



*A DCAT-AP v3.0 extension for cataloguing health datasets
under the European Health Data Space (EHDS) Regulation*

Pascal Derycke – Innovation Manager

Innovation in health information systems unit – Sciensano Data Governance

FedOSC Webinar - Data 101 - What is metadata and what does it do? 24/04/2026

Regulatory context: EHDS & Article 77

Why a dedicated health application profile is needed?



EHDS Regulation

*Obligations for describing
& cataloguing health datasets*

The European Health Data Space (EHDS) Regulation introduces specific obligations for describing and cataloguing datasets made available for the secondary use of health data - creating the need for a harmonised, machine-readable metadata standard across Member States.

Article 77 - HDAB obligations

Health Data Access Bodies (HDABs) must provide a catalogue that is:

- ✓ **Public** - Openly accessible to data users
- ✓ **Standardised** - Built on shared, interoperable models
- ✓ **Machine-readable** - Consumable by automated tooling
- ✓ **Descriptive** - Source, scope, features, nature of data
- ✓ **Transparent** - Conditions under which data is available

An extension of DCAT-AP v3.0

An application profile for health datasets within HealthData@EU/EHDS



Definition

HealthDCAT-AP is explicitly defined as an extension of DCAT-AP v3.0. It addresses health-specific requirements while preserving full interoperability with the wider DCAT-AP ecosystem.

Core approach

Reuses the core DCAT-AP classes

Adds classes, properties & constraints

Introduces controlled vocabularies for health

Reused core DCAT-AP classes

Kept intact for interoperability with the wider ecosystem

dcat:Catalog

dcat:CatalogRecord

dcat:Dataset

dcat:Distribution

dcat:DataService

+EXT

Explicit access-level model

Differentiated metadata requirements to handle complex health scenarios



NON-PUBLIC

Mediated access

Personal data. Strict mediated-access model requiring HDAB-related distributions and precise governance metadata.



RESTRICTED

Conditional access

Non-personal data released under defined conditions. Controlled, conditional distribution terms apply.



PUBLIC

Open access

Openly available datasets with standard DCAT-AP baseline metadata and lighter governance requirements.



Governance properties: mandatory Health Data Access Body (HDAB) and refined publisher semantics identifying intermediation entities in alignment with EHDS.

Enhanced dataset characterisation

Properties tailored to the specific needs of health data analysis



Health categories

Mandatory classification

Mandatory health categories structure datasets into domain-aware groupings, supporting discovery across Member States. (Art. 51-1)



Population context

Who the data describes

Minimum and maximum typical age, population coverage, record counts, and unique-individual magnitudes.



Semantic interpretability

What the data means

Properties that explicitly state the coding systems and code values used within the data - no hidden semantics.

Specialised vocabularies for governance

Machine-readable standards for the high sensitivity of the health domain

DPV



Data Privacy Vocabulary

Indicates personal data and declares legal bases for processing.

DQV



Data Quality Vocabulary

Attaches machine-readable quality annotations to datasets.

PROV-O



Provenance Ontology

Traces the origin, lineage and derivation history of data.

ODRL



Usage Policies

Expresses permissions, prohibitions and duties around data use.

CSVW



Variable Documentation

Uses Table and Column entities to describe variables in a structured way.

SAMPLE



Synthetic / Anonymized Data

Links to sample distributions - improves discoverability without exposing sensitive data.

Conformance: a layered validation framework

Built on DCAT-AP prerequisites, refined with health-specific SHACL shapes

Layered SHACL validation

Non-public shapes

Mediated-access enforcement

Restricted shapes

Conditional-access rules

Public shapes

Open-access baseline

DCAT-AP prerequisites

Core structural foundation



What the framework enforces

- **SHACL shapes per access level**
Shapes tailored to non-public, restricted and public tiers.
- **Automated structural checks**
Validates structural constraints programmatically.
- **Controlled-vocabulary checks**
Ensures required vocabularies and code lists are used.
- **HDAB distributions required**
Non-public datasets must include HDAB-related distributions.
- **Precise governance metadata**
Guarantees governance properties are present and complete.

Takeways

HealthDCAT-AP in a nutshell

DCAT-AP v3.0

Extension, not fork

Reuses core classes - fully interoperable with the wider DCAT-AP ecosystem.

3 tiers

Access-level model

Non-public, restricted and public - each with tailored metadata requirements.

RDF Vocab

Governance by design

DPV, DQV, PROV-O, ODRL and CSVW integrated for machine-readable governance.

SHACL

Layered conformance

Per-tier shapes validate structure, vocabularies and mediated-access rules.

Resources@IHIS

HealthDCAT-AP

European Health Information Portal

HealthDCAT-AP editor – validator – literacy – dataset categorization tool, etc.

HealthDCAT-AP catalogue sandbox

Publication

Designing DCAT-AP Extensions for Common European Data Spaces: The EHDS HealthDCAT-AP Case Study
(NeXt-generation Data Governance workshop 2025 (NXDG 2025))

Coming publications

- Implementing a HealthDCAT-AP metadata management infrastructure for the European Health Data Space: A Sciensano Use Case. (Semantics in Data Spaces 2026)
- HealthDCAT-AP: a Linked Data application profile for describing health datasets in the EHDS (Manuscript in preparation)

Contact

ehds2@sciensano.be