SCIENSANO
CONNECTS HEALTH, SCIENCE AND SOCIETY

Sciensano - Belgium's national public health institute

Pathways to Generative AI Innovation in Public Administration

.be

Innovation in health information systems

Wisdom

Knowledge

Information

Data

HEALTH INFORMATION SYSTEM

Human resources, Legal framework, Political setting, ICT infrastructure

Innovating and pioneering in data technologies for secondary use of health information for research and evidence- based health policies

Engaging with and shaping the (European) health data ecosystem

Building **strong partnerships for knowledge exchange** and collaboration in health information

in **Europe**, in **Belgium** and in **Sciensano**

sciensano

.be

1/ Metadata AI Assistant

- We are building the next generation of interoperable data catalogues for the European Health Data Space.

- We co-created with European partners the HealthDCAT-Application Profile which extends the DCAT standard for the health domain.

- Our approach is standards-first—grounded in DCAT and its application profiles—augmented by a metadata AI assistant that turns partial or free-text inputs into consistent, compliant, machine-actionable records.

- The expected result is faster onboarding, higher-quality metadata, and seamless cross-border discovery and reuse across EU Data Spaces.

sciensano

.be

WHY — Reduce the administrative burden on health data holders by turning complex EHDS/Data Act/DGA/HVD requirements into guided, machine-assisted steps. — Raise metadata quality and interoperability so datasets are FAIR-by-default and reliably discoverable and reusable across EU Data Spaces and HDAB workflows.

WHAT — A metadata AI assistant that analyses documentation, DCAT-AP records and proposes compliant completions (i.e.: EHDS categories, health themes, keywords, applicable legislation, access conditions, sensitivity level). — It learns from open RDF catalogs and controlled vocabularies to auto-suggest and validate against cardinalities and rules, producing explainable justifications with confidence and exporting standards-compliant RDF via persistent URIs.

2/ Semantic Analyser for FAIR datasets

- We're building a semantic analyser that pairs CSV on the Web (CSVW) with HealthDCAT-AP to capture table/column structure (data types, units, concepts)

- It uses GenAI to propose precise semantic annotations (RAG/GraphRAG).

- It turns plain tables into self-describing, machine-actionable assets ready for EHDS and EU Data Spaces.

sciensano

.be

WHY — Understanding and reusing health datasets is hard because sensitivity limits what can be exposed, and structure and semantics are often not captured explicitly in metadata. — Automating column-level descriptions and mappings to controlled vocabularies reduces onboarding time and raises consistency, FAIRness, and compliance.

WHAT — A CSVW-driven extension to HealthDCAT-AP that records schema, units, data types, and links between datasets and variables, plus a GenAI semantic analyser that suggests concepts, code lists, categories with confidence and rationale. — The outputs are validated, explainable RDF/JSON-LD + CSVW files, published via persistent URIs and ready for SPARQL and catalog integration.

Contact

Pascal Derycke • Innovation Manager • pascal.derycke@sciensano.be

.be