do nhu cầu nói kết mở rộng với các DWS – cũng kết xuất dữ liệu định dạng JSON, nên việc lựa chọn WFS là rất phù hợp cho Mô hình chuẩn hóa dữ liệu hạ tầng.

Cùng quan sát một mẫu tin được kết xuất từ WFS theo định dạng GeoJSON cho dữ liệu PHÒNG HỌC/PHÒNG LÀM VIỆC, cấu trúc dữ liệu trả về như sau:

```

{"uri": "https://ctu.edu.vn/quanlyphong/",
 "type": "Feature", "id": "tang2_lau1.7", "geometry": {"type": "MultiPolygon", "coordinates": [[[[[1.177368071096346E7, 1122461.97905321], [1.177368385335498E7, 1122465.72928437], [1.177369052889346E7, 1122460.06237269], [1.177368738650183E7, 1122456.31214269], [1.177368071096346E7, 1122461.97905321]]]]]}, "geometry_name": "geom", "properties": {"id": "7", "phong": "Phòng làm việc 3", "dia diem": "Khu nhà thực tập chuyên ngành thủy sản", "dien_tich": "41.28", "suc_chua": null, "donvi_quanly": "KHOA THỦY SẢN", "cong_nang": "PHÒNG LÀM VIỆC, VĂN PHÒNG"}}
                    
```

Trong đó, các thành phần của tập giá trị trả về được phân tích như sau:

Values	Giải thích các thuộc tính
<pre> { "uri": "https://ctu.edu.vn/quanlyphong/" "type": "Feature", "id": "tang2_lau1.7", "geometry": { "type": "MultiPolygon", "coordinates": [[[[[1.177368071096346E7, 1122461.97905321], [1.177368385335498E7, 1122465.72928437], [1.177369052889346E7, 1122460.06237269], [1.177368738650183E7, 1122456.31214269], [1.177368071096346E7, 1122461.97905321]]]]]] </pre>	Định danh nguồn dữ liệu Web Feature Service (WFS) Mã định danh Tọa độ dạng hình học Kiểu dữ liệu không gian (Hình học) Các tọa độ

<pre> }, "geometry_name": "geom", "properties": { "id": 7, "phong": "Phòng làm việc 3", "dien_tich": 41.28, "suc_chua": 15, "donvi_quanly": "KHOA THỦY SẢN", "cong_nang": "PHÒNG LÀM VIỆC, VĂN PHÒNG" } } </pre>	<p>Tên kiểu dữ liệu (Geom) Các thuộc tính phi không gian</p> <p>Mã đối tượng (Dùng trong nối kết thuộc tính phi không gian với DWS)</p> <ul style="list-style-type: none"> Phòng Diện tích Sức chứa Đơn vị quản lý Công năng <p>Các thuộc tính phi không gian sẵn có của đối tượng</p>
--	--

Trong kết xuất của WFS, có hai nhóm thuộc tính cần được lưu ý, đó là **“geometry”** và **“id”**. Thuộc tính **“geometry”** chứa các thông số không gian của đối tượng. Trong khi đó, thuộc tính **“id”** xác định rõ mã đối tượng, hay **mã thực thể phân tán**. Mã này sẽ

được ánh xạ vào **Object Key** (Mã đối tượng) của DWS để truy vấn và mở rộng các thuộc tính phi không gian (bên cạnh một số thuộc tính phi không gian sẵn có). Có thể nói **“id”** chính là chìa khóa nối kết giữa WFS và DWS trong mô hình đề xuất.



a) Trạng thái về trang thiết bị, vật chất và hoạt động ngoại khóa của sinh viên (“Không có hoạt động”) tại Phòng học có mã “101/C1”.

b) Trạng thái về hoạt động ngoại khóa của sinh viên (“Sinh hoạt Chi hội Huyện Tam Bình – Vĩnh Long”) tại Địa điểm “DTN1– Sân VP Đoàn thanh niên”.

Hình 8: Giao diện web kết nối hai dịch vụ WFS và DWS trong quản lý cơ sở hạ tầng tại Khu II – Đại học Cần Thơ

Hình 8 trình bày một số ví dụ minh họa về việc nối kết giữa WFS và DWS áp dụng trong quản lý cơ sở hạ tầng Khu II – Đại học Cần Thơ. Trong đó, hai đối tượng được quan tâm là *Phòng học* và *Địa điểm sinh hoạt ngoại khóa*. Đối với *Phòng học* (Hình 8a), thông tin mà dịch vụ DWS kết xuất bao gồm trạng thái về cơ sở vật chất, trang thiết bị có bên trong phòng và trạng thái sinh hoạt ngoại khóa của sinh viên (nếu có đăng ký). Đối với *Địa điểm sinh hoạt ngoại khóa* (Hình 8b), chỉ duy nhất trạng thái sinh hoạt ngoại khóa của sinh viên được kết xuất.

5 KẾT LUẬN VÀ ĐỀ XUẤT

5.1 Kết luận

Bài viết đã hệ thống hóa tất cả các vấn đề khó khăn nhất liên quan để việc chuẩn bị nền tảng dữ liệu cho việc xây dựng một hệ thống bản đồ có tương tác động. Kết quả của bài viết là một phần trong

nghiên cứu sâu hơn về việc xây dựng bản đồ nền Web cho quản trị hệ thống cơ sở hạ tầng Khu II – Đại học Cần Thơ. *Mô hình chuẩn hóa dữ liệu hạ tầng* được đề xuất dựa trên nguyên lý thực thể học phân tán (distributed ontology). Mô hình này có thể áp dụng trong quản lý các đối tượng cơ sở hạ tầng của một cơ quan tổ chức hoặc các dự án SDI ở phạm vi đô thị. Dữ liệu *phòng học* và các *địa điểm sinh hoạt ngoại khóa* trong khuôn viên Khu II – Trường Đại học Cần Thơ cũng được minh họa để làm rõ hơn các khía cạnh kỹ thuật.

Với mô hình đề xuất dựa trên hai nhóm dịch vụ Web: *Domain-based Web Services (DWS)* và *Web Features Services (WFS)* (The Open Geospatial Consortium, n.d.), nghiên cứu đã đạt được mục tiêu chuẩn hóa hai nhóm dữ liệu đặc thù là phi không gian và không gian, sẵn sàng vận hành trên môi trường phân tán để kết xuất các bản đồ tương tác.

Quy trình và mô hình chuẩn hóa dữ liệu hạ tầng đề xuất đã góp phần giải quyết hai vấn đề nan giải của các bài toán quản lý bản đồ trong một tổ chức có quy mô lớn về diện tích và các nhóm đối tượng chuyên ngành, bao gồm: (1) vấn đề thiếu hoặc tồn tại các dữ liệu bản đồ chuẩn Spagetti, sẵn sàng tham gia vào các hệ thống GIS phân tán kiểu mẫu và (2) vấn đề không thể truy cập trực tiếp đến các hệ quản trị CSDL có sẵn mặc dù chúng đã tồn tại trên các hệ thống thông tin đang vận hành của tổ chức, chủ yếu liên quan đến tính phức tạp của các hệ thống này hoặc yếu tố về bảo mật hệ thống.

5.2 Đề xuất

Việc áp dụng mô hình chuẩn hóa dữ liệu hạ tầng được đề xuất mang tính quy trình, không bắt buộc về cấu trúc dữ liệu của các đối tượng cần quản lý. Do vậy, việc áp dụng mô hình mang tính gợi mở và cần linh động trong việc áp dụng. Bên cạnh đó, cách tiếp cận này cũng góp phần làm đơn giản hóa các giải pháp quản lý dữ liệu cơ sở hạ tầng, qua đó có thể phát triển các ứng dụng GIS phân tán một cách hữu hiệu và mang tính khả thi cao cho các tổ chức có quy mô lớn. Việc xây dựng dịch vụ tương tự *Domain-based Web Services (DWS)* là một xu hướng đã rất phổ biến đối với các tổ chức công nghệ chuyên về bản đồ (chẳng hạn, các *Google Map API*, hay *Viet Map API*). Giải pháp này nên được phát triển mạnh và lưu ý nâng cao các vấn đề liên quan đến bảo mật dữ liệu và ứng dụng.

TÀI LIỆU THAM KHẢO

Belussi, A., Catania, B., Clementini, E., and Ferrari, E., 2007. Spatial Data on the Web: Issues and Challenges. *In: Belussi, A., Catania, B., Clementini, E., Ferrari, E. (Eds.). Spatial Data on the Web.* Springer, Berlin, Heidelberg, pp. 1–12. https://doi.org/10.1007/978-3-540-69878-4_1.

Galarza, V.S., 2015. Semantic Annotation of RESTful and WFS OGC Services. Doctoral Thesis. Madrid, Technical University of Madrid. <http://oa.upm.es/35030/>.

Gruber, T.R., 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition.* 5 (2): 199–220. <https://doi.org/https://doi.org/10.1006/knac.1993.1008>.

Guarino, N., Oberle, D., and Staab, S., 2009. What Is an Ontology? *In: Staab, S., Studer, R. (Eds.). Handbook on Ontologies.* Springer, Berlin, Heidelberg, pp. 1–17. <https://doi.org/10.1007/978-3-540-92673-3>.

Hu, Y., Li, W., 2017. Spatial Data Infrastructures. *In: John P. W. (Ed.). Geographic Information Science & Technology Body of Knowledge (2nd*

Quarter 2017 Edition). <https://doi.org/10.22224/gistbok/2017.2.1>.

Kresse, W., and Danko, D.M., eds. 2012. *Springer Handbook of Geographic Information.* Springer Handbooks. Springer. <https://doi.org/10.1007/978-3-540-72680-7>.

Narula G.S., Wason R., Jain V., and Baliyan A., 2018. Ontology mapping and merging aspects in semantic web. *International Robotics & Automation Journal.* 4(1): 23–27. DOI: 10.15406/iratj.2018.04.00087

Nguyễn Văn Kiệt, Trương Xuân Việt, Lê Quyết Thắng. 2011. Xây dựng Dịch vụ Bản đồ Tương tác với các WebServices dựa trên Kiến trúc SOA. *Trong: Kỷ yếu Hội nghị Tổng kết 5 năm nghiên cứu Khoa học và Đào tạo Khoa CNTT&TT – Đại học Cần Thơ, tháng 12/2011, Cần Thơ. Đại học Cần Thơ, 67-76.*

Open Geospatial Consortium (OGC). n.d. OGC Standards, accessed on 27 September 2020. Available from <https://www.ogc.org/docs/is>.

Object Management Group (OMG). 2018. *Distributed Ontology, Model, and Specification Language*, accessed on 27 September 2020. Available from <https://www.omg.org/spec/DOL/1.0/>.

Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal.* 10 (1): 439–46. <https://doi.org/10.32614/RJ-2018-009>.

Campbell, J.E., Shin, M. 2012. Data Models for GIS. *In: Essentials of Geographic Information Systems (Textbooks. 2), pp. 75-101.* <https://digitalcommons.liberty.edu/textbooks/2>

Open Geospatial Consortium (OGC). n.d. Web Feature Service (OGC Standards), accessed on 27 September 2020. Available from <https://www.ogc.org/standards/wfs>.

Truong, X.V., Drogoul, A., Huynh, X.H., and Le,N.M., 2011. Modeling the Brown Plant Hoppers Surveillance Network Using Agent-Based Model: Application for the Mekong Delta Region. *In: Rahwan I., Wobcke W., Sen S., Sugawara T. (Eds.) PRIMA 2012: Principles and Practice of Multi-Agent Systems. PRIMA 2012. Lecture Notes in Computer Science, vol 7455, pp. 228–242.* Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-32729-2_16

Vilches-Blázquez, L.M., and Saavedra, J., 2019. A Framework for Connecting Two Interoperability Universes: OGC Web Feature Services and Linked Data. *Transactions in GIS.* 23 (1): 22–47. <https://doi.org/10.1111/tgis.12496>.