



Funded by the
Nordic Council
of Ministers

Final metadata catalogues report

Report 2.8
Value from Nordic health data – VALO project

November 2025





Document info

0.1 Authors

Author	Partner
Mari Mäkinen	Finnish Institute for Health and Welfare (THL)
Zoltan Thinsz	National Board of Health and Welfare (NBHW), Sweden
Michael Peolsson	Swedish eHealth Agency
Per Wiklander	Swedish Research Council
Derya Akcan	Swedish eHealth Agency
Rebekka Björg Guðmundsdóttir	Directorate of Health, Iceland
Védís Helga Eiríksdóttir	Directorate of Health, Iceland
Hans Aage Huru	Norwegian Institute of Public Health (NIPH)
Truls Korsgaard	Norwegian Institute of Public Health (NIPH)
Minna Liikala	Finnish Institute for Health and Welfare (THL)
Peija Haaramo	Finnish Social and Health Data Permit Authority Findata
Martin Lund	Danish Health Data Authority
Commentators	
Michel Silvestri	Swedish eHealth Agency
Siri Larønningen	Norwegian Institute of Public Health (NIPH)

0.2 Keywords

Keywords	Health data, Nordic, EHDS, metadata, dataset catalogues
-----------------	---

Disclaimer

The content of this report represents the views of the author(s) and do not necessarily reflect the views of the experts interviewed.

Copyright Notice

Copyright © 2024 Sitra and the Nordic Council of Ministers. All rights reserved. For more information on the project, please see Sitra's website sitra.fi/en.



Contents

Executive summary	3
1 Introduction	6
2 History of the Nordic co-operation regarding national health data catalogues.....	6
3 HealthDCAT-AP metadata specification	7
4 National health data catalogues in the Nordic countries	8
4.1 Denmark.....	8
4.2 Finland	10
4.3 Iceland.....	12
4.4 Norway	13
4.5 Sweden	15
5 Subtasks and achievements of T2.6 Tracking progress on and supporting national metadata catalogues.....	17
5.1 Goals and organization of the work in T2.6.....	17
5.1.1 Monthly Health Metadata Network meetings	17
5.1.2 VALO Nordic EHDS2 Competence Forum Metadata break-out sessions	18
5.2 Subtasks of T2.6	18
5.3 Outcomes of T2.6.....	19
5.3.1 Subtask 1: Ensuring a shared understanding of HealthDCAT-AP	19
5.3.2 Subtask 2: Metadata data models vs. HealthDCAT-AP: Trying to find good and preferably shared ways to implement HealthDCAT-AP datasets model nationally.....	22
5.3.3 Subtask 3: Developing the shared Nordic variable level metadata specification	26
5.3.4 Subtask 4: Sharing information on progress in national metadata projects and EU projects.....	26
6 The way forward.....	29
6.1 Planned next steps.....	29
6.2 Identified long-term visions	30
Annex 1: Nordic common variable metadata specification v.0.8.....	0



Executive summary

This report describes the work done in the task 2.6 of the VALO project funded by the Nordic Council of Ministers. It describes the current situation of the national health data catalogues in the Nordic countries, the development plans, and the preliminary planning on how to implement HealthDCAT-AP metadata specification in the national health data catalogues. It also touches upon to the main challenges seen at the moment regarding HealthDCAT-AP specification and the future goals seen in the Nordic health metadata co-operation.

There has already been previous co-operation in the Nordic countries regarding metadata issues supporting secondary use of health data. This project builds on the previous work, and also the informal Nordic Health Metadata Network was incorporated into the task's work.

Every Nordic country has published metadata on key health data sources already for some time, and a national or semi-national health data catalogue exists in Finland, Norway and Sweden. Denmark is currently developing a national health data catalogue, Sweden is developing a prototype for a national catalogue building on the existing one and Finland is further developing the existing national health data catalogue.

HealthDCAT-AP is the metadata specification being developed for use in the European Health Data Space (EHDS2). The key challenges of HealthDCAT-AP draft from the Nordic perspective include lack of shared understanding on the information model, lack of hierarchy in the information model, ambiguity about how datasets should be defined and some technical complexity and terminology inconsistencies. For upcoming work it is proposed to continue building shared understanding of HealthDCAT-AP and looking for as harmonious way as possible to implement the HealthDCAT-AP in the national health data catalogues. Proposed solutions include creating possible enrichments of HealthDCAT-AP in the national catalogues in a harmonious way, including common Nordic variable metadata specification, implementing data quality and utility labelling in a harmonious way and aiming for harmonised dataset definitions.



Tiivistelmä

Tässä raportissa kuvataan tehtävän 2.6 työ VALO-hankkeessa, jota rahoittaa Pohjoismaiden ministerineuvosto. Raportti kuvaa Pohjoismaiden kansallisten terveysdatan katalogien nykytilaa, kehittämissuunnitelmia sekä alustavaa suunnittelua siitä, miten HealthDCAT-AP-metatietomäärittelyä voidaan toteuttaa kansallisissa terveysdatakatalogeissa. Raportissa käsitellään myös keskeisiä haasteita, jotka liittyvät HealthDCAT-AP-määrittelyyn, sekä tulevaisuuden tavoitteita pohjoismaisessa terveysdatan metatietoyhteistyössä.

Pohjoismaissa on jo aiemmin tehty yhteistyötä metatietokysymyksissä, jotka tukevat terveysdatan toissijaista käyttöä. Tämä hanke rakentuu aiemman työn pohjalta, ja myös epävirallinen Nordic Health Metadata Network otettiin mukaan tehtävän toteutukseen.

Jokainen Pohjoismaa on julkaissut metatietoa keskeisistä terveysdatan lähteistä jo jonkin aikaa, ja Suomessa, Norjassa ja Ruotsissa on olemassa kansallinen tai puolikansallinen terveysdatan katalogi. Tanska kehittää parhaillaan kansallista terveysdatan katalogia, Ruotsi kehittää prototyyppiä kansallisesta katalogista nykyisen pohjalta, ja Suomi jatkaa olemassa olevan kansallisen terveysdatan katalogin kehittämistä.

HealthDCAT-AP on metatietomäärittely, jota kehitetään käytettäväksi Eurooppalaisen terveystietoalueen (EHDS2) yhteydessä. HealthDCAT-AP-luonnoksen keskeiset haasteet Pohjoismaiden näkökulmasta liittyvät muun muassa yhteisen ymmärryksen puutteeseen tietomallista, hierarkian puuttumiseen tietomallissa, epäselvyyksiin siitä, miten tietoineistot tulisi määrittellä, sekä tekniseen monimutkaisuuteen ja terminologian epäjohdonmukaisuuksiin. Jatkoa varten ehdotetaan, että yhteistä ymmärrystä HealthDCAT-AP:stä rakennetaan edelleen ja etsitään mahdollisimman yhtenäisiä tapoja toteuttaa HealthDCAT-AP kansallisissa terveysdatan katalogeissa. Ehdotettuihin ratkaisuihin kuuluu muun muassa mahdollisten HealthDCAT-AP-laajennusten luominen kansallisiin katalogeihin yhtenäisellä tavalla, yhteisen pohjoismaisen muuttujametatietomäärittelyn sisällyttäminen, datan laadun ja hyödynnettävyyden merkintöjen toteuttaminen yhtenäisesti sekä harmonisoitujen tietoineistomääritelmien tavoittelu.



Sammanfattning

Denna rapport beskriver arbetet som utförts inom uppgift 2.6 i VALO-projektet, som finansieras av Nordiska ministerrådet. Rapporten beskriver den aktuella situationen för de nationella hälsodata-katalogerna i de nordiska länderna, utvecklingsplanerna samt den preliminära planeringen för hur metadata-specifikationen HealthDCAT-AP ska implementeras i de nationella hälsodata-katalogerna. Rapporten berör även de största utmaningarna som för närvarande identifierats kring HealthDCAT-AP-specifikationen samt de framtida målen för det nordiska samarbetet kring hälsometadata.

Det har redan tidigare funnits samarbete mellan de nordiska länderna kring metadatafrågor som stödjer sekundär användning av hälsodata. Detta projekt bygger vidare på det tidigare arbetet, och det informella nordiska nätverket för hälsometadata (Nordic Health Metadata Network) integrerades i uppgiftens arbete.

Varje nordiskt land har publicerat metadata om centrala hälsodata-källor under en längre tid, och en nationell eller semi-nationell hälsodata-katalog finns i Finland, Norge och Sverige. Danmark utvecklar för närvarande en nationell hälsodata-katalog, Sverige utvecklar en prototyp för en nationell katalog baserad på den befintliga, och Finland vidareutvecklar den befintliga nationella hälsodata-katalogen.

HealthDCAT-AP är metadata-specifikationen som utvecklas för användning inom European Health Data Space (EHDS2). De största utmaningarna med HealthDCAT-AP-utkastet ur ett nordiskt perspektiv inkluderar brist på gemensam förståelse av informationsmodellen, avsaknad av hierarki i informationsmodellen, oklarheter kring hur dataset ska definieras samt viss teknisk komplexitet och terminologiska inkonsekvenser. För det kommande arbetet föreslås att man fortsätter bygga en gemensam förståelse av HealthDCAT-AP och söker så harmoniska lösningar som möjligt för att implementera HealthDCAT-AP i de nationella hälsodata-katalogerna. Föreslagna lösningar inkluderar att skapa möjliga utökningar av HealthDCAT-AP i de nationella katalogerna på ett harmoniskt sätt, inklusive en gemensam nordisk metadata-specifikation för variabler, implementering av märkning för datakvalitet och användbarhet på ett harmoniserat sätt samt att sträva efter harmoniserade definitioner av dataset.

1 Introduction

This is the final report from VALO Work Stream 2 (WS2) Task 2.6 Tracking progress on and supporting national metadata catalogues of the Value from Nordic Health Data (VALO) Project.

The Nordic Council of Ministers funded the VALO project, which aim was to determine common Nordic principles for the implementation of the Regulation on the European Health Data Space (EHDS), and to investigate how to exploit the potential of Nordic cooperation in research, development and innovation related to secondary use of health data. The goal was to strengthen the Nordic countries' position as forerunners and their competitiveness, with regards to secondary use of health data.

The project partners in VALO were initially Finland, Iceland and Sweden, with Denmark and Norway acting as observers. However, Denmark and Norway became active members after the first Competence Forum in Helsinki and Estonia and Lithuania accepted the invitation to join the project as observers.

Task 2.6, related to metadata and data description matters in the Nordic countries, was part of WS2 in the VALO project. The main goal of WS2 was to establish and organize an EHDS2 Competence forum, where the Nordic countries meet to exchange information related to, and coordinate preparations for, the implementation of the EHDS. Part of coordinating the implementation of the EHDS was trying to find a congruent way to implement the EHDS HealthDCAT-AP metadata specification into national health data catalogues and to ensure a shared Nordic view on metadata solutions for secondary use of health data.

The aim of this report is to report on the work done during the project and show where the Nordic countries stand at the moment regarding national health data catalogue work and HealthDCAT-AP implementation in the national health data catalogues, what are the main challenges seen at the moment regarding HealthDCAT-AP and what kind of goals are seen in the Nordic health metadata co-operation in the future.

2 History of the Nordic co-operation regarding national health data catalogues

The Nordic countries already have quite a long history of co-operation on secondary use metadata issues. The VALO project was preceded by two Nordic Commons projects funded by the Nordic Council of Ministers, both of which had a task force on metadata issues.

The first Nordic Commons project took place between 2017 – 2019. The main output regarding metadata task force was to formulate a Nordic action plan on metadata. The report [A vision of a Nordic secure digital infrastructure for health data: The Nordic Commons](#) was published in 2019. It has as visions for metadata work to create a Nordic common metadata repository ecosystem for harvesting and consuming Nordic health data resources, having Nordic health data described with rich metadata according to the FAIR principles and establishing a Nordic health metadata catalogue.

As an annex to the report was a document with a table showing the metadata situation for different data categories or types in each Nordic country: Annex to the report “A vision of a



Nordic secure digital infrastructure for health data: The Nordic Commons” [Metadata Status for the Health Data Domain in Nordic Countries](#) (August 2018).

The second Nordic Commons project was planned to last from 2022 to 2024, but it was suspended in 2023. The main objectives in the Task force on Metadata, Data quality and Semantics were to prepare the functional, technical and semantic specifications and description of content for a Nordic health metadata ecosystem including a Nordic metadata catalogue and central services. The scope was both in standardized metadata about the data sources, the belonging variables, and the code lists/value sets the relevant variables are based on. During the project, the sharing of machine-readable metadata from endpoints was piloted and a common Nordic variable-level metadata specification was built up to draft version 0.7.

After the second Nordic Commons project was suspended in 2023, the Nordic health metadata experts who had worked together in these projects, decided to continue co-operation, learning together and discussion in an informal manner. The group decided to take its name Nordic Health Metadata Network and aimed for regular monthly meetings. The main goal of the group has been to share information and discuss related to national health metadata catalogues and EHDS and find synergies. Nordic Health Metadata Network is an open group where everyone working with health metadata matters in Nordic governmental agencies can join.

3 HealthDCAT-AP metadata specification

HealthDCAT Application Profile (HealthDCAT-AP) is a metadata specification being developed for use within the EHDS framework. It is an extension of the DCAT-AP specification (DCAT Application Profile for data portals in Europe), extending it with the additional elements needed to describe health datasets. There are already other domain-specific extensions (e.g. GeoDCAT-AP for geospatial data and StatDCAT-AP for statistical data) or national application profiles of DCAT-AP, and also new extensions are planned for other sector-specific European data spaces in the future. Read more about DCAT-AP: <https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/dcat-ap>.

HealthDCAT-AP aims to provide information on health data catalogues, datasets and data services across Europe. Its refinements to DCAT-AP include the introduction of new classes and properties, variable level metadata (called data dictionary in HealthDCAT-AP draft and provided by Sample distribution property), the use of controlled vocabularies for specific fields, and the definition of mandatory elements.

An unofficial draft version of HealthDCAT-AP was published in December 2023 as a recommendation. This draft was developed under the work package 6 of the HealthData@EU Pilot project. Further refinements to the specification were made as part of the work package 5 of the TEHDAS2 joint action and introduced in the [TEHDAS2 D5.1 Guideline for health data holders on data description](#). Recently, DG SANTE (The Directorate-General for Health and Food Safety of the European Commission) has taken ownership of the development of HealthDCAT-AP and a [European Commission Draft Specification of HealthDCAT-AP Release 5](#) was published in 22.9.2025. It is worth noting that the differences between HealthDCAT-AP Release 5 and the previous draft have not been thoroughly analysed and therefore any changes it may bring have not been taken into account in this report.



During the VALO project, the HealthDCAT-AP draft evolved quite a bit. When the VALO project was started, the HealthDCAT-AP draft at that time focused mainly on metadata of the datasets and did not describe the actual data level and variables at all. However, the variable level was added as a proposal in a later phase of the TEHDAS2 work, after many countries, including the Nordic countries, emphasized that describing the data at a high level was not enough and information on variables is needed also for the data application processes.

As part of HealthDCAT-AP, and thus the data descriptions, a dataset quality and utility label will be introduced, as specified in Article 78 in the EHDS regulation. This label is currently being developed in the EU-funded [QUANTUM project](#).

An implementing act, which will describe the final format in which metadata must be provided to the EU Central Dataset Catalogue, will be provided at the latest in March 2027. For secondary use, implementation will begin, after a transition period, in March 2029.

4 National health data catalogues in the Nordic countries

The aim of this section is to provide an overview of the situation in each Nordic country regarding the national health data catalogue. The information presented in the tables is partly based on the HealthData@EU Pilot output Milestone M6.1: Report on the landscape analysis of available metadata catalogues and the metadata standards in use (no longer public, the version found in the project's internal Google Drive was used).

4.1 Denmark

Table 1: Description of the project building Danish national health data catalogue

HDAB Denmark Metadata catalogue : ongoing project for building national health data catalogue	
URL	Project website: https://english.sundhedsdatastyrelsen.dk/health-data-and-registers/european-health-data-space/metadata-catalogue
Owner	Danish Health Data Authority
Description	<p>The EU-funded HDAB Denmark project is building a setup for a national health data catalogue and an application portal in Denmark as required in the EHDS regulation. The project is currently in the implementation phase.</p> <p>Variable descriptions are already implemented and currently used in application processes. The variable descriptions of all health registries are published in an Excel file and researchers pick variables from this file.</p> <p>Denmark also previously had a pilot version of a national health data catalogue (The Danish Meta Map), but it is no longer in use.</p>
Software	The solution is custom build, using a no-sql mongo database, and a set of API's build in c#.



Metadata standard	HealthDCAT-AP
Key functionalities	<p>Receive metadata about datasets from data holders and redistribute metadata to the EU Dataset catalogue as well as national information point (Datavejviser.dk).</p> <p>Browse and search metadata on both dataset- and variable level.</p>
Key properties	<p>In the MVP version of the dataset catalogue, all mandatory fields from HealthDCAT-AP draft, as well as selected recommended fields, are included. Currently, the following fields are implemented:</p> <ul style="list-style-type: none"> • On Dataset level: <i>Dataset ID, title, acronym, keywords, description, origin, purpose, geographical coverage, dataset theme, health theme, category, dataset type, applicable legislation, number of unique individuals, number of records, collection frequency, population coverage, collection period, publisher (name, website, email, type, note, trusted), contact point (name, website, email), distributions (title, url, rights, size, format, applicable legislation).</i> • On Table level: <i>Technical name, title, description, temporal.</i> • On variable level: <i>name, title, dataType, description, codeset, linkTable, seriesNote, variableNote, supportingElements, source, isPrimaryKey, yearFrom, yearTo, fieldLength, linkedTerms.</i>
Coverage of Art. 51 data categories (preliminary analysis)	Population-based health register data: The first version of the Danish dataset catalogue will focus on the national registries, which hold key national health data.
Languages	Danish
Status as national catalogue	Yes
Key differences compared to HealthDCAT-AP	It is based on HealthDCAT-AP draft. Metadata about datasets, tables and variables are included.
Development plans	<p>A MVP version of the Danish metadata catalogue to be operative by January 2026, when the project officially ends, and this MVP version is to be expanded the following year through January 2027, as the project has applied for an extension till then.</p> <p>The MVP is to be able to receive metadata from data holders and send it to the EU central services/platform and to the Danish register Datavejviser.dk, where metadata from various sectors are gathered.</p>



4.2 Finland

Table 2: Description of the Finnish national health data catalogue

Data Resources Catalogue (Aineistokatalogi)	
URL	https://aineistokatalogi.fi
Owner	Finnish Institute for Health and Welfare (THL) and Findata - Finnish Social and Health Data Permit Authority
Description	<p>Data resources catalogue (Aineistokatalogi) is the national health data catalogue in Finland. All data under the Act on the Secondary Use of Health and Social Data are described in the catalogue in a way specified in the Regulation of the Health and Social Data Authority: Data content, concepts and data structured for data descriptions. There are also some other data described in the catalogue, especially from the field of education.</p> <p>The catalogue is maintained in co-operation by the Finnish Institute for Health and Welfare (THL) and Findata - Finnish Social and Health Data Permit Authority. It was originally developed in co-operation between THL, Statistics Finland, Finnish Social Science Data Archive and Sitra in a project during years 2016 – 2018.</p> <p>Data holders make their own descriptions in the Data resources editor user interface and publish them. CSV-import for variable and code list descriptions is possible. Metadata descriptions play a key role in Findata's permit application process and the data extraction processes of data holders.</p> <p>Currently there are approximately 725 data resource descriptions by 42 different data holders.</p>
Software	<p>The Data resources catalogue and the Data resource editor are logically distinct information systems, but they are technically implemented as a single application. This application includes web interfaces and Representational State Transfer (REST) APIs for both systems.</p> <p>In the backend, the application utilises the Termieditori system, a web-based tool developed by THL for maintaining vocabularies and metadata. In practice, Termieditori functions as a graph database, which the application accesses via its REST API for data retrieval and storage. Additionally, Termieditori features a web interface used for maintaining the Data resource editor.</p>
Metadata standard	The Generic Statistical Information Model (GSIM) as information model. The metadata specification is mapped to DDI-L.
Key functionalities	<ul style="list-style-type: none">• Search data resources by name. Filter by the language of the metadata.• Filter/browse data resource descriptions by organisation.



	<ul style="list-style-type: none"> • Search variables by variable name, description, keyword or concept variable. Filter by language of the description or variable group. • Download the variable and code list descriptions in a CSV file (Excel).
Key properties	<p>In Data resource and Dataset level: <i>name, alternative name, abbreviation, description, organisation, contact information, data resource type (ex: registry data, survey data), terms of use (controlled list), terms of use, further information (free text), links, observation unit type, number of observation units, universe/population, reference date (reference period start date, reference period end date), regional coverage, target group, sample size, non-response, data resource life cycle phase, keywords, free keywords, series.</i></p> <p>In variable level: <i>Name, technical name, description, concept variable, variable group, keywords, free keywords, reference date (reference period start date, reference period end date), domain type (continuous/categorical), label for missing value, code for missing value, information about data quality, variable data source, data type, data format, observation unit type, variable related instance question, code list (references to external code lists: external identifier, name, description / self-defined code lists: name, description, owner, codes: code, value).</i></p>
Coverage of Art. 51 data categories (preliminary analysis)	<ul style="list-style-type: none"> • EHR data • Socio-economic, environmental and behavioral determinants of health • (Aggregated healthcare data) • Pathogen data • Healthcare administrative data • (Genetics, epigenetics and genomics) • (Human molecular data) • Population-based health register data • Medical and mortality register data • Other health data medical devices • Registries medicinal products & medical devices • Research data • (Biobank data)
Languages	FI, SV, EN
Status as national catalogue	Yes
Key differences compared to HealthDCAT-AP	<p>The information model differences: In Data resources catalogue data resources, datasets and variables & code lists are described. In HealthDCAT-AP datasets, distributions and also variables and code lists as a distribution are described.</p> <p>The differences in attributes: Quite many need to be changed/added as new.</p>



	See more: nHDsC: Requirements and Specifications in the EU Funding & Tenders Portal.
Development plans	Currently ongoing development under EU-funded FinHITS Direct grant project: Usability and search function development for the catalogue: search functions, user interface improvements etc. Development of the application portal and metadata integration. Later: Development to be HealthDCAT-AP compliant.

4.3 Iceland

Table 3: Description of the metadata of Icelandic health registries

Data descriptions including variable descriptions of health registers in a webpage	
URL	https://island.is/en/gagnasofn-embattis-landlaeknis/eydublod-breytulistar
Owner	Directorate of Health
Description	<i>This description only applies to the data that consists of publicly available variable descriptions in Iceland. Most of this data is maintained by the Directorate of Health.</i> Description of registers including variable description in Excel files at webpage (HTML). Very similar to a metadata catalogue but no DCAT-AP compliant technology behind. The variable descriptions are used in data applications.
Software	Description of registers including variable description in Excel files at webpage (HTML).
Metadata standard	None
Key functionalities	Metadata on register level in a webpage. Link to Excel file with variable descriptions.
Key properties	In dataset level: <i>Title, responsible party, processor, purpose, content, period, source of data, registration details, comparable or related datasets, processing and publication, history.</i> List of variables available as downloads (xls files).
Coverage of Art. 51 data categories (preliminary analysis)	<ul style="list-style-type: none"> • (Socio-economic, environmental and behavioral determinants of health) • (Aggregated healthcare data) • Pathogen data • Healthcare administrative data • Population-based health register data



	<ul style="list-style-type: none"> • Medical and mortality register data • Research data
Languages	IS
Status as national catalogue	Non-existent
Key differences compared to HealthDCAT-AP	<p>In-depth analysis has not yet been done, but there are some similarities in the dataset level properties.</p> <p>Icelandic team has piloted HealthDCAT-AP editor with registry descriptions.</p>
Development plans	Planning on applying for a direct grant for starting to plan a national health data catalogue.

4.4 Norway

Table 4: Description of the Norwegian national health data catalogue

Norwegian health metadata catalogue (Helsedata.no)	
URL	https://helsedata.no/en/
Owner	Norwegian Institute of Public Health (NIPH)
Description	<p>National health data catalogue with data source, variable and code list descriptions.</p> <p>Structure: Data sources with drill down to variable lists with quite rich metadata, code lists/value sets and statistics.</p> <p>Services: data source explorer, variable explorer, overview open data, create your own variable list, application service/form.</p> <p>Content: Administrative registries, national health registries, national medical quality registries, national and regional health studies and other registries.</p> <p>Metadata shared with:</p> <ul style="list-style-type: none"> • Norwegian Cross Sectorial Metadata Catalogue, example: https://data.norge.no/datasets/caf3d50c-1d5c-3909-860b-0f491b24177b • The official portal for European data, example: https://data.europa.eu/data/datasets/https-helsedata-no-no-forvaltere-kreftregisteret-kreftregisteret-?locale=en <p>Variable descriptions are defined in health data applications, but it is not yet mandatory.</p>
Software	Off-the-shelf software Rhapsody (previously HealthTerm). The software is soon to be replaced by an in-house build software.



	Helsedata.no, where the health data sources and variables are displayed, is built mainly on Optimizely CMS system.
Metadata standard	<p>Dedicated metadata schema National metadata specification for Health Data sources for secondary purposes (Nasjonal spesifikasjon for metadata om helsedata).</p> <p>The health metadata content in Rhapsody are taken out from Rhapsody by its API and converted programmatically to the DCAT-AP standard and then shared with the national dataset catalogue, which again share it with the European data catalogue.</p>
Key functionalities	<p>Search for data sources and variables by name and all other reported properties by a standard internet search engine.</p> <p>Filters in data source level: <i>Access level, type of data source, category, data holders, data source.</i></p> <p>Filters in variable level: <i>Type of data source, data source, variable groups, category, different variable themes/categories, temporal coverage.</i></p> <p>“Drill down” to detailed metadata on variable level. In the variable explorer you can create variable lists, share them, and download them and use them in applications.</p>
Key properties	Data source, sub sources and data records: <i>title, publisher, theme, temporal coverage, variables, open data (description), criteria for data access, apply for access to data, contact, landing page.</i>
Coverage of Art. 51 data categories (preliminary analysis)	<ul style="list-style-type: none"> • Socio-economic, environmental and behavioral determinants of health • (Aggregated healthcare data) • Healthcare administrative data • Population-based health register data • Medical and mortality register data • (Clinical study data) • Registries medicinal products & medical devices • Research data • (Biobank data)
Languages	NO, EN
Status as national catalogue	Yes
Key differences compared to HealthDCAT-AP	<p>Metadata is already mapped to DCAT-AP and can be distributed in that format.</p> <p>The national standard has a defined property set for variables, whilst the HealthDCAT-AP has a looser definition at the moment.</p>



	HealthDCAT-AP is an open internet standard, while the national standard is proprietary.
Development plans	Extend the national standard to include biobanks, aggregated data etc. Map and possibly align the national standard to HealthDCAT-AP.

4.5 Sweden

Table 5: Description of the Swedish existing health data catalogue and the project building a prototype of a national health data catalogue

Ongoing SENASH project & already existing RUT (Register Utiliser Tool)	
URL	<p>SENASH: https://www.ehalsomyndigheten.se/languages/english/senash/</p> <p>RUT: https://rut.registerforskning.se/</p>
Owner	<p>SENASH is an EU-funded project. The Swedish E-health Agency is the Competent Authority.</p> <p>RUT: Swedish Research Council</p>
Description	<p>SENASH:</p> <ul style="list-style-type: none"> • The <i>Sweden National services for Access to Swedish Health Data</i> project is building a MVP version of a national health dataset catalogue and an application system. • Building on solutions and knowledge from the existing RUT Metadata system, the initiative brings together five Swedish government agencies in close collaboration. • The forthcoming phase of the project will evaluate whether the MVP constitutes the most appropriate solution for establishing a future national health dataset catalogue. <p>RUT: RUT is a metadata system for register-based research. It provides a quick overview of the content of various register datasets and offers tools tailored for both data holders and researchers.</p> <p>Structure:</p> <ul style="list-style-type: none"> • RUT Metadata Administration – a tool for data holders to create and manage metadata, • RUT Metadata Explorer – a tool for researchers providing rich metadata and semantic descriptions, • RUT Metadata Catalogue – a public website giving researchers an overview and listing of available register datasets.
Software	In house development built upon open-source components.



Metadata standard	SENASH: GSIM and HealthDCAT-AP RUT: GSIM
Key functionalities	RUT enables advanced searches across register datasets and allows users to compare variables from multiple perspectives. Access requires authentication: accounts can be created only for researchers affiliated with recognized research organizations and for staff at data-holding authorities. While RUT provides rich metadata and search functionality, SENASH expands on this by enabling applicants to use the metadata to specify exactly which subset of a dataset they want to apply for, and to submit their application to the correct data holder.
Key properties	RUT: <ul style="list-style-type: none"> In the metadata catalogue overview https://rut.registerforskning.se/metadatakatalog/register/698318b7-fada-4543-9751-85c65f7d9b4b/: <i>Title, record keeping organisation, website, description, purpose, type of registry, connection status, data collection period, metadata in RUT (timestamp).</i> In the metadata tool (which requires log-in): <i>Metadata on variable level structured according to GSIM; register structure according to the GSIM model, variable name, variable definition, (statistical) object type, population covered by the variable, code list, value domain, measurement scale, period of measurement etc.</i>
Coverage of Art. 51 data categories (preliminary analysis)	<ul style="list-style-type: none"> Socio-economic, environmental and behavioral determinants of health (Aggregated healthcare data) Healthcare administrative data Population-based health register data Medical and mortality register data (Clinical study data) Registries medicinal products & medical devices Research data (Biobank data)
Languages	SV, EN
Status as national catalogue	RUT: "Semi-national"
Key differences compared to HealthDCAT-AP	The SENASH project follows the development of HealthDCAT-AP and aligns with EHDS requirements, while adapting the model to the needs of Swedish health data holders. Compared to HealthDCAT-AP, SENASH extends the model by integrating GSIM concepts, enabling a more precise semantic description of health data. Specifically, SENASH allows to describe:



	<ul style="list-style-type: none"> • What has been measured – by semantically specifying variables and the units they apply to, • How it has been measured – by linking variables to value domains and, where relevant, to code lists and statistical classifications, • When and where it has been measured – by relating variables to populations that delimit the observed units in time and geography. <p>The project’s working hypothesis is that this level of detail is essential for accurately processing data access requests and for ensuring that applicants are granted access to exactly the data they need, while still maintaining interoperability with the European HealthDCAT-AP standard.</p>
Development plans	SENASH is under development, and the final report will be delivered in January 2027.

5 Subtasks and achievements of T2.6 Tracking progress on and supporting national metadata catalogues

5.1 Goals and organization of the work in T2.6

The task 2.6 description has been formulated in the VALO project plan as follows: Engage with the informal Nordic Health Metadata Network, focus on metadata data models vs. DCAT-AP Health Extension. Contribute updates on metadata catalogues related work and projects to T2.1 (on progress report template) and T2.5 (on horizon scanning on future projects) and respective outputs. In terms of the output, it has been formulated: Analysis of metadata data model options and alternative pathways for transition to DCAT-AP Health extension adoption in national metadata catalogues.

The task 2.6 work was promoted through meetings of the Nordic Health Metadata Network and metadata break-out sessions in the VALO Nordic EHDS2 Competence Forums.

5.1.1 Monthly Health Metadata Network meetings

From the end of 2024, the Nordic Health Metadata Network's activities were integrated into the VALO project, and much of the metadata work of the VALO project task 2.6 was carried out through the Network's activities. This also provided more continuity for the Network's work, as a voluntary approach without funding is seen to be more difficult to maintain. Consequently, some people outside the VALO project also participated in the discussions and information exchange in the Nordic Health Metadata Network.

All the meetings were held online, and the Network also had its own workspace set up in THL's Teams environment. In addition, meeting materials and notes were also shared in the VALO project's Teams channel. All together 12 Network meetings were held during the VALO project, two of them being joint meetings between the Network and the Nordic DCAT-AP experts.



The Network members state that the Network and VALO EHDS2 Competence Forum Metadata break-out sessions have been important forums, especially because not all participants are involved in the EU-level Health Data Access Bodies (HDAB) Community of Practice (CoP) metadata subgroup. It has been seen that the Network and VALO project have created kind of a Nordic equivalent to the EU-level CoP.

5.1.2 Nordic EHDS2 Competence Forum Metadata break-out sessions

The VALO project's quarterly Nordic EHDS2 Competence Forums were also a key part of promoting the metadata work. The work in T2.6 was kicked off in the first EHDS2 Competence Forum in Helsinki in September 2024 and all subsequent forums had their own break-out sessions dedicated for metadata matters.

When this report is being drafted, four Nordic EHDS2 Competence Forums have been arranged: in Finland (Helsinki), in Denmark (Copenhagen), in Sweden (Gothenburg) and in Iceland (Reykjavik). A fifth and final Competence Forum will be organised in Finland (Helsinki) by the very end of the project. Presentations and notes from the sessions can be found in the internal VALO Teams channel.

5.2 Subtasks of T2.6

Four unofficial subtasks were formulated within the Task 2.6. The subtasks and the related planned activities are listed below.

Table 6: Unofficial subtasks and planned activities of T2.6

Subtask	Planned activities
<p>Subtask 1: Ensuring a shared understanding of HealthDCAT-AP</p>	<ul style="list-style-type: none"> • Getting to know HealthDCAT-AP. <ul style="list-style-type: none"> ○ Familiarization with the HealthDCAT-AP draft. ○ Discussions about how to interpret different elements, sharing questions to the group. ○ Review and giving shared feedback of HealthData@EU Pilot D6.2 on HealthDCAT-AP. ○ Also, possibility to participate in some of the Joint Action TEHDAS2 Task 5.1 Technical working group discussions. • Piloting and testing HealthDCAT-AP with the HealthDCAT-AP editor produced in HealthData@EU Pilot WP6 by Sciensano. <ul style="list-style-type: none"> ○ Participating in TEHDAS2 WP5 rolling campaign for different stakeholders to make data descriptions. ○ Sharing experiences and questions inside the group. • Giving feedback on HealthDCAT-AP. <ul style="list-style-type: none"> ○ Feedback to the European Commission by testing the releases of HealthData@EU Central Platform EU Dataset Catalogue and giving feedback with the EU form.



	<ul style="list-style-type: none"> ○ Public consultation of the TEHDAS2 metadata deliverables in February 2025. ○ Feedback to TEHDAS2 through the technical working groups of HealthDCAT-AP, either by participating in discussion or via the Network members who participate in the technical working group.
Subtask 2: Metadata data models vs. HealthDCAT-AP: Trying to find good and preferably shared ways to implement HealthDCAT-AP datasets model nationally (the choice of metadata data model)	<ul style="list-style-type: none"> ● Mapping the national metadata data models to HealthDCAT-AP draft. ● Sharing and discussing about the mappings in the Nordic Health Metadata Network. ● Discussion on the best practices on how to define the datasets. This has implications for the data provided for secondary use because the data in datasets are applied for secondary use in EHDS.
Subtask 3: Developing the shared Nordic variable level metadata specification	<ul style="list-style-type: none"> ● Developing further and testing the common variable level specification developed in the Nordic Commons project. ● Following closely the discussions and possible development around HealthDCAT-AP variable description model. The QUANTUM work also needs to be followed regarding the discussions on variable descriptions.
Subtask 4: Sharing information on progress in national metadata projects and EU projects	<ul style="list-style-type: none"> ● Sharing information on national projects. ● Sharing information on EU projects TEHDAS2, QUANTUM, HDAB Community of Practice etc.

5.3 Outcomes of T2.6

5.3.1 Subtask 1: Ensuring a shared understanding of HealthDCAT-AP

The development of a shared understanding of the HealthDCAT-AP draft specification had started during the project. The task 2.6 and Network members participated in TEHDAS2 WP5 Technical working group on developing HealthDCAT-AP, shared related information for others, read HealthData@EU Pilot project's D6.2 on HealthDCAT-AP and shared comments on TEHDAS2 Milestone 5.1 Guideline for data holders on data description and Milestone 5.3 Technical specification on the national metadata catalogue. Many countries, also countries like Iceland who didn't participate in the TEHDAS2 work, also piloted HealthDCAT-AP descriptions with HealthDCAT-AP editor provided by Sciensano.

Members from the task group also actively participated in developing HealthDCAT-AP by providing Sciensano, the organisation leading HealthDCAT-AP development work in TEHDAS2 WP5, with the Nordic variable metadata specification v.0.8. The Nordic common variable metadata specification development started in the Nordic Commons project and was further developed during the VALO project. It contains a Nordic compromise view on the most important metadata properties to describe variables of a dataset. The Nordic suggestion seems to be well considered in the deliverable 5.1 [Guideline for data holders on data](#)



[description](#) of the TEHDAS2 (see page 83, Table 27: Feedback from the technical working group on metadata attributes essential and important to be included in a CSVW data dictionary).

In addition to the above, an important collaboration was started with Nordic DCAT-AP experts in spring 2025. Some countries had already established communication inside the country regarding HealthDCAT-AP draft with the national DCAT-AP experts before the meeting, but some had not. The plan is to continue this collaboration with regular meetings. The aim of the co-operation is to learn from the already existing DCAT-AP based national catalogues and their experiences in application of the metadata specification, get valuable feedback of HealthDCAT-AP draft specification and share experiences and learn from each other regarding metadata catalogues.

The challenge with the HealthDCAT-AP related work has been that the draft specification has changed during the VALO project period, and it most likely will continue to evolve also in the future. At the time of writing, the TEHDAS2 task on developing HealthDCAT-AP draft is finished and it is a bit unclear how the further development of HealthDCAT-AP will take place and what kind of participating opportunities there will be. It is now under the authority of DG SANTE and also the latest draft version was published by DG SANTE during the time of writing (see: <https://healthdataeu.pages.code.europa.eu/healthdcat-ap/releases/release-5/>).

Reflections on HealthDCAT-AP draft: identified challenges and areas requiring clarification

The following is a list of issues that have been raised in the Nordic discussions and work related to HealthDCAT-AP draft. The list is a working list in nature, and the work will continue.

Information model understanding

- Lack of shared understanding across the community on the HealthDCAT-AP information model.
- HealthData@EU Pilot and TEHDAS2 work on HealthDCAT-AP focused primarily on the Dataset class. There is a lack of shared understanding on different classes in the information model.
 - What actually is Catalog, what is Data service? This all needs to be clear when setting up national catalogues.
 - Especially in countries that have already mapped their health metadata to DCAT-AP there may be previous views and practices on the matter.
 - There is only Dataset level in the information model. Upper hierarchical level for registry/data resource/data collection is missing. It is unclear how the Dataset series property can be utilized and if it could provide the needed higher hierarchical level above a Dataset.
- How to define Dataset is unclear. Some of the suggestions presented in HealthData@EU Pilot D6.2 and other work do not seem to be sustainable and functional for health data holders and users, as they would break up data that are currently described as a single entity into several different datasets.
- HealthDCAT-AP only takes into account making data findable, not data management (as GSIM model does).



Terminology and naming issues

- **Sample Distribution:** Misleading attribute name - primary distribution should be named "Data dictionary" instead, as it is primarily required instead of sample/synthetic data.
- **Data dictionary:** "Variable description" could be a more understandable term and should be used instead of or in addition to "Data dictionary" for better clarity.
- **Type:** Confusing attribute name as it actually refers to Dataset type.
- **Health Category:** Confusing attribute name as it actually refers to Health data category or EHDS2 data category (EHDS Art. 51).

Technical implementation and other challenges

- **Too many mandatory properties:** Too many required fields creating burden for data holders.
- **Too many different profiles:** When using Sciensano's HealthDCAT-AP editor, it has proven to be a bit confusing for the user to know which profile to select: non-public, restricted or public. The profiles are related to different legislation and have different mandatory fields. It was also commented by the Nordic DCAT-AP experts that there are too many profiles.
- **Wikidata vocabulary:** Wikidata, a free linked database of secondary data that can be read and edited by both humans and machines, is proposed to be used as an ontology in HealthDCAT-AP. Questions have been raised about governance, quality assurance, metadata editor integration, and selection accuracy, as Wikidata has multiple entries for the same concepts.
- **Vocabulary development:** Multiple vocabularies need further development.
- **Suitability for dynamic data:** Many attributes are currently unsuitable for daily updated data (Number of unique individuals, Number of records, Modified date).
- **Has Coding Systems:** No connection between Dataset-level Has Coding Systems and variable code list descriptions.
- **Has Code Values:** Unclear relationship between coding systems and code values. However, the situation has improved with the TEHDAS2 output D5.1: This property is for semantic annotation of the dataset. If this is also the ultimate understanding of the purpose of this property, the name "has code values" seems a bit misleading.
- **Distribution usage:** Does it currently align properly with DCAT-AP standard? Are the distributions sufficiently understandable for the health data holders describing their data and the catalogue users?
- **CSVW for variable descriptions:** CSV (Comma-Separated Values) on the Web (CSVW) standard is suggested for variable descriptions (called data dictionary in the HealthDCAT-AP draft). No mandatory properties are defined yet for CSVW. The properties suggested in public consultation and TEHDAS2 technical working group discussions are provided in the TEHDAS2 D5.1. However, the possibilities of CSVW are not yet sufficiently known. Further examination, clarification and testing are needed, especially from the technical side on how these files should be produced.

HealthDCAT-AP implementation considerations

- The current understanding is that EHDS does not per se require that national health data catalogues must have HealthDCAT-AP implemented. National contact points for secondary use (NCP) can be responsible for transforming the metadata into HealthDCAT-AP format for the EU Central Dataset catalogue. This might result to different approaches in each country.



- The current view in most of the Nordic countries seems to be that they want to implement HealthDCAT-AP in the national health data catalogue, while also retaining other properties that have already been found to work. The approach Norway is currently planning is to start by just transforming the metadata they already have to HealthDCAT-AP and not ask the data holders to change the descriptions. They hope it can be done using Sciensano's HealthDCAT-AP editor or another tool. When revising the national metadata specification, new fields would be incorporated to make it HealthDCAT-AP compliant. This might be a way forward in the beginning for other Nordic countries as well.
- When considering the information/data model approaches, some countries currently consider the possibility of keeping the already existing data model in the national catalogue. An important point to consider here is to think how the data descriptions will look in the national catalogues compared to the EU Dataset Catalogue at HealthData@EU Central Platform – if they would be too different, it can cause misunderstandings.

5.3.2 Subtask 2: Metadata data models vs. HealthDCAT-AP: Trying to find good and preferably shared ways to implement HealthDCAT-AP datasets model nationally

The key descriptive element in the HealthDCAT-AP specification is dataset. Datasets can have different kinds of relationships to other datasets, but currently there is no upper binding element above the dataset in the information model. This is a key difference compared to the information/data models in many current Nordic health data catalogues. For example, Norway and Finland currently have a metadata information model which constitutes from data resource/registry/data collection level and datasets under them. Also, Sweden has been looking at DCAT element Dataset series as a possibility to provide the upper level to link datasets related to the same data resource when building a prototype of national health data catalogue in the SENASH project.

Also, how the datasets should be divided in the catalogue and thus in which kind of entities the data should be offered for secondary use, is not yet very clear. It is also not yet known whether the Commission will give guidance on this, or whether the decision will be ultimately left to the data holders to make when they describe their data and provide a data quality and utility label.

During the VALO project, there was a lot of discussion on the topic and consideration was given to dataset division with a few different data categories and from different perspectives. The views on the matter are preliminary and discussion and work on this will continue. The possible continued discussion around and development of HealthDCAT-AP can greatly influence the views.

Current Dataset definitions landscape

The following is a compilation of definitions for the dataset mentioned in various sources related to EHDS.

- **DCAT-AP & HealthDCAT-AP:** Dataset is a conceptual entity that represents the information published.
- **EHDS Regulation:** Dataset is a structured collection of electronic health data.
- **TEHDAS2 D5.1 Specifications** (p. 66):



- **dcat:Dataset**: This class describes a coherent collection of data, whether numerical, textual, visual, or otherwise. It could correspond to a database, a file, a report, or another structured data asset intended for use or reuse.
- **dcat:DatasetSeries**: This class is used for datasets that form part of a logical sequence, such as regular statistical releases or time-based collections. It enables grouping datasets with shared themes or structures along a recurring dimension (e.g. time, geography).
- **TEHDAS2 D5.1** (p. 94-95):
 - Two guiding considerations should inform this process (of structuring data in the catalogue):
 - Dataset granularity and discoverability:
 - Dataset descriptions should be created at a level that is meaningful for potential users. For example, in the case of cohort datasets with multiple collection events or subpopulations, it is often more useful to define datasets at the level of a collection event.
 - A practical principle is to organize datasets around a common aspect, such as population, collection period, or data collection method, and then use HealthDCAT-AP's relation properties to link related datasets and provide an overview of their structure.
 - Relations in the current HealthDCAT-AP draft version (source TEHDAS2 D5.1, p. 66-67):
 - Source: Identify dataset/document from which current dataset was derived,
 - In series: Specify dataset belongs to series/collection (recurring publications, periodic releases, thematic collections),
 - Qualified relation: Describe specific role relationships ("updates," "complements," "is replaced by"),
 - General relation: Indicate general link when more specific properties do not apply.
- **QUANTUM project definition** (source: QUANTUM Data Quality & Utility Labelling Tool, only available after login):
 - Dataset Definition: A Dataset is a collection of data, published or curated by a single source, and available for access or download in one or more formats.
 - Datasets Examples:
 - (1) A complex dataset: A set of three databases related to a specific cohort of cancer patients:
 - A relational database of the patients' medical history;
 - A set of the patients' cancerous cell images;
 - The logs of the microscope used to generate these images at the molecular level.
 - These three databases can be used together via shared identifiers (patient identifier in medical history & cancerous cell images, images belonging to batches whose identifier can be found in both images & microscope logs).
 - (2) A simple dataset: database, a CSV file, etc.
- **Nordic DCAT-AP experts' preliminary guidance**: Dataset should be maintained by a single agent; data holders typically know best how to present their content based on their experience.



Identified guiding principles in dataset definition from the Nordic perspective

- **Taking user friendliness as a guiding principle.** Secondly, the division also has to be reasonable from data holder's perspective (not too finely divided/laborious, not too broad/managing huge number of variables in one dataset) and also from the HDAB perspective (when receiving the application, it has to make sense to them).
- **Having only datasets in the information model might not be enough.** How can you get an overview of the whole data resource if the underlying metadata model does not support hierarchy? The intention from the Nordic health metadata community is to try to affect HealthDCAT-AP development to better support hierarchy, if possible.
- **Consider a multi-stakeholder perspective:**
 - Identified considerations from health data user perspective
 - How do users perceive and search for data?
 - How to ensure discoverability of all needed information in catalogue? If for example one national registry is divided into several datasets, how to make sure users do not miss needed information? In the Nordic countries national registers are a fairly established concept (and the guidance so far from Sciensano has been that they should not be described as one dataset). The Nordic priority is to ensure that users will continue to find the data they need in clearly understandable entities.
 - Variable-level understanding vs. high-level descriptions.
 - User-friendly access and comprehension requirements.
 - What is a useful entity when considering data access applications?
 - Identified considerations from health data holder perspective
 - How do data holders perceive and manage data in their internal processes?
 - How does it integrate to possible data warehouses and other technical environments?
 - What divisions are not too laborious for implementation? Both HealthDCAT-AP descriptions, including variable descriptions, and separate data quality and utility label evaluation and quality reporting will be implemented from datasets.
 - Possible existing definitions shaped by previous legislation must be considered.
 - Balance needed between meaningful segmentation and operational feasibility.
 - Clear boundaries needed between data at different stages in the data holder's internal processing: final datasets provided and described for secondary use.
 - Identified considerations from HDAB's perspective
 - Is the entity of data described understandable for HDAB evaluating the data access and data request applications?
 - Identified possible considerations from IT system sustainability perspective
 - Are there concerns about managing dataset descriptions with large numbers of variable descriptions?
 - Are there other technical limitations and resource constraints?
 - How does it integrate with existing infrastructure and workflows?
- Other considerations
 - **Consider hierarchical level impact:** Does the possibility of including higher hierarchical level (data collection/resource level) change the dataset definition approach?

- Potential use of `dc:DatasetSeries` class for logical grouping.
 - Balance between granular detail and overview structure.
- **Take into account different EHDS Art. 51 data categories**
 - The work so far in the VALO project has been around the data categories more familiar to the group that are currently described in the Nordic catalogues, mainly national health registry data and research data.
- **Data quality and utility label (a suggestion currently being developed in QUANTUM project) integration:** Alignment with HealthDCAT-AP – the data quality and utility label with links to quality reporting will be given to most of the datasets described at dataset level. The choice of dataset definition changes the data quality evaluation result.
- **Consider the possibility of providing user friendly data products** – rethinking the datasets?
 - Most used variables: might be risky as it might steer the users too much?
 - How to define data products?
- In the end it's always the person describing their data who decides, we can only guide (unless there will be more specific guidance from the Commission).

Current preliminary views on dataset division

The view presented by Sciensano, which has led the HealthDCAT-AP work in HealthData@EU Pilot project and TEHDAS2, has been that *dataset should have unique temporal and geographical aspects*. This approach is understandable from the clarity perspective, but it does not seem workable as such. For example, for the Swedish National patient registry, it would mean dividing the data into approximately 18 different datasets based on analysis by experts from the National Board of Health and Welfare (NBHW).

It is quite common for many national registries that the geographical coverage is not the same throughout the registry and also the temporal aspects might differ, as some data has been collected from some areas earlier or later than others for example for piloting or other reasons. Based on the analysis by Swedish experts, following Sciensano's suggestion would mean that when a researcher would want to select for instance all inpatient data for region Uppsala between 1964-2008 in the catalogue, they would need to select eleven different datasets for the application. That would not be customer friendly when the description in the catalogue would be too fragmented and also it would set many requirements for the Data Access Application Management Systems (DAAMS).

Based on their preliminary analysis, the experts from NBHW would either divide the national patient registry into three datasets (Inpatient care data, Outpatient care data, and Compulsory care data) or divide based on population-centric view to five datasets (Inpatient care, Day surgery, Specialist outpatient visits, Outpatient psychiatric visits with other medical professions than doctors, and Compulsory care). These datasets would have individual lists of variables, but also same kind of variables could be found in all the datasets.

In Norway the preliminary identified starting point is to take a practical approach to EHDS implementation. It would mean sharing metadata on the current data sources/registries (as described in the national catalogue currently) as datasets in the EHDS catalogue. It would mean that data holders would not be asked to adapt to anything new. The only change would be to eventually develop the existing metadata systems by adding mandatory HealthDCAT-AP properties. This could be a practical start and by experimenting they would see if the Commission will start to steer them in other direction. CSV-files could be shared as they are shared already. On later stage they would be looking at constructing data products to be shared as datasets.



In Finland the team has looked at different ways to divide the current data resource and dataset descriptions into datasets. Different approaches have been looked at, and some discussions have also been held with data owners. The problem currently is that the data holders have decided quite freely how to define a data resource and datasets in the national health data catalogue. Therefore, very varied solutions are currently seen in the catalogue, for example Health care and Social care data resource division with hundreds of datasets under them being common for the data from wellbeing services counties.

Lately the thinking in the Finnish team has started to steer towards the same as in Norway – the EHDS dataset being closer to current data resource than dataset in Finnish Data resources catalogue information model. Dataset would then be serving as a systematic resource containing the data on the registered subjects required for a specific purpose, as the definition of data resource in the current Findata's regulation on data description is. If the dataset would be too big with too many variables, then it would be considered practically how to divide it. However, this view might be affected when knowing how DCAT:DatasetSeries can be applied in HealthDCAT-AP and how it can be presented in the catalogue for users.

Iceland does not have a health data catalogue yet, but their data descriptions on the website, including Excel sheets with variable descriptions, are already quite similar as a catalogue. Icelandic team sees that the current way of presenting registries as one entity has worked well, but also further discussions are needed.

5.3.3 Subtask 3: Developing the shared Nordic variable level metadata specification

When the work on the VALO project began, the HealthDCAT-AP draft did not yet include variable-level metadata. However, all Nordic countries recognize that variable-level description is an essential part of metadata. It has also been recognized that it would be important for variable descriptions to be interoperable across the Nordic countries.

The development of a common Nordic variable metadata specification started already in the Nordic Commons project. During the VALO project, the draft version was developed further until version 0.8. The specification was delivered to Sciensano, the organisation responsible of HealthDCAT-AP draft development in TEHDAS2 project, and the Nordic suggestion of variable attributes was well received in the TEHDAS2 deliverable D5.1 on data descriptions.

The goal is to continue working on the Nordic specification in the VALO2 project and to actively participate in any further development of HealthDCAT-AP. The revised version 0.8 is attached to this report (Annex 1).

5.3.4 Subtask 4: Sharing information on progress in national metadata projects and EU projects

Information about different international projects has been actively shared at every meeting. Discussions were held especially on metadata related aspects in HealthData@EU Pilot project, TEHDAS2, QUANTUM and HDAB Community of Practice metadata subgroup.

Direct Grant projects from the EU4Health program have played a key role in the development of metadata catalogues at the national level. The following is a brief description of the current Nordic projects and their key metadata development goals and achievements.

Metadata catalogue development in Direct Grant projects

Denmark: HDAB Denmark (Danish Health Data Authority)

- Duration: February 2024 – January 2026 (applied for an extension till January 2027).
- Metadata related goals: To map which categories of health data, i.e. metadata, are available in Denmark and compile them into a digital metadata catalogue. The purpose is to provide a better overview of existing health data for research use in the EU.
- Outcomes so far: Currently working closely with a small selection of data holders to include their first datasets in the MVP version. Formats, rdf, for example .ttl of json-ld are being tested.
- Further information: <https://english.sundhedsdatastyrelsen.dk/health-data-and-registers/european-health-data-space/metadata-catalogue>

Finland: FinHITS – Strengthening Finnish Health Data ICT for Secondary Use (Findata)

- Duration: November 2023 – October 2027.
- Metadata related goals:
 - Usability improvements on Data resources catalogue (search function and other usability improvements).
 - Variable descriptions integration into the data access application management system (DAAMS) and applications.
 - Plan for spring 2026: More detailed planning on steps needed on implementing HealthDCAT-AP in the already existing national Data resources catalogue.
- Outcomes so far:
 - Detailed plan and start of the Data resources catalogue IT development cycle in autumn 2025.
 - Metadata integration into the DAAMS proceeded already quite far.
 - Start of the integration plan and requirements on HealthData@EU integration.
- Further information:
 - <https://findata.fi/en/finhits/>
 - <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/projects-details/43332642/101126512/EU4H>

Norway: The SPUHiN project: FAIR Secure Provision and Use of Health Data in Norway (Norwegian Directorate of Health)

- Duration: January 2023 – December 2025.
- Metadata related goals:

- Ensure that metadata for NIPH's national health registries are complete, translated, and aligned with the Norwegian Health Metadata Specification and HealthDCAT-AP. Enrich metadata by adding information on accessibility, interoperability, and reusability. The goal is to improve findability and discoverability across the Norwegian Health Metadata Catalogue, the European Data Portal, and the EU Health Dataset Catalogue.
- OMOP mapping of key variables of national health registries. Metadata description of these variables in the Norwegian Health Metadata Catalogue, The European Data Portal and the EU Health Dataset Catalogue (marked as "conforms to": OMOP).
- Define and publish metadata properties for endpoints, secure infrastructures, and Secure Processing Environments (SPEs). The goal is to enhance their discoverability in the Norwegian Data Catalogue using standards like DCAT and CPSV-AP.
- Outcomes so far:
 - Metadata update for NIPH's national health registries in progress.
 - Chosen key variables are mapped to OMOP based on registry owner input and the article "*Nordic Health Registry-Based Research*". New OMOP-based properties are proposed to improve findability, interoperability and reuse. The mapped variables are ready to be exposed on helsedata.no.
 - Metadata properties for endpoints: Version 0.75 of the specification has been completed and regrouped following feedback from the Norwegian Directorate for Digitalization (DigDir). Further work on extended properties and visual examples are needed before review by VALO, CoP, and TEHDAS2 stakeholders.
- Further information:
<https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/projects-details/43332642/101128232>

Sweden: SENASH - Sweden National services for Access to Swedish Health Data (Swedish eHealth Agency)

- Duration: February 2024 – January 2027.
- Metadata related goals:
 - Piloting MVP version of a national health dataset catalogue.
 - Integration to DAAMs for better user experience.
 - Support for variable metadata that can be used for effective data access processes.
 - Plan for autumn 2025 and 2026: Planning advanced and user-friendly functionalities based on variable-level metadata.
- Outcomes so far: First pilot successfully completed, leading to improvements including tool tips for properties in Health DCAT-AP and better search functionalities.
- Further information: <https://www.ehalsomyndigheten.se/languages/english/senash/>

6 The way forward

The metadata work will continue in the VALO2 project. The following objectives have been identified in the project plan: As part of WS2, the Nordic Health Metadata Network has been further formalised with regular meetings to discuss metadata and metadata catalogue issues in a specific Nordic light and in relation to the EHDS, as well as the work in the parallel EU projects TEHDAS2 joint action and QUANTUM. This increases the potential for a joint Nordic perspective and collaboration on the metadata matters.

The VALO2 task 2.5 named Metadata and metadata catalogues in relation to EHDS will be tracking progress on and supporting national metadata catalogues, engaging with the informal Nordic Health Metadata Network and focusing on current or under development metadata data models alignment with DCAT-AP Health Extension and other EHDS metadata requirements. Also, a preliminary analysis of the needs AI places on metadata will be included.

6.1 Planned next steps

In the shorter term, especially during the VALO2 project, the following goals and next steps have been identified in the metadata work:

- Address DCAT Dataset series methodology and implementation, preferably with the Nordic DCAT-AP experts who already have experience of it. If it is found to be a workable solution after a more detailed investigation, utilizing dataset series concept for providing data resource level for datasets belonging to same register, research data etc. could be suggested as a common approach in the Nordic countries. Also try to influence the development of HealthDCAT-AP in this regard if possible.
- Continue working on the implementation of a consistent dataset definition model in the Nordic countries and try to influence the possible EHDS guidelines in this regard.
- Continue defining minimum common elements for Nordic common variable metadata specification for harmonized approach.
- Address EHDS Article 51 data category coverage gaps. Start grouping what data resources countries have with different data categories in different data holders, possibly with the VALO legal group. Preferably include more data categories and data owners in the discussions.
- Start doing a preliminary analysis of the needs AI places on metadata. In the longer term, try to ensure that the requirements are considered in the national metadata specifications, the common Nordic variable specification, and in the HealthDCAT-AP.
- Look at the possibility of the Nordic Health Metadata Network commenting on the metadata guidelines in the EHDS implementing acts and building connection to EHDS Comitology work. Adequate stakeholder consultation is critical in the development of the implementing acts.



6.2 Identified long-term visions

For longer perspective, the Nordic health metadata experts have currently identified the following goals and visions:

- Establish a sustainable Nordic metadata collaboration model.
- Create possible enrichments of HealthDCAT-AP in the national catalogues in a harmonious way – possibility of “HealthDCAT-AP Nordic extension”?
- Implement data quality and utility labelling in as common way as possible. Goal: Standardized workflow across Nordic countries while respecting national differences.
- Create common Nordic framework to define similar data products and describe most-used variables across countries.
- As possible goal in the future: A common Nordic or Nordic-Baltic health data catalogue. This should be built only if the upcoming EU Central Dataset Catalogue doesn't fulfil the needs of the Nordic or Nordic-Baltic users.
- Interoperability across Nordic countries would require standardized registers and variables. Consider proposing building totally new registries with the same structure in all Nordic countries? It would bring opportunity for standardized Nordic data collection from inception.

Annex 1: Nordic common variable metadata specification v.0.8

FAIR Principle	Group	Class	Property name	Controlled vocabulary	Status	Description (Suggested)	Questions and comments
Findability	Title, description and findability		Dataset identifier		Mandatory	Persistent Unique Uniform Resource Identifier for the dataset the variable is part of.	Need for a logical information model that explains the relationship(s). Previous Property names: IsPartOf, Parent.
Findability/Reusability	Administrative purposes	Variable	Variable identifier		Mandatory	Unique name string for a variable.	Preferably an OID. Unique within the dataset / registry provider.
Findability	Title, description and findability	Variable	Variable name		Mandatory	Short descriptive name of the individual variable, e.g. Gender. The "human understandable" name for a variable.	Previous Property name: Title.
Reusability	Title, description and findability	Variable	Variable technical name		Mandatory	The name of the variable as it is in the dataset. This can be used to connect metadata with data. If the variable does not have a separate "human understandable" name and a technical name, the same information should	Previous Property name: TechnicalTerm.

						be entered in both fields.	
Findability	Title, description and findability	Variable	Variable description		Mandatory	Detailed description of the variable.	
Interoperability	Formats and data types	Variable	Data type	FHIR Primitive Types	Mandatory	Data type as specified in FHIR Primitive Types (https://www.hl7.org/fhir/datatypes.html).	-
Reusability	Content and quality	Variable	Data start date		Mandatory	Used to specify from when there is data available / a variable is valid. Prerequisite for being able to facilitate searches on valid variables on a given date or changes on a variable in a period of time. "Valid" means that there is data available for the variable. New derived variables may be valid back in time.	ISO 8601 (Date). https://www.iso.org/iso-8601-date-and-time-format.html Previous. Property name: DataValidFrom.
Reusability	Content and quality	Variable	Data end date		Recommended	Used to indicate from when there is no data anymore / that a	Previous Property name: DataValidThrough.

						variable became invalid from a given date.	
Findability			Object type		Optional	Type of object / observation unit type the variable is measuring.	
Interoperability	Formats and data types	ValueDomain	Value domain type	Described, Enumerated (from GSIM)	Mandatory	The permitted range of values for a characteristic of a variable https://statswiki.unec.org/display/GSIMclick/Value+Domain	
Interoperability	Classification, value set or terminology	ValueDomain	Value set		Recommended	A linked ValueSet (OID, URI etc) where you can find more detailed information about the codes and values. Used for enumerated variables based on a classification, code list, value set or questionnaire (for example EQ-5D) which can be linked.	Either a link, or a list encoded as text. Previous Property name: ValueSetURL, also separate ValueSetID which was removed.
Interoperability	Classification, value set or terminology	ValueSet	Value set values		Recommended	If the classification, codelist, value set or questionnaire is not published in an outer	Not used if the ValueSetURL is added. Suggested format: quoted

						source and is small, the values may be presented here as comma separated values.	CSV. Previous Property name: CodeValues.
Interoperability	Formats and data types	ValueDomain	Measurement unit	Yes	Recommended	Which unit of measure the variable is based on. E.g. year, month, kg, gram, beat/minute, mmol/l.	Controlled vocabulary?
Interoperability	Formats and data types	ValueDomain	Measurement type	Yes	Recommended	What the variable measures. E.g. time, age, weight and heart rate.	Controlled vocabulary?
Interoperability	Formats and data types	ValueDomain	Minimum allowed value		Optional	Lower limit value (minimum value). Can be used for variables with data types integer and decimal. Indicates that lower values are highly likely to be incorrect and should be excluded from analysis.	Previous Property name: LimitValueLow.
Interoperability	Formats and data types	ValueDomain	Maximum allowed value		Optional	Upper limit value (maximum value). Can be used for variables with data types integer and decimal. Indicates that higher values are	Previous Property name: LimitValueHigh.

						likely to be incorrect and should be excluded from analysis.	
Interoperability	Formats and data types	ValueDomain	Maximum length		Optional	Maximum length of the variable in number of characters/digits.	Previous Property name: MaxLength.
Interoperability	Formats and data types	ValueDomain	Measurement scale	Yes (for example: nominal, ordinal, interval, ratio)	Optional	The type of information provided by numbers. Each of the four scales (i.e., nominal, ordinal, interval, and ratio) provides a different type of information.	The description from: https://www.britannica.com/topic/measurement-scale
Interoperability	Formats and data types	ValueDomain	Precision		Optional	Maximum decimal places.	
Findability			Source		Recommended	Source of the information in the column. Especially needed if there are variables in a national register from another national register, like the Population information system.	Preferably a link to variable description.
Reusability	Content and quality	Variable	Code for missing value		Recommended	Enter the code used for the missing information in the	

						data in the field. Example: -1	
Reusability	Content and quality	Variable	Quality description		Recommended	Overall textual description of the quality of the data that the variable represents, e.g. completeness in the form of completeness (encoding quality) and/or code quality.	Previous Property name: Quality annotation.
Findability			Missing		Recommended	A measure of non-response. (If the measure is stable over time and across the entire dataset. Otherwise, this is presented in the detailed quality description for the dataset.)	
Findability	Findability (Categorization and filtering purposes)	Variable	Keywords	Defined ontologies	Optional	Keywords describing the key content and form of the variable. Preferably use ontologies.	
Interoperability	Formats and data types	Variable	Data representation format		Optional	Data representation format (eg date format YYYY-MM-DD).	Previous Property name: DataFormat, Format.

Reusability	Content and quality	Variable	Replaces		Optional	Link to the variable that is replaced by this variable.	
Reusability	Content and quality	Variable	Is replaced by		Optional	Link to the variable that has replaced this variable.	
Reusability	Content and quality	Variable	Variable lineage		Optional	How calculated, derived and coded variables are derived.	Could link to the original variable, if it is published in a catalogue. Previous Property name: OriginDerivation. Also term Provenance could be used as in HealthDCAT-AP.
Interoperability	Content and quality	Variable	Conforms to		Optional	This property is used to refer to a standard used / mapped to (for example FHIR, OMOP, SNOMED CT, dictionaries).	Previous Property name: MappedToSource.
Findability	Form	Questionnaire	Question in questionnaire		Optional	The question / text in the questionnaire or reporting form for each variable.	Previous Property name: TextInForm.
Reusability	Legal Information	Variable	Degree of identification	Yes	Optional	The data manager's classification of contributions to risks for identification of individuals. Useful information for the researcher when the project needs to consider measures	Need a common controlled vocabulary. Previous Property name: Sensitivity.

						for data minimisation and for the HDAB.	
Findability			Linked variable		Optional	Indicates whether the variable/column is a linked variable/column, i.e. whether it is used to link to other datasets.	
Findability			Linking description		Optional	Describes how the Linked variable is used in different links.	
Findability	Title, description and findability		Presentation order		Optional	Used if you want to present the variables in a custom order that differs from an alphabetical one. Valid values are positive integers and the variables are presented in ascending order based on these numbers.	For all the variables. Depends on how you structure your data.
Findability		Variable	Comment		Optional	Comment.	