



BÁO CÁO NGHIỆM THU NHIỆM VỤ KHOA HỌC VÀ CÔNG NGHỆ

Xây dựng Cổng Dữ liệu Mở VUSTA Đáp ứng Yêu cầu Chuyển đổi Số

Báo cáo Tổng kết & Khảo chứng Kết quả Thực hiện Thuyết minh

Chủ nhiệm đề tài: TS. Nguyễn Thị Bích Lan

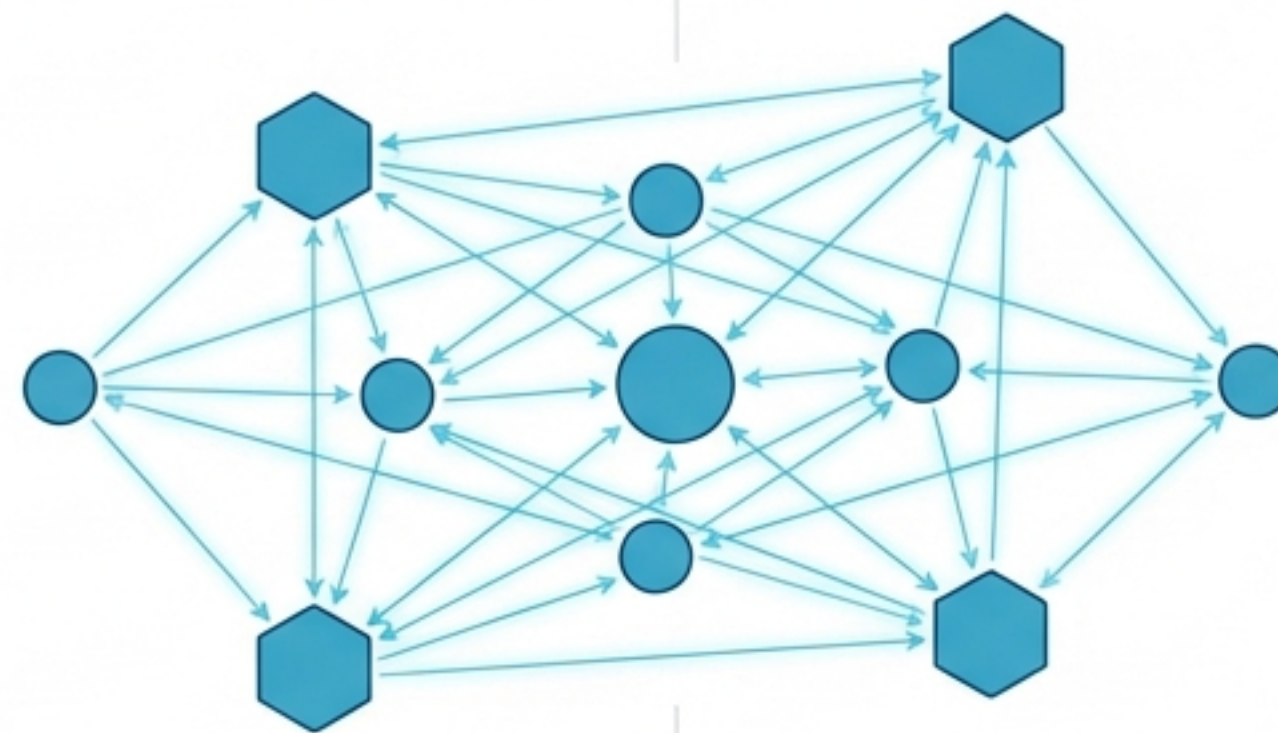
Tổ chức chủ trì: Viện Sáng tạo và Chuyển đổi số (VIDTI)

Cơ quan chủ quản: Liên hiệp các Hội Khoa học và Kỹ thuật Việt Nam



Vấn đề Hiện tại - Khoảng trống Thực thi

- **90% dữ liệu** trên Cổng Dữ liệu Quốc gia (data.gov.vn) đang ở **định dạng PDF** (đóng, máy không đọc được).
- **Thực trạng VUSTA**: Tài sản tri thức khổng lồ nhưng phân tán thành các “ốc đảo” (Word, PDF rời rạc), chưa tạo ra giá trị liên kết.



Mục tiêu Nhiệm vụ - Giải pháp Chiến lược

- **Định hướng**: Tuân thủ Nghị định 47/2020/NĐ-CP về kết nối, chia sẻ dữ liệu.
- **Giải pháp**: Chuyển đổi từ “**Kho lưu trữ tĩnh**” sang “**Hệ sinh thái Dữ liệu mở Liên kết 5-sao**”, kiến tạo hạ tầng số làm bệ phóng cho đổi mới sáng tạo.

ĐỐI CHIẾU 1: Mục tiêu Nghiên cứu

Cam kết từ Thuyết minh

Kết quả Báo cáo Tổng kết

Nghiên cứu tổng quan và quy định pháp lý về dữ liệu mở.



Chương 1 & 2: Khung pháp lý, mô hình W3C, khảo sát thực trạng hạ tầng VUSTA.

Xây dựng danh mục và quy trình xuất bản dữ liệu.



Chương 3: Danh mục 10 nhóm dữ liệu trọng điểm & Quy trình kỹ thuật 6 bước vòng đời dữ liệu.

Xây dựng cổng dữ liệu mở chuẩn quốc tế.



Chương 4: Kiến trúc hạ tầng CKAN, thiết kế DB lai, bảo mật RBAC.

Cung cấp dịch vụ dữ liệu tại opendata.vusta.vn.



Triển khai thực tế: Hệ thống live với API tự động, Data Explorer, UI/UX chuẩn hóa.

ĐỐI CHIẾU 2: Bảng Điều Khiển Sản Phẩm Giao Nộp



10 Báo cáo Chuyên đề (Đạt 100%)

Từ tổng quan W3C, tiêu chuẩn D-CAT, đến thiết kế UI/UX và kiến trúc phần mềm.



03 Hội thảo Khoa học (Đạt 100%)

Góp ý tiêu chuẩn, quy trình xuất bản, và đánh giá cổng dữ liệu mở.



02 Báo cáo Chuyên gia (Đạt 100%)

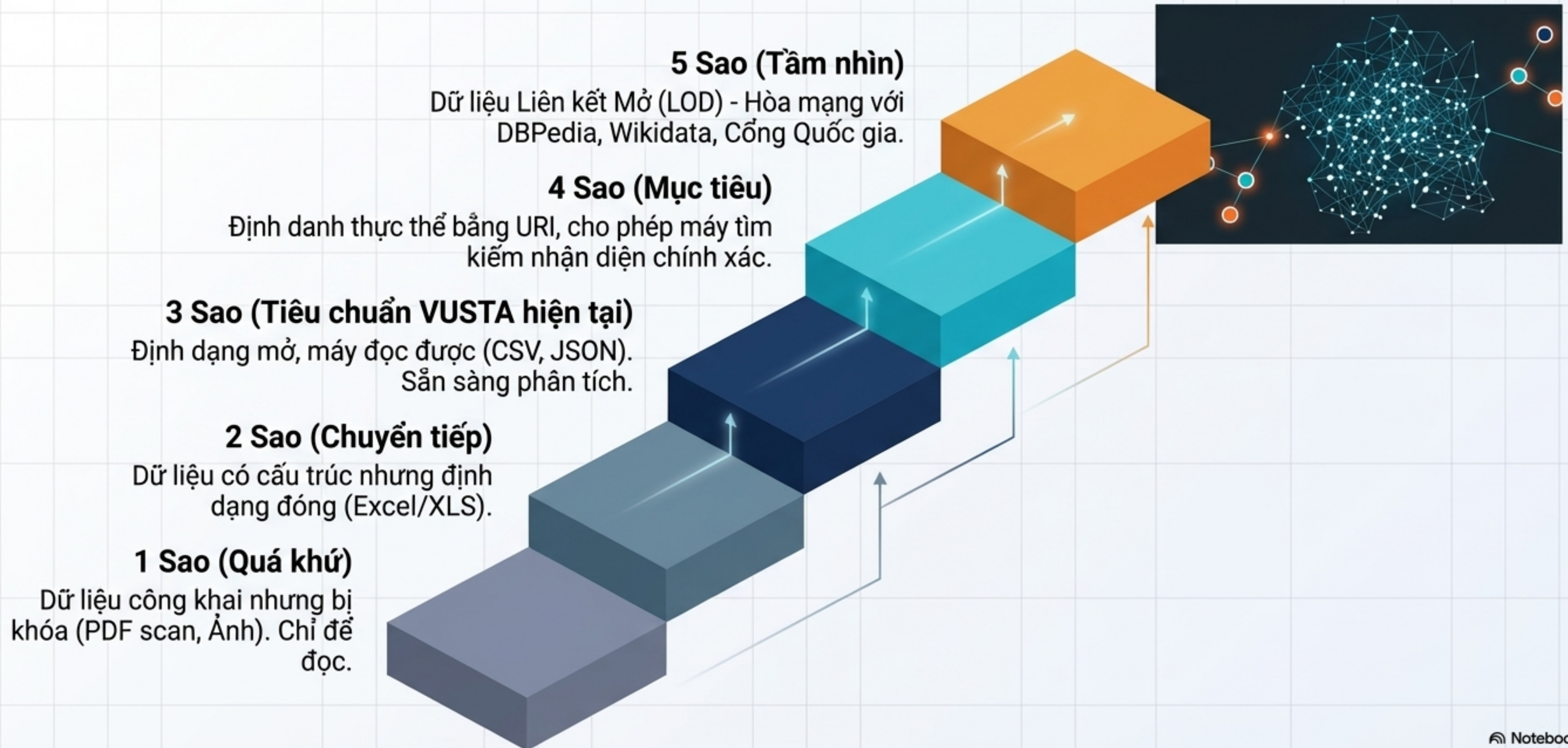
Tư vấn kiến trúc hệ thống và tích hợp dữ liệu Đổi mới Sáng tạo.



01 Cổng Phần mềm (Đạt 100%)

Đã triển khai thực tế trên hạ tầng đám mây tại tên miền opendata.vusta.vn.

Khung Tiêu chuẩn Kỹ thuật: Lộ trình Dữ liệu W3C 5-Sao

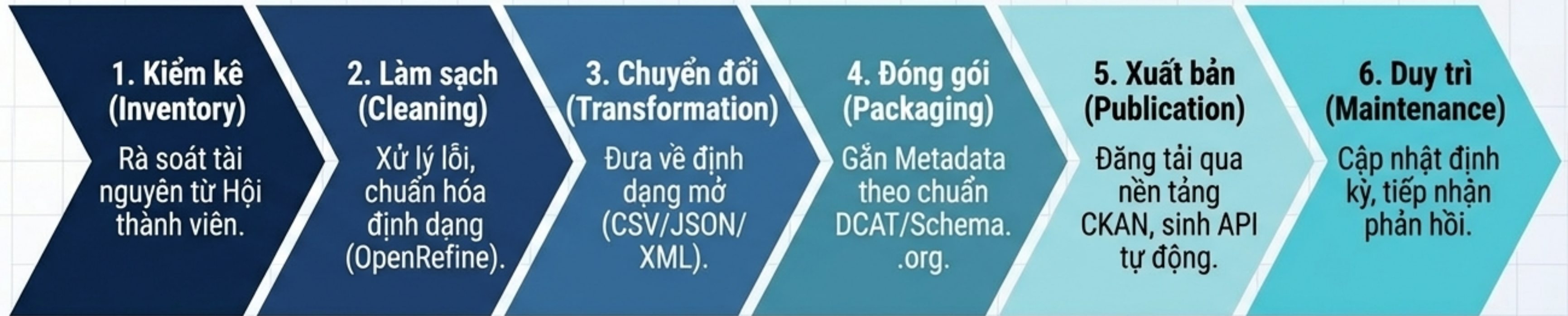


Phân loại & Chuẩn hóa: 10 Nhóm Dữ liệu Trọng điểm



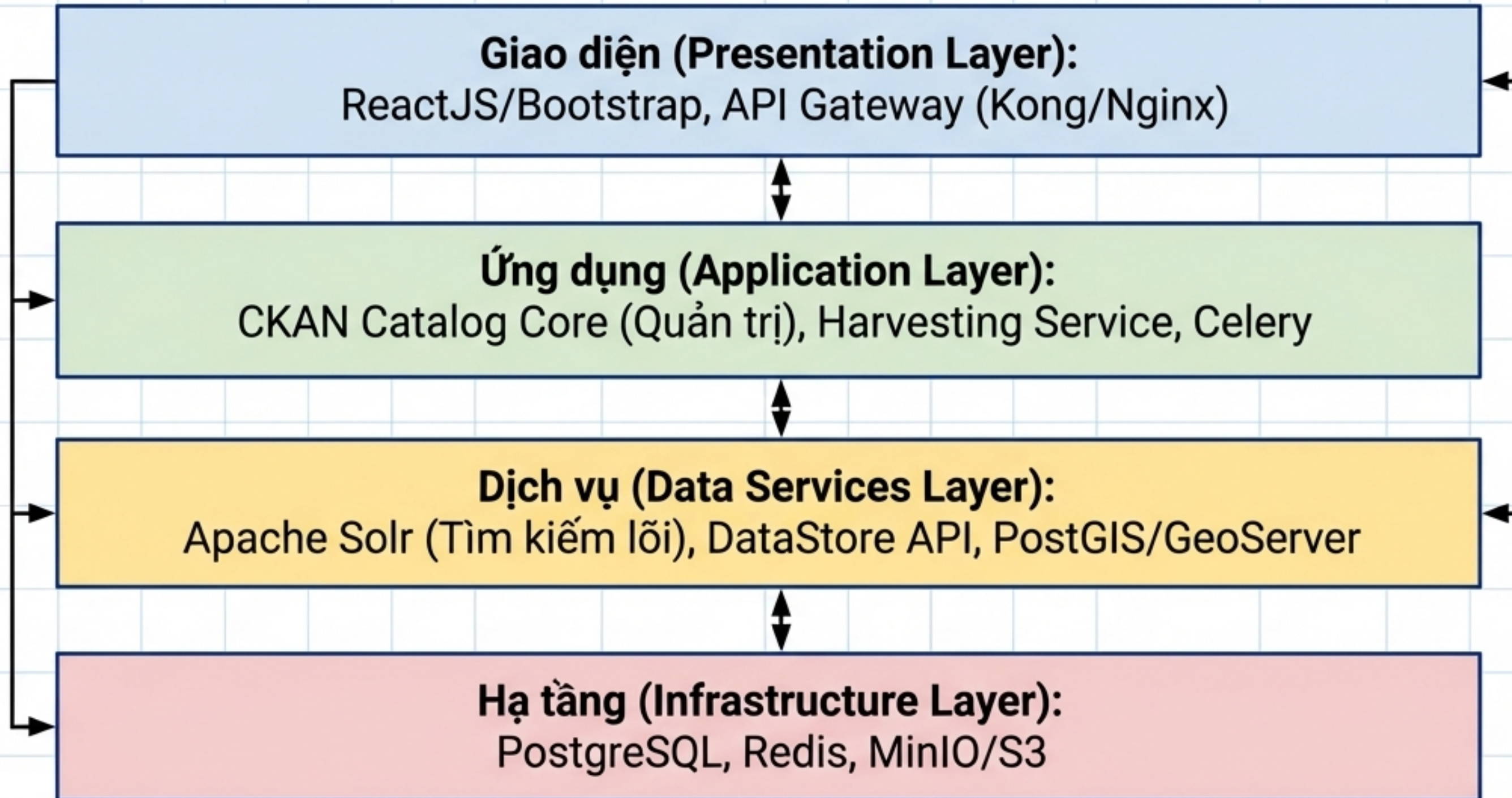
Áp dụng chuẩn DCAT-AP giúp 100% siêu dữ liệu (metadata) của VUSTA tự động tương thích với Cổng Dữ liệu Quốc gia.

Quy trình Kỹ thuật: Vòng đời Dữ liệu Khép kín



Quy trình chuẩn hóa đảm bảo giải quyết triệt để vấn đề 'Rác dữ liệu' (Dirty Data) trước khi lên hệ thống.

Kiến trúc Hệ thống: Nền tảng Tiêu chuẩn Vàng CKAN



Việc sử dụng nền tảng mở CKAN đưa VUSTA sánh ngang với tiêu chuẩn hạ tầng lưu trữ của data.gov.uk và data.gov.vn.

Thiết kế Dữ liệu Động (Hybrid Database) & API Tự động

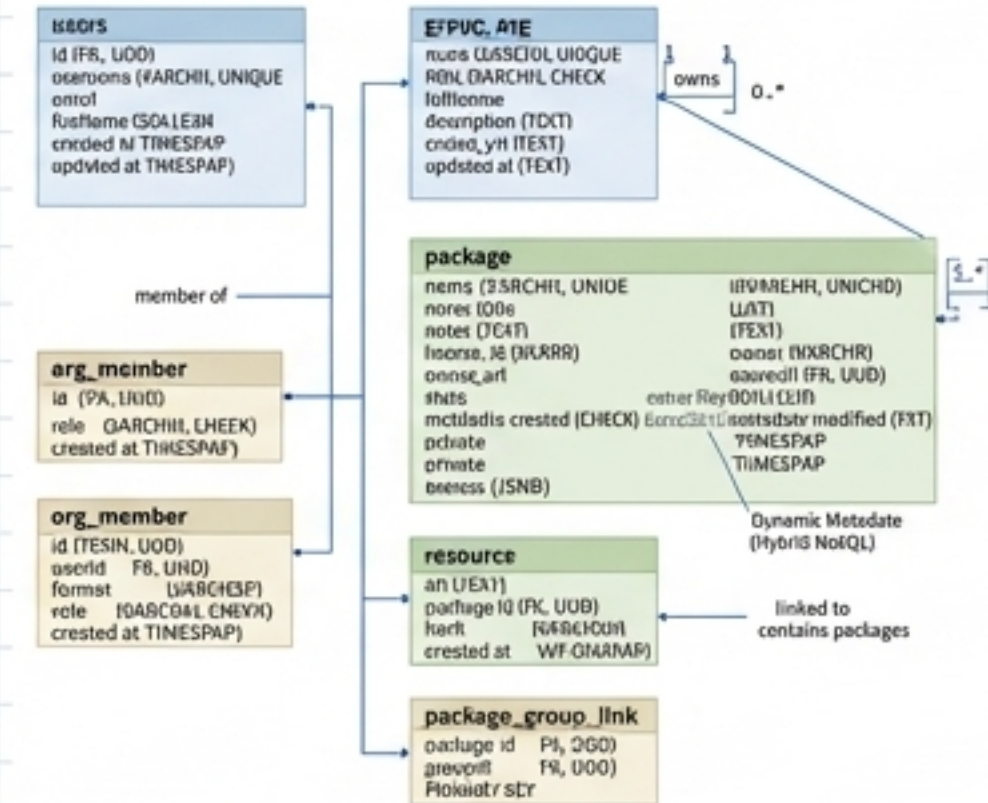
Vấn đề



Vấn đề: Dữ liệu KH&CN cực kỳ đa dạng (Y tế có mã bệnh, Khí hậu có tọa độ). CSDL quan hệ cứng nhắc không thể đáp ứng.

Giải pháp

VUSTA PostgSGroup Hybrid Relational Schema

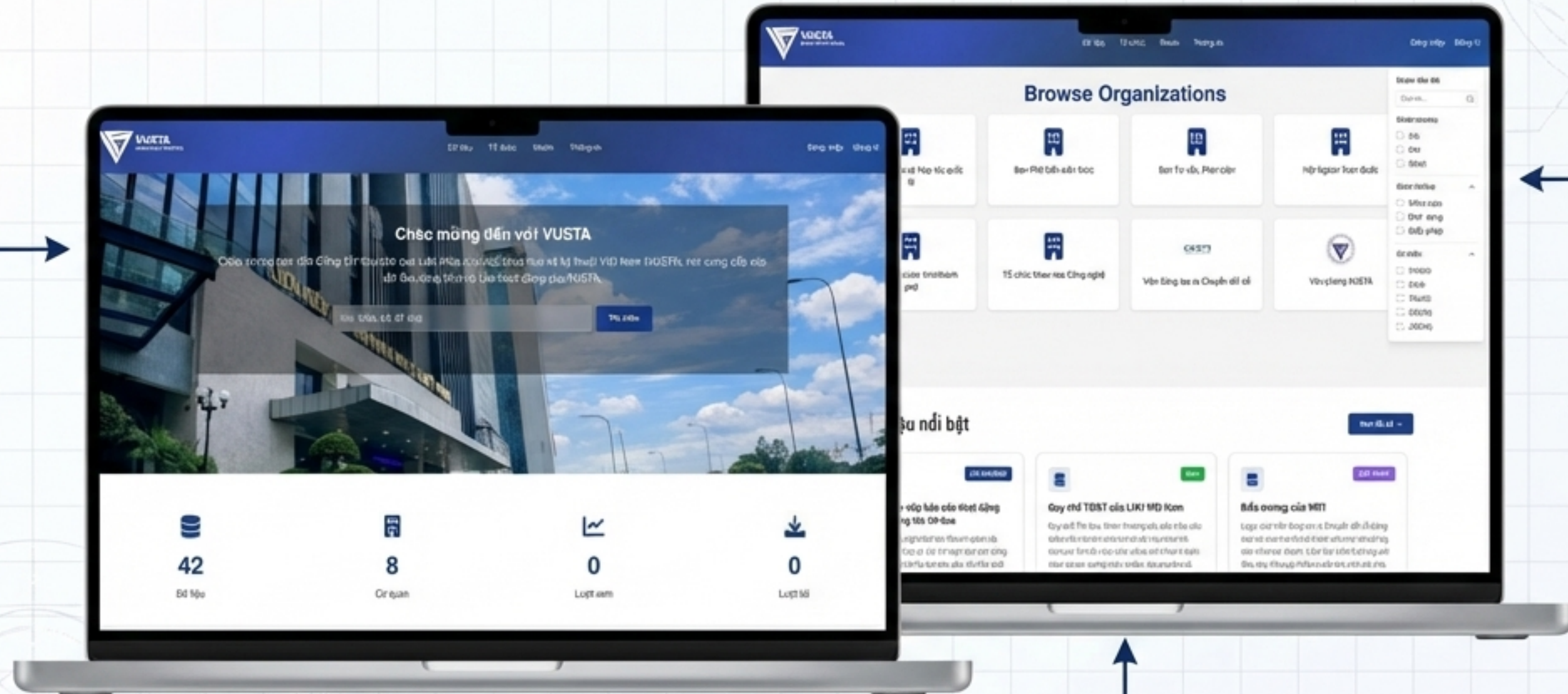


- Giải pháp Hybrid (PostgreSQL JSONB):**
- **Cấu trúc Cốt lõi (SQL):** Giữ tính toàn vẹn cho quản trị user, tổ chức.
 - **Siêu dữ liệu Động (NoSQL/JSONB):** Cho phép mỗi tập dữ liệu tự định nghĩa trường thông tin.

API Generation: Hệ thống tự động sinh RESTful API cho mọi bộ dữ liệu dạng bảng, cho phép máy móc truy vấn trực tiếp (row-level access).

Giao diện & Trải nghiệm Người dùng (UI/UX)

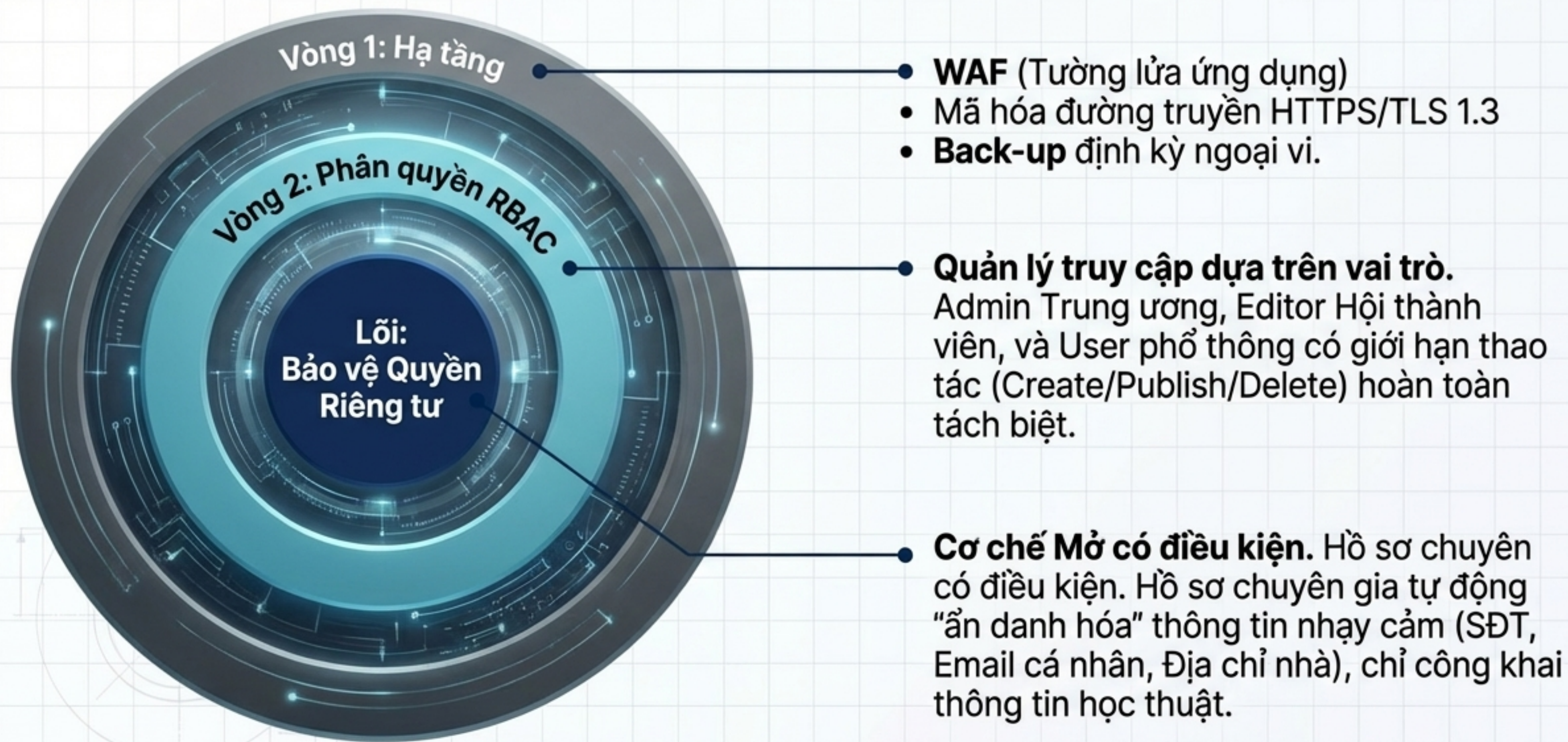
1
Trang Chủ: Thiết kế sang trọng, Dashboard thống kê thời gian thực (Lượt tải, Số bộ dữ liệu).



2
Tìm Kiếm Đa Chiều (Faceted Search): Lọc siêu tốc theo Tổ chức cung cấp, Định dạng (CSV, JSON), và Giấy phép.

3
Khám Phá Trực Quan (Data Explorer): Xem trước dữ liệu ngay trên trình duyệt dưới dạng Lưới (Grid), Biểu đồ (Graph), hoặc Bản đồ (Map) không cần tải về.

Mô hình Bảo mật & An toàn Thông tin Cơ sở Dữ liệu



Tổng hợp Đánh giá Tác động & Giá trị Lỗi

Giá trị về Dữ liệu

Mở khóa Tri thức:
Biến 500+ 'kho lưu trữ câm' thành mạng lưới tri thức sống động, có thể tái sử dụng và phân tích bằng AI/ML.

Giá trị về Kết nối

Tương tác Hệ sinh thái: Sẵn sàng liên thông API trực tiếp với Cổng Dữ liệu Quốc gia nhờ tuân thủ tuyệt đối chuẩn W3C và DCAT-AP.

Giá trị về Chính sách

Khẳng định Vị thế: Đưa VUSTA trở thành tổ chức đi tiên phong trong việc thực thi Nghị định 47, đóng góp vào nền kinh tế số quốc gia.

Hệ sinh thái Số VUSTA

Kiến nghị Triển khai & Lộ trình Tiếp theo

Bước 1: Thể chế hóa Dữ liệu (Ngắn hạn)

Ban hành quy chế
"Dữ liệu là Tài sản Số"
trong toàn VUSTA.
Bắt buộc công bố dữ liệu
là nhiệm vụ quản trị
thường xuyên.

Bước 2: Triển khai Thí điểm Trọng tâm (Trung hạn)

Ưu tiên xuất bản các
nhóm dữ liệu 'Sẵn sàng
cao' trước: Danh mục
Tổ chức (Nhóm 1) và
Nhiệm vụ KH-CN
(Nhóm 3).

Bước 3: Đầu tư Con người & Kết nối (Dài hạn)

Xây dựng đội ngũ
"Quản trị viên dữ liệu"
(Data Stewards).
Mở rộng API kết nối
tích hợp hệ thống Trục
liên thông văn bản
quốc gia.