Health Data Access Body-NL

Denk mee. Praat mee. Bouw mee.

# HDC Requirements and Specifications

**Project acronym:** HDAB-NL.

**Project title:** *Establishing the coordinating Health Data Access Body in the Netherlands and its supporting digital business capabilities.*

**Number of grant agreement:** 101128662.

**Call identifier/ID of the action:**

**Topic:** D6.1 HDC: Requirements and Specifications

**Starting date of the project:** December 1st 2023

**Duration of the project:** 4 years (dec. 31st 2027).

**Work package (task):** WP6 Catalogue & Metadata (T6.1)

**Submission date of deliverable:** 28 February 2025

**Dissemination level:** PU

Co-funded by the European Union

*Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.*

# 0 Document info

## 0.1 Authors

| Lead Author(s) | Lead organisation |
|---|---|
| **Marianne Knoop Pathuis - Baarda** | HDAB-NL WP6 lead, Health-RI |
| **Hannah Neikes** | HDAB-NL WP6, Health-RI |
| **Cor Melse** | HDAB-NL WP6, RIVM |
| **Harold Kroeze** | HDAB-NL WP6, CBS |

## 0.2 Document History

| Date | Version | Editor | Change | Status |
|---|---|---|---|---|
| 14 November 2025 | 0.1 | Marianne Knoop Pathuis - Baarda, Hannah Neikes, Cor Melse, Harold Kroeze | First Draft for Internal Milestone | Draft |
| 20 January 2025 | 0.85 | Marianne Knoop Pathuis - Baarda | Second draft, shared for HDAB-NL internal review and with stakeholders | Draft |
| 5 February 2025 | 1.0 Concept | Marianne Knoop Pathuis - Baarda | Feedback of internal review processed.<br>Concept v1.0 shared for review by Tactisch Overleg. | Concept deliverable |
| 24 February 2025 | 1.0 | Marianne | Processed TO feedback to deliver final version to HDAB-NL steering board | Definitive deliverable |

# 1 Executive Summary

As part of the European Health Union, the European Union (EU) is advancing the secondary use of health data to drive research, innovation, and evidence-based policymaking.

A key initiative in this effort are the direct grants that aim to prepare the way for national HDABs in the member states as mentioned in EHDS.

This deliverable is a deliverable by work package 6 (WP6) of HDAB-NL. This work package aims to enable The Netherlands' Health Data Access Body (HDAB-NL) to provide data users with a way to discover and assess the suitability of existing health datasets for the data users' intended purpose, as well as data holders with the functionality needed to publish information about their datasets as will be required for the European Health Data Space (EHDS).

Defining requirements and specifications is the subject of this first deliverable.

Most importantly we have taken the statements in the EDHS concerning the cataloguing capability as a starting point for national catalogue requirements. Besides this, stakeholder requirements known to the consortium partners and through stakeholder sessions have been collected and are part of this document, even though they will still have to be checked against the common understanding of EHDS that is still being created, through analysis in the program and cooperation within Europe.

The document has been written during a time when the EHDS texts were still under negotiation and TEHDAS2 and the HealthData@EU pilot were still working on their deliverables. A common understanding of EHDS text is still to be discussed further in the Community of Practice (CoP) for EHDS, and national law still needs to be adjusted and interpreted. Therefore, this deliverable must be seen as a first iteration of requirements and specifications, to be further elaborated based on the deliverables from the projects and initiatives mentioned.

# 2 Table of Contents

# 3 Introduction to Work Package 6: Catalogue and Metadata

## 3.1 Objectives of this work package as described in Grant Agreement

This work package aims to enable The Netherlands' Health Data Access Body (HDAB-NL) to provide data users with a way to discover and assess the suitability of existing health datasets for the data users' intended purpose, as well as data holders with the functionality needed to publish information about their datasets as will be required for the European Health Data Space (EHDS). Tasks include:

- Define requirements and specifications for a national Health Dataset Catalogue (HDC);
- Create a design and architecture for the HDC based on existing solutions among the project partners.
- Implement and pilot the HDC with representative use cases, both from the data user and the data holder perspective.
- Launch an operational national HDC.

Defining requirements and specifications is the subject of this first deliverable and due per 28 February 2025, as described below.

# 4 Description of the First Deliverable

## 4.1 T6.1 National Health Dataset Catalogue (HDC): Analysis and design

**The following is described in the Grant Agreement for the HDAB-NL program but has been slightly adjusted because of current insights.**

The first step for a successful implementation of the Health Data Catalogue (HDC) is to **collect the requirements** among the projected **data holders**, as well as the anticipated **user community**. The deadline for this deliverable is 28 February 2025.

This analysis encompasses:

a. The features supporting users to discover datasets meeting their needs, and
b. The functionality needed for data holders to efficiently publish information on their datasets (in conjunction with WP9 to ensure data quality standards). This information will include data access meta data fields (according to DUO/ODRL standards).

These activities will consider the final text of the EHDS, the outputs from TEHDAS (joint action Towards European Health Data Space) and upcoming joint actions on secondary use, as well as on the lessons learnt on the HealthData@EU Pilot, to ensure alignment and interoperability with the other HDABs HDCs.

The resulting requirements will be mapped onto the existing, and fairly mature, HDC solutions present among the consortium partners, as was stated previously, though in our direct grant efforts we are still working on more detailed user flows, partly based on the business capacities recognized

by the European Commission to be needed to make the overall processes work. Based on these findings, the IT architecture for the future HDC, in the context of the HDAB infrastructure, will combine these existing building blocks with new features to arrive at a network of catalogues that will act as one integrated HDC solution.

Also, we will consider the milestone M6.1 from the EHDS Healthdata@eu pilot, which is a report on the landscape analysis of available metadata catalogues and metadata standards in use.

We will try to consider the strategic choices made for primary use of health data and align with the parties there as much as possible, to achieve an infrastructure that makes sense

Participants: VWS (BEN), ICTU (AE), CBS (AE), Health-RI (AP), RIVM (AE)

# 5 Approach

## 5.1 Overall Work Plan for WP6

To meet the objectives as mentioned above, a work plan was created as a first step in this approach. Based on this plan we have:

- Shared knowledge between the consortium partners on existing cataloguing solutions in the Netherlands. And we had a closer look at the capabilities that are already available at the consortium partners, specifically cataloguing capabilities.
- Worked closely with the other technical work packages within the program to align on requirements: WP5 DAAMS (Data Access Application Management), WP7 SPE (Secure Processing Environment), WP9 Data Quality, WP8 cross border infrastructure. This helps us to take interoperability between components into account, and data quality aspects as a part of requirements for metadata in the catalogue, including functional requirements to make quality of datasets clear to users.
- Learned from our partners working on WP10 about the EHDS.
- Joined the EHDS2 CoP working groups to exchange knowledge with member states on cataloguing and metadata solutions in the member states.
- Organised a breakout as part of an orienting HDAB-NL meeting with a small group of stakeholders in June 2024. Due to the status of EHDS at that time, we could not yet make this a big kick-off meeting unfortunately, which has caused a delay in the project.
- Organised a breakout as part of the HDAB-NL stakeholder kick-off meeting in September 2024.
- Organised two stakeholder sessions specifically for WP6 Catalogue and Metadata, including a presentation from WP9 on data quality and the developments for that in Quantum.
- Processed feedback and comments into our requirements inventory.
- Started a catalogue & metadata working group within an online community for HDAB-NL created by our Ministry of Health.
  To reach this deliverable by February 2025, an internal Milestone was set at 30 November 2024, to deliver internally in the project a first draft of this deliverable and review it within the HDAB-NL team, to determine next steps towards the deliverable in February.

Performed a legal check on the requirements, on interpretation of the newer EHDS text and to consider broader legislation that may apply.

# 6 Existing knowledge amongst consortium partners and stakeholders

Via our Ministry of Health, a stakeholder platform has been set up, so that we can reach out to stakeholders to think along on specific measures and have them review our deliverables. We did notice that the most urgent questions from stakeholders are about the scope of data holders, data categories and datasets. These questions could not yet be fully answered with certainty without the final publication of the EHDS. Stakeholders are concerned that not all data holders impacted by EHDS already know that they will be impacted. This will be a point of attention throughout the rest of the project.

**Health-RI: National Health Data Catalogue and central infrastructure**

Because of the National Growth Fund (NGF) assignment for Health-RI to realise a national health data infrastructure for improvement of the secondary use of health data, a part of the stakeholders of the EHDS in the Netherlands were involved in the starting phase of creating a Dutch national health data infrastructure. In this process functional requirements of a national catalogue have been gathered amongst researchers and several health research data holders. The aim of the NGF project is to realise a central catalogue and a central request process, in which regional nodes are onboarding their metadata: https://catalogus.healthdata.nl/. The requirements from this project have been considered for the HDAB catalogue requirements. This catalogue is still in early stages of development, having a limited set of minimal mandatory DCAT elements as metadata scheme, and an important lesson learned in the process is that data holders face several internal challenges, before they can be ready to share their metadata with a central catalogue, which makes this a lengthy process.

Health-RI's national catalogue has funding till 31-12-2028 to be further developed into a national health data infrastructure including a central request application and SPE functionality. So, it would be an efficient use of public funding to reuse this catalogue as an MVP example for part of the HDAB process, as far as it can support the EHDS requirements. It can play a role in the pilots that are part of the second deliverable within the HDAB-NL project by connecting with other components as described in this document. It is aimed at automated harvesting. Metadata for this catalogue has initially been set up according to minimal mandatory DCAT-AP metadata properties and is being further adjusted for HealthDCAT-AP, to ensure it will comply to the EHDS. This helps data holders to prepare to deliver HealthDCAT-AP compliant metadata by the time the EHDS becomes effective. Furthermore, the link with the Cumuluz program is being investigated, which is focused on primary data use. Linking to that can make primary data findable for secondary use.

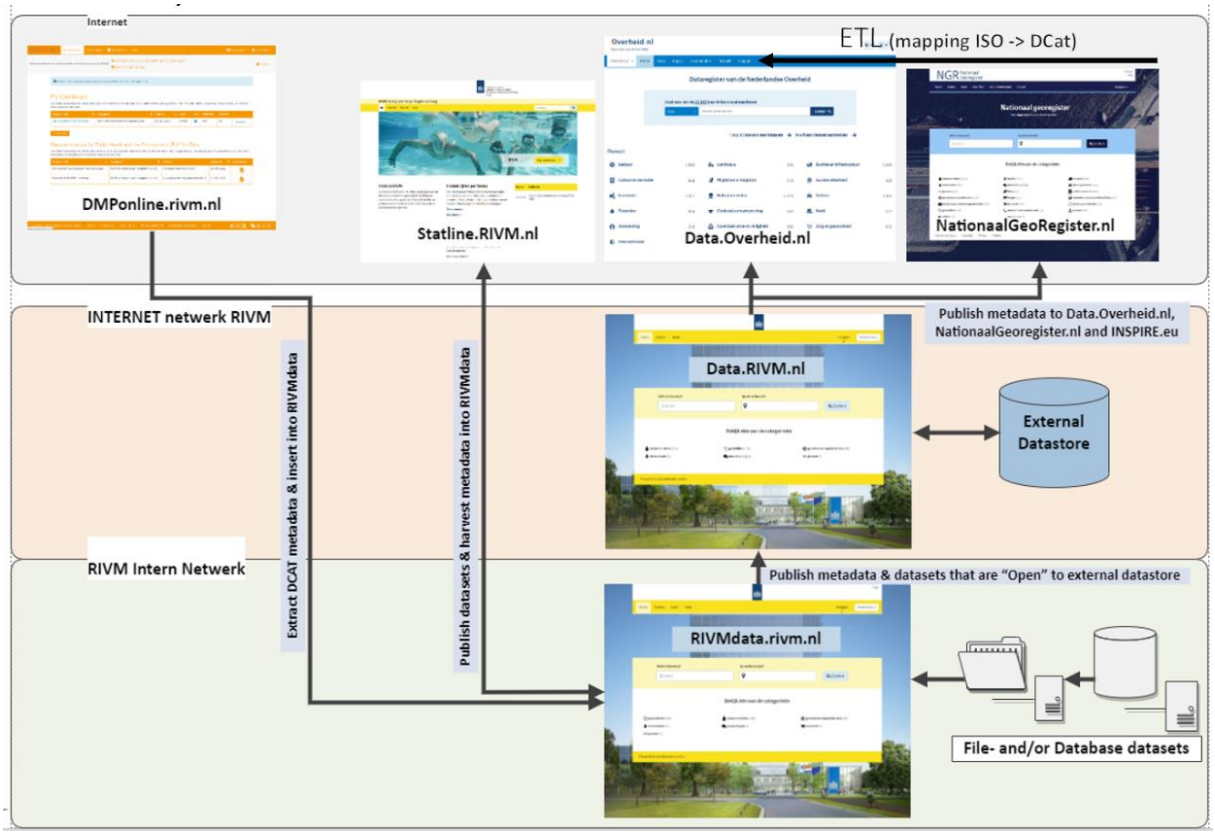
Still, we must keep in mind that there are requirements in EHDS that are not foreseen in the NGF project and therefore need to be addressed in the scope of the HDAB-NL program.

**RVIM and CBS: Cataloguing, DAAMS and SPE expertise**

There is extensive knowledge within the National Institute for Public Health and the Environment (RIVM) and Statistics Netherlands (CBS) - including expertise of the geo-domain and health related (micro)data and statistics - along with a long standing established processes for handling requests for statistical data and public health data catalogues. This offers us a starting point with a lot of requirements known, based on current infrastructures and legislation. Investigation of the potentially additional expected EHDS requirements was an important part of the effort for this first year of the HDAB-NL program.

**RIVM national catalogues**

RIVM has centralized catalogues like https://data.rivm.nl/ and https://bronnen.zorggegevens.nl/Bron/Zoek. The picture below visualises the ways several types of metadata are shared to national catalogues, making available RIVM open data and tabular data.




**CBS national catalogues**

Statistics Netherlands (CBS) publishes open data via https://opendata.cbs.nl/StatLine/#/CBS/nl/. Besides that they offer services around microdata, and a subset of descriptions are made available on the ODISSEI portal https://portal.odissei.nl/ portal (where a number of large collections can be found) and https://data.overheid.nl/, where a part of CBS data is described and can be made available on request to researchers. See also there the theme "Zorg en Gezondheid", containing aggregated tables rather than data files. The data are also findable in the form of PDF reports including code lists, on Gezondheid en welzijn | CBS, with a total overview on Gebruikershandleiding Zorgregisters: Informatie gebruik | CBS . Metadata descriptions here have been automatically derived from CBS's data service center, with some editorial additions.

**Consortium partners and stakeholders**

The knowledge of the consortium partners and their stakeholders forms the basis of our requirements, that we have adjusted to EHDS legislation, as far as current knowledge allowed us. Adjustments will be necessary because the EHDS has been published during the review period of this document. Clarifications for common understanding are still to come as we are finishing up this documentation in February 2025. Furthermore the TEHDAS2 guidelines that we align with are still under public consultation.

Stakeholders we have contacted based on a collective overview of stakeholders through the consortium partners including the Ministry of Health. Identified stakeholders have been invited to think along with the requirements in the HDAB-NL program and several of them are volunteering to be involved in further HDAB-NL pilots. It must be noted that a thorough stakeholder inventory will be carried out, initiated by the ministry of health, and additional stakeholder therefore may bring new requirements and challenges.

## 6.1 Information on existing other catalogues and their properties

From the Healthdata@eu pilot we have received the landscape analysis of metadata catalogues and the standards used. This was the basis for the pilot to start the work on HealthDCAT-AP. We cannot yet point out specific catalogues matching all criteria for HDAB, as HealthDCAT-AP is still in development and more detailed capabilities are still under investigation and discussion. After having worked out the more detailed user processes and functionality needed in the HDAB infrastructure, it will be a logical moment to again review the catalogues out there, and in how far they would match, or have a roadmap to match, the requirements and specifications. Different catalogues have been pointed out to us at stakeholder sessions, usually domain specific with extra useful functions for those domains, but not capable (yet) of exchanging the HealthDCAT-AP format. We should not directly want to replace these specialised catalogues but rather connect them, as they offer a lot of functionality in the domain that are invaluable to their data users. Also, we should consider that these catalogues will also be evolving with EHDS in mind and therefore again inventory their status and roadmap later in the process.

Besides this, we will evaluate the code that will become available in March 2025 for the Central Services catalogue, and the components being developed by the Data Spaces Support Centre and any other initiatives that seem promising and map them to the capabilities that we need for an HDAB catalogue to evaluate if they fit our purpose.

## 6.2 Delays during the project

Unfortunately, we had a slow start due to the delay of the final EHDS publication. This means, as mentioned before in the approach section, that we could not kick off the HDAB-NL stakeholder engagement fully before the summer, as we would have liked, but we did start involving stakeholders at a broad scale right after summer of 2024.

We expect that after the delivery of this deliverable in February 2025, the same month that the EHDS is being signed, following further explanations for common understanding thereafter, we will need to revisit the requirements and see if any requirement needs to be adjusted. As noted above, also TEHDAS2 guidelines and implementing acts will still have impact on the requirements. Therefore, it must be noted that any requirements and specifications currently delivered cannot be regarded as set in stone. Updated versions are to be expected throughout the project. This does not only apply to upcoming EU law, but also when national legislation is adjusted based on EHDS. And of course, based on lessons learned during the pilot phase, we will need to keep updating the requirements and specifications accordingly. These delays may impact (the scope of) other deliverables in the project as well, as more time will need to be spent after February 2025 to work on the requirements based on the definitive texts.  Also, some stakeholders have trouble thinking along with us as they are unsure of the exact scope of the EHDS. Therefore, we can also expect more stakeholder feedback once they become more actively engaged inworking towards the expected changes that the EHDS will bring.

## 6.3 Collaboration within EHDS2 CoP SG2

We have joined the EHDS2 CoP Subgroup 2 on Catalogue and Data Quality where we share knowledge with other member states, share presentations about the different implementation choices, and exchange expectations about changes needed for a good implementation of the EHDS. Comparing solutions between member states will continue throughout the project and can have impact on the Dutch deliverable as we finetune requirements based on lessons learned from other member states.

## 6.4 Collaboration with TEHDAS2

The TEHDAS2 program will be delivering guidelines that, when made definitive, will be the basis for implementing acts that accompany the EHDS. This makes it important for our program to stay closely aligned, share experiences and feedback with the TEHDAS2 program. Through the EHDS CoP we're also in contact with TEHDAS2 members. And starting November 2024, we have started close alignment of work between HDAB-NL consortium members and organisations in the Netherlands that take part in TEHDAS2, exchanging requirements and joining the TEHDAS2 working group meetings. And of course, we will be actively taking part in TEHDAS consultations to share the lessons we learn along the way with the TEHDAS2 project members.

## 6.5 Collaboration with Quantum

Through close collaboration with WP9, we also are in close contact with members of the Quantum project, so that we can stay aligned with requirements about data quality. This has a very close relationship with the metadata model required for the catalogue. Also, it will impact the user interface of the catalogue because the catalogue is required to show to the user clearly the quality of datasets if that information is available.

## 6.6 Collaboration with DSSC

We started alignment with DSSC (Data Spaces Support Centre) because they are preparing general components for data spaces to ensure interoperability and efficiency in creating data space resources and infrastructure.

## 6.7 Collaboration with HealthData@EU

We've closely followed developments in the HealthData@EU pilot and will keep reviewing information about the deliverables from this pilot and the next developments when the results are handed over to the EHDS2 CoP.

# 7 Functional and non-functional Requirements inventory

We have created a list of more detailed requirements for the catalogue and metadata scheme. These are maintained in Excel in the HDAB-NL project space and are included in this document as appendices. In the appendices you can see that we refer to a so-called RS37 requirement framework from the requirements we have gathered. This is a Dutch framework based on OpenDEI which we have used in the requirements processing to check if there are still categories of requirements that need our attention, and which we will further elaborate on in later iterations, based on a cooperation between our work package and Work Package 8, concerning the infrastructure and cross-border functionality.

In the next chapter of this document, we explain the requirements on high level, using visualised flows to explain expected requirements.

In December 2024 the European Commission has shared with us their visualization of the expected Digital Business Capabilities needed. This overview, found within the CoP file sharing environment, is also being used within the project to make sure we do not overlook capabilities that are needed for the cataloguing of datasets and dependencies with other capabilities.

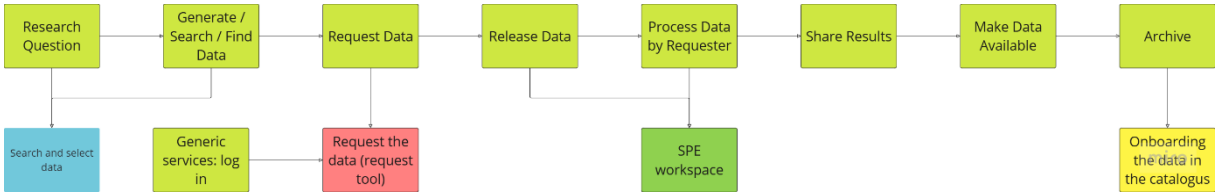
# 8 Data User Flow: Searching for Data(sets) in the Catalogue

At the basis of our functional requirements list is an understanding of the type of users who will need to use the applications we envision to make the health data infrastructure successful. Therefore, we have investigated the processes needed within a health data infrastructure, either national or international.

The aim of the work package is to deliver an infrastructure that will be actively used by the researchers and other intended end users (from now on **user**), to search for available datasets and then either request open data or a statistical answer or apply for permits. We aim to prevent creating an infrastructure that users will try to avoid at all costs. Therefore, we are investigating the processes on the users' side and how to support them best with the intended business capabilities.

An example of a user's journey and how the HDAB components can support this, used as a basis for further discussion with our stakeholders:

*Please note that we are still comparing user journeys with the digital blueprint of business capabilities that the European Commission has presented, and this is only one of the representations that we have used so far. We are working jointly with the other technical work packages to create more extensive and more in-depth visualisations of the user flows to be supported.*



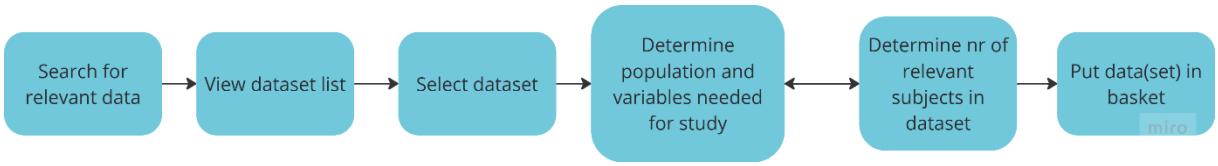

*Note: Light Green is the basic flow that was suggested by a group of stakeholders that consist mostly of university medical centres, and one generic log in step. Blue is the process that concerns the catalogue. Red is the step that corresponds to DAAMS process. Dark green is the SPE process and yellow is the supporting process to onboard data holder metadata into the catalogue.*

Explanation of the general flow, from left to right:

1. A user has a research question
2. The user finds existing data to answer the research question (or needs to generate it, which is a different process).
3. The user requests access to the datasets found
4. The data is released to the user
5. The data is processed by the user (e.g. doing analyses and/or enriching the data)
6. The user shares the research results.
7. The user makes the (enriched) research data available.
8. The user archives the data.

Especially at the **first two steps**, the **catalogue** comes into play. The users want to find available datasets and assess their suitability for the research question.

Zooming in on the user processes in the catalogue:

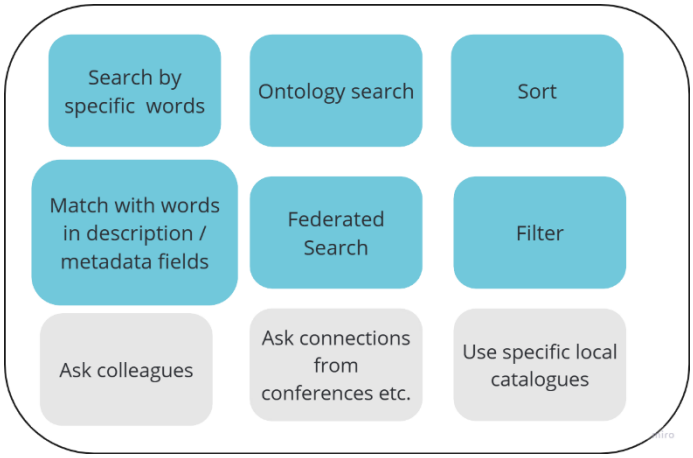



From left to right:

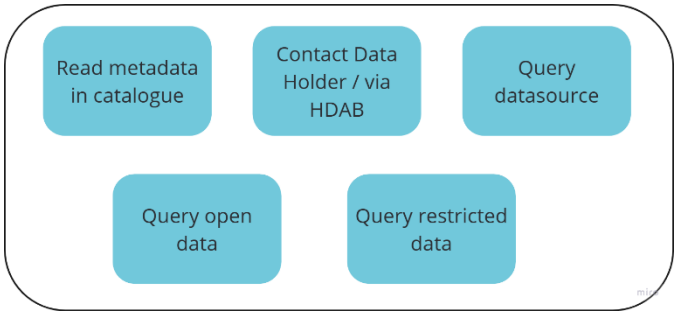
1. The user searches for relevant datasets → A suitable **search** functionality is needed
2. The user views the resulting datasets → The results need to be **shown** in a useful way (N.B. in this step the user should be able to evaluate the quality of the data and whether it fits the research question, which will be further elaborated on in later iterations, and by the work of WP9.
3. The user selects datasets that seem interesting at a first glance → Results must be **selectable**
4. The user determines what population and variables are specifically needed to answer the research question → **Sufficient detail** in the metadata, and/or data discovery/deep search can be necessary, and/or **contact** with the data holder for deeper understanding of the dataset.
5. The user wants to know, for their specific research question, if the (combined) dataset includes **enough subjects.** If not, the user might revisit the needs for the study or continue to other datasets. If so:
6. The user wants to put the dataset in a basket, or in some other way **start an application process** for a **permit** to access the data or start a **statistical request** for an answer about the data.

## 8.1 Suitable Search functionality

How do users now or in the future want to search for data? As the EHDS states that users must be able to determine if a dataset is relevant for their research, the catalogue should offer users the functionality they need to do so. NB: Also searching based on identifiers is a subject of later investigation and is expected to be added to the visual representation of search options later.




How do users want to determine if there are enough subjects in the dataset, to suit their research question sufficiently and to start the request or permit process? The processes mentioned are not all explicitly mentioned by EHDS but have been mentioned by users to indeed make sure a dataset is relevant for their research. They may therefore not all be mandatory part of an MVP but would be expected to be made possible at some point.

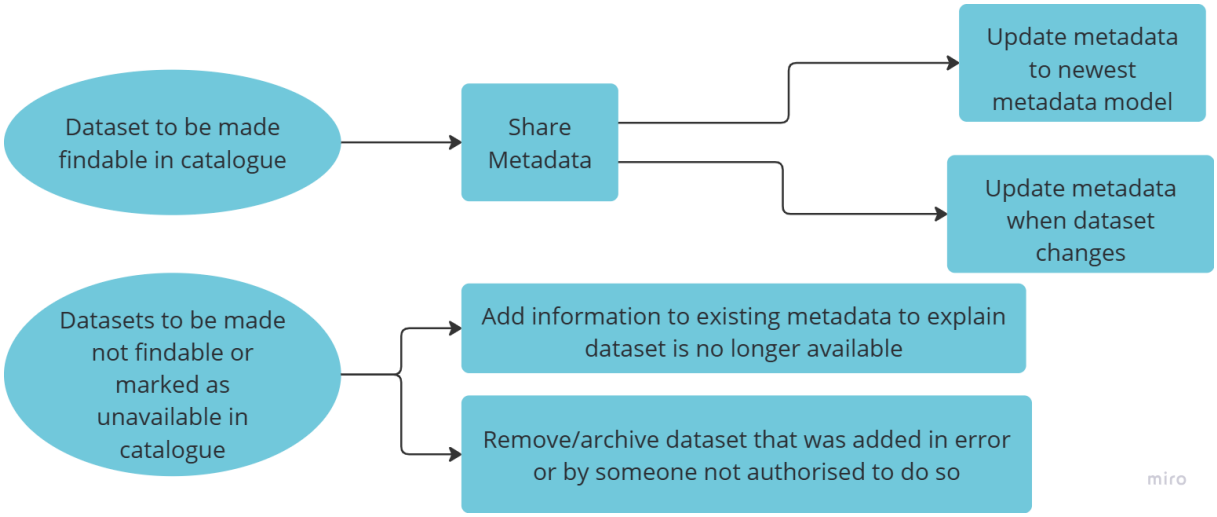



## 8.2 Linking to DAAMS

If a dataset is interesting enough, the user wants to go from the catalogue to the Data Access Application Management System (DAAMS) for a request for a statistical answer, or wants to request a permit to get access. For this, we expect to need a connection from the catalogue to the DAAMS, so that the user can select datasets and then proceeds to the DAAMS, if it would concern datasets of one member state only. The route that must be available would be making the datasets findable in the central European catalogue, from there creating a request, and from there going to the national DAAMS.

# 9 Data Holder User Flows: Managing metadata in the catalogue

Processes that need to be supported for data holders:




To make information about datasets available in the catalogue:

- Share metadata:
  - Manually (e.g. an import function or a form), and/or
  - Automated: Onboarding process for not-yet connected data holders – the automated process is likely the most scalable solution.
- Update process for connected data holders
  - Manually, and/or
  - Automated
- Metadata versioning
  - Manually, and/or
  - Automated

The options that should be developed for manual and automated dataset onboarding (= adding information a.k.a. metadata about a dataset) must partly depend on what stakeholders are able to do by the time EHDS requires them to share metadata.

Removal of a dataset, or in the catalogue marking them as no longer available:

- This could be because a dataset is obsolete (one might still want to refer to its previous existence, though)
- Or the user needs to remove a dataset because it was added in error or not by a person authorised to do so.

## 9.1 Share metadata automatically: Onboarding a data holder's information into a catalogue

The EHDS requires data holders to be able to place information about datasets in a national catalogue, that will in turn be connected to the European central catalogue. The EHDS does not explicitly states that this needs to be an automated process. However, the opinion is shared among many participants of the CoP that automation is essential to make these processes scalable.

There are several examples of automated harvesting out there. The problem can be that this is complex for data holders, especially smaller ones. Therefore, sufficient support for users is needed to guide them in delivering information about their datasets to the catalogue. The following is an example of an onboarding process to get metadata harvested into the national catalogue from data holders that did not yet previously expose their metadata. In other words, data holders that do not yet have technical connections to publish their metadata into the catalogue.

Some remarks on this process:

- It is important to note that this can be quite a taxing process for a data holder, and therefore, it needs to be guided and supported for this process to succeed. Also, there can be multiple solutions depending on the size of the data holders' collection of metadata, the frequency the metadata is updated by the data holder, etc.
- The process that is used as an example here, is still under improvement together with stakeholders and is expected to evolve further if more diverse data holders will use the process to make their datasets findable. **Please note that in the HDAB program, we are still in discussions with other member states on the best ways to support data holders to get their metadata into a national catalogue**. Fair Data Points are one of the options being discussed, though at this moment is not at a sufficient maturity level to be put into production.

For an **example** of how to support data holders to onboard their datasets on a national catalogue in an automated way, see https://health-ri.atlassian.net/wiki/spaces/FSD/pages/279150593

Health-RI data / **Metadata onboarding on the National Catalogue** 1 Jira link

Ready for review

(see What is the future of the National Health Data Catalogue)

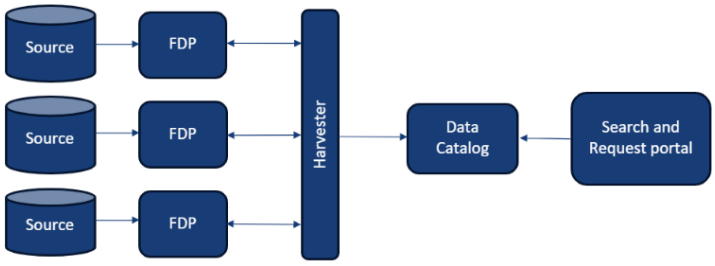



Figure 1. Connection of data (source) to the National Catalogue via an FDP

## How to onboard your metadata to the Catalogue?

There are several steps needed to publish your metadata on the National Catalogue. Here we show the basic steps. You can find some examples of onboarding and scenarios here. For technical documentation please refer to the Health-RI Github: GitHub - Health-RI/health-ri-metadata at master.

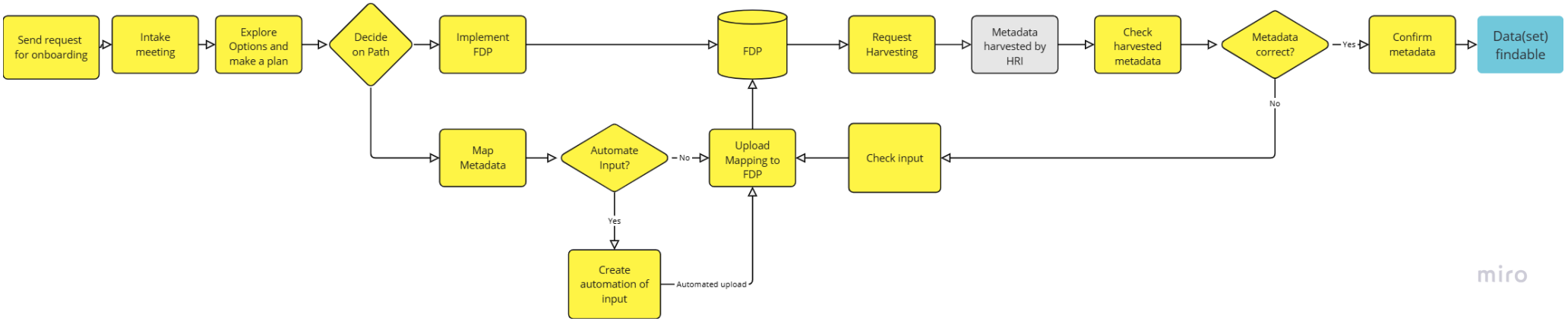
### General onboarding steps

**1.Request**

In this step a data holder/provider reaches out to Health-RI to request onboarding of the metadata, via our service desk: servicedesk@health-ri.nl. A Health-RI contact person is assigned to the data holder/provider and the request is internally registered.

The Wiki page mentioned on the previous page is an **example** of support to onboard data holders so that they can make an automated connection from a FAIR Data Point (FDP) to the national catalogue, giving them the means to keep updating their metadata through a contact point within their own infrastructure. The FAIR Data Point is one of the solutions that is being discussed within the EHDS CoP subgroup 2 on Catalogue and Data Quality. Our consortium partners in their experience also have other options to include information about datasets in their catalogue.

The process to realise this can be depicted as:

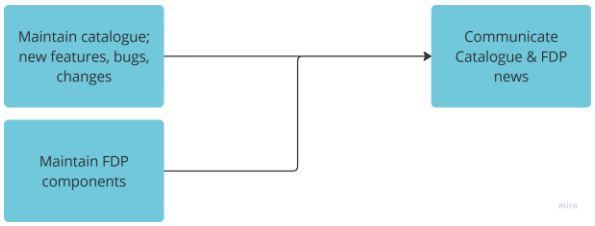


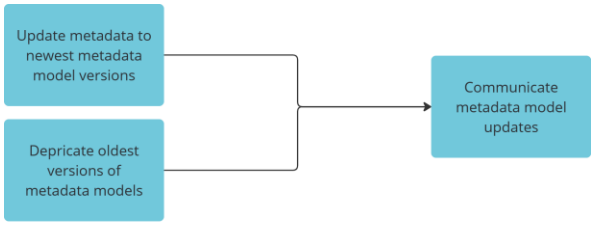

- First, you send a request for onboarding.
- Then, you'll have an intake meeting to determine there is a shared understanding of what is needed.
- The options will be explored together, and a plan is made.
- The data holder decides which path to take, either:
    - The data holder implements a local FDP. The local FDP will be linked to a central FDP, or:
    - The data holder maps metadata and chooses either to:
        - Manually upload the mapping to the central FDP, or:
        - Create automation for uploading mapping to the FDP (which automates updates of metadata).
- Once metadata has been offered to the central FDP, a request is done to harvest it.
- Metadata is harvested
- Then the metadata is checked against the model. If correct:

- Metadata is confirmed.
- The dataset is findable in the catalogue.

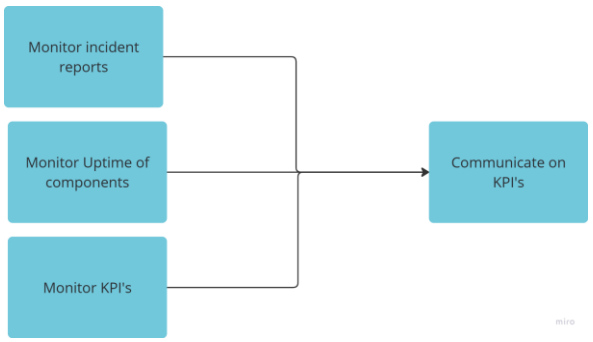
# 10 HDAB Staff User flows: Maintaining the catalogue components

The following processes need attention in creating and delivering the catalogue and to make sure that the ecosystem will be maintainable.



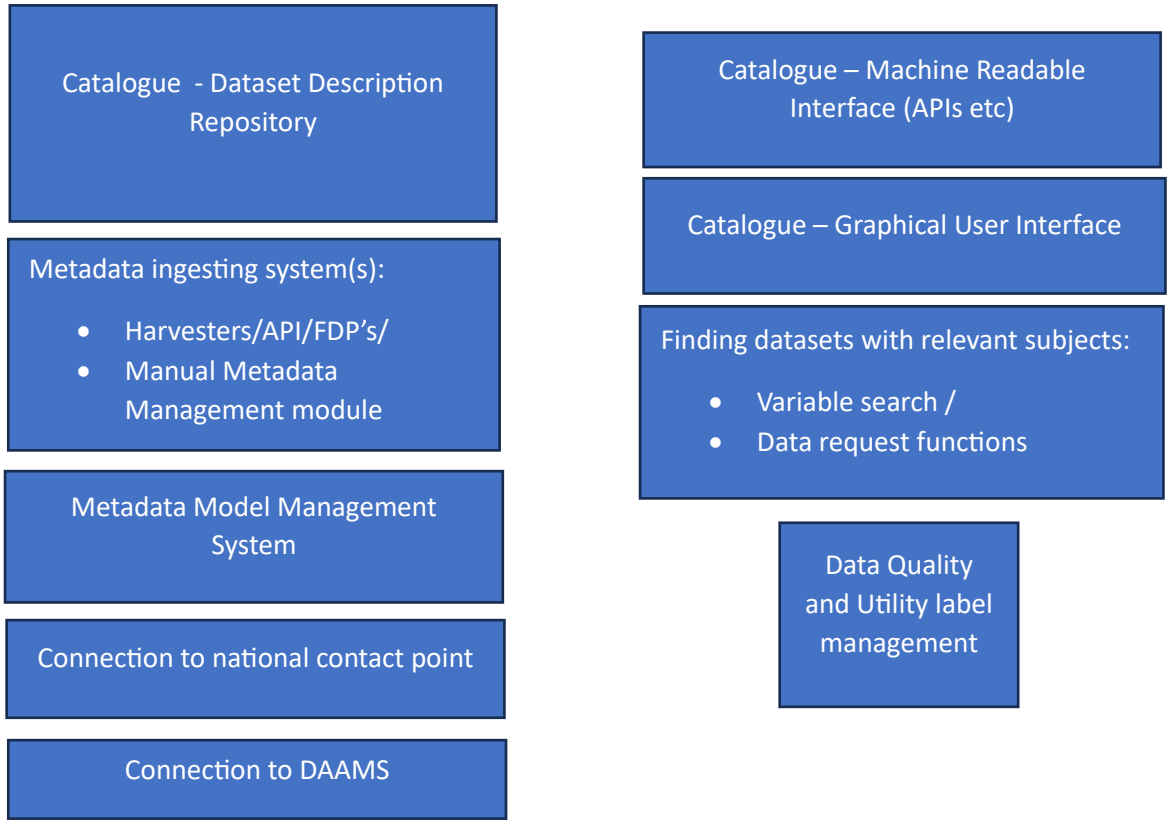




# 11 Overview of expected components needed for cataloguing

This chapter shows an overview of main components needing for the cataloguing capability. Mind that the exact functions needed are also still under debate and needs to be revised within the HDAB-NL program and with its stakeholders, while further investigating the expected digital business capabilities (also incorporating the overview of expected business capabilities that the European Commission has provided), EHDS common understanding that will be further discussed within the EHDS CoP, TEHDAS2 guidelines and implementing acts.

**Core components for the cataloguing capability**




**Supporting Services for the cataloguing capability**


**Catalogue – Dataset Description Repository**

The core functionality of a dataset catalogue is to contain the dataset descriptions that make health datasets findable as the EHDS requires. Therefore, the catalogue as a dataset description repository is the first component to mention. There are still debates on the technical implementation of this and

best methods to use (e.g. discussion on relational database versus triple store and the impact of that on functionalities the catalogue can offer).

**Catalogue – Machine Readable Interface (APIs etc)**

To allow for interaction between systems in the infrastructure and a scalable process the catalogue must have a machine-readable interface. This must also allow the infrastructure to be readied so that we can all profit from the capabilities that AI will offer us to improve health care research.

**Catalogue – Graphical User interface**

To make the catalogue accessible for all envisioned types of users, the catalogue must have a user-friendly interface to enable users to find relevant datasets for their research.

**Metadata ingesting system(s): Harvesters/API/ Fair Data Points (FDP)/ Manual Metadata Management module**

To fill the catalogue with dataset descriptions (a.k.a. metadata about the datasets) there must be a way to enter them. There are several ways to do this and what is the best way (or ways) to go about this is also still subject of international discussions.

- Harvesters / APIs/ Fair Data Points: these are ways to automate the ingestion of dataset descriptions, which is a must have for the infrastructure because we need to make it scalable. However, technical implementation on the data holder side can be complex and for smaller data holders could be too difficult.
- Manual metadata management module: to allow any data holder to get their dataset description into the metadata catalogue, a manual option should be provided. A form purely to submit will not be sufficient, as data holders will need to be able to update their previously entered metadata (and multiple users from a data holder organization will need to be able to do this. Therefor a module within the system willbe needed for manual metadata management.

**Data Quality and Utility label management**

The EHDS requires to describe the quality and utility of datasets. And to make this visible in the catalogue. Therefore, there must be funcitonality for managing data quality and utility labels, which includes the posibility for the HDAB to revoke them, as the EHDS requires.

**Finding datasets with relevant subjects: Variable search / Data request functions**

This doesn't sound much like a (technical) component; however, the debate is still going on how to exactly enable users to determine which datasets they come across after a search on metadata, have relevant subjects for their specific research questions. There are multiple options being discussed and we need to find out in these discussions which option is the most useful and if available, if that makes the other options unnecessary.

- Create specific metadata fields for specific health care domains to further narrow their search
- Allow users to effectively access the variables available in the datasets and use them to set up their request for data from the selected datasets.
- Use tools to ask questions to the data and get back the number of subject available based on the question (from the central catalogue for EHDS this option would be covered by performing the

data request, however the scalability of answering these data requests by data holders is subject of debate, therefore several member states are investigating the options for automating this as is being done in the Genomic Data Infrastructure (GDI) project.

**Metadata model management system**

Because the catalogue will need to support specific metadata models, it is needed to maintain these in a way the catalogue can use them, validate them, and versioning of metadata models will be necessary because these models will evolve over time. For data holders and data users it must be clear which data model(s) are current, which will be deprecated, and how to use that knowledge when searching for datasets.e.g.to get combined results of datasets entered in an earlier version and datasets entered in a new version of the metadata model.

**Connection to DAAMS**

To be able to request data from datasets found in the catalogue, it must be possible to communicate which datasets have been selected in the catalogue, to the Data Access Application Management System. Therefore at least an identifier of that dataset must be communicated. And for usability and efficiency purposes, DAAMS should be able to fetch more details of the dataset from the catalogue.

Also, the other way around, it will be relevant to show to both data users and data holders which datasets have been requested before and how long that took, as they search for datasets in the catalogue.

**Connection to national contact point**

To communicate to the EU Central Catalogue, a connection to the national contact point is needed. Exactly how this should work is to be tested and piloted in the next phase of the HDAB-NL project.

**Supporting Services for the cataloguing capability**

The most important supporting services are the authentication and authorisation of the different types of users that the catalogue will service. This includes

- Data users being able to login to save their previous searches and selections
- Data holders being able to login to maintain metadata
- HDAB staff being able to login and perform their tasks like revoking data quality and utility labels

# 12 Support needed for the users of the catalogue

From a cataloguing point of view, the following training needs have been shared with the work package for training and support, high over:

- Training for users on how to use the catalogue to find the right data for their research
- Training for users on their rights and obligations within the EHDS ecosystem
- Training for data holders on how to map to the right metadata
- Training for data holders on how to technically connect to the ecosystem, or in other words support for **dataset onboarding** (i.e. making datasets findable by adding metadata) in the catalogue
- Training for HDAB staff to maintain the ecosystem

# 13 Legal considerations

A legal check was performed on the requirements as listed in the appendices, with a legal staff member involved in the HDAB-NL project. The following considerations were added, to be further discussed, checked, and considered in further iterations and in the pilot, which we will need to give sufficient attention in a next iteration.

- Consider that according to GDPR it may not be allowed to publish contact names for anyone to see in the catalogue but instead protect them and only show them to the HDAB itself when needed in the process.
    - On the language used in the catalogue, Artikel 2:6 Awb determines that communication must be available in Dutch.
- Nis2, Cyber Security Act and Cyber Resilience Act do not have direct impact on the requirements as listed at the moment.
- The Dutch implementation of the Data Governance Act determines that a central information point will be appointment, for now we do not yet know the result of this. This needs to be checked later. (Translates from the Dutch the statement: "Uitvoeringswet - Data Governance Verordening bepaalt in artikel 2 lid 5, oftewel UDGA art 2(5) bepaalt:
    - Bij besluit van Onze Minister van Binnenlandse Zaken en Koninkrijksrelaties wordt het centraal informatiepunt, bedoeld in artikel 8, eerste lid, van de datagovernanceverordening, aangewezen" )
- In the end the appointed HDAB must comply to https://wetten.overheid.nl/BWBR0037987/2024-04-01. Further evaluation of this is still needed.

As most important basis for the requirements the EHDS articles have been selected that have consequence for the cataloguing capability: This was done based on the definitive text received in November 2024 findable by https://www.europarl.europa.eu/meetdocs/2024_2029/plmrep/COMMITTEES/ENVI/DV/2024/12-04/2022_0140COR01_EN.pdf

| REQID | Quote from EHDS text | Further comments | MoSCoW for piloting MVP |
|---|---|---|---|
| REQ-017.1 | Data holders as specified by EHDS must be enabled to share their metadata with the HDAB catalogue, by themselves, **or** by intermediate entities appointed by national law. | See chapter I Article 2 definition of data holder (page 119). Makes clear that this can also be done by intermediate entities. So, these must also be identified as stakeholders for the catalogue. | M |
| REQ-017.2 | Extra data holders specified in national law must be enabled to share their metadata with the HDAB catalogue. | Dependent on additional national law. To be further investigated later. | M |
| REQ-017.3 | Data holders must be enabled to share their datasets that are described by the minimum categories of electronic health data. | See chapter IV article 51 minimum categories of health data (page 208-209). As HealthDCAT-AP is still in development, this must also be revisited later and will need to be iteratively developed. | M |
| REQ-017.4 | Data holders must be enabled to share their datasets of extra categories appointed by national law. | Dependent on additional national law. To be further investigated later. | M |
| REQ-017.5 | Data holders must be able to correct, annotate or enrich data based on a permit depending on additional national rules | Dependent on additional national law. To be further investigated later. Not sure if the full scope of this statement affects the catalogue itself. But at the least the metadata itself in the catalogue needs to be managed in these cases. | M |

| | | | |
|---|---|---|---|
| REQ-017.6 | Data holders must be able to take measures and additional safeguards for categories of data under paragraph 1 points (f), (g), (i) and (q) in correspondence with additional national law | Dependent on additional national law. To be further investigated later. | M |
| REQ-017.7 | Data holders must be enabled to share metadata making clear extra protections for electronic health data protected by intellectual property rights, trade secrets or covered by the regulatory data protection right | At the latest following a request received from the health data access body. Needs to be investigated further how to make this sufficiently clear in the metadata. | M |
| REQ-017.8 | **HDAB shall make public through electronic means a national dataset catalogue that includes details about the source and nature of electronic health data, in accordance with Articles 77, 78 and 80, and the conditions for making electronic health data available;** | Core requirement for HDAB-NL WP6 Catalogue and Metadata | M |
| REQ-017.9 | (vii) information, at a minimum on an easily accessible website or web portal, on the connection to HealthData@EU of national contact points for secondary use of a third country, or of a system established at international level by an international organisation, as soon as the third country or the international organisation becomes an authorised participant in HealthData@EU. | Not sure if this applies to catalogue itself, so included in this list at least till we're sure it's not picked up elsewhere. Concerns transparency about the third countries that can become part of the ecosystem. | S |
| REQ-017.10 | The national dataset catalogue referred to in point (j)(i) of this paragraph shall also be made available to single information points under Article 8 of Regulation (EU) 2022/868. | Not sure yet what Article 8 of Regulation (EU) 2022/868 says. To be further investigated. | M |
| REQ-017.11 | cooperate with all relevant stakeholders, including patient organisations, representatives of natural persons, health professionals, researchers, and ethics committees, where applicable in accordance with Union or national law; | This is required of the HDAB, so what the HDAB-NL program must start to do in relations to piloting the catalogue to give the HDAB a good basis. | M |
| REQ-017.12 | cooperate with other national competent bodies, including the national competent authorities supervising data altruism organisations under Regulation (EU) 2022/868, the competent authorities under Regulation (EU) 2023/2854 and the national competent authorities under Regulations (EU) 2017/745, (EU) 2017/746 and (EU) 2024/1689, where relevant. | These organisations may be relevant for sharing metadata in the catalogue, or in using it. | M |
| REQ-017.13 | Health data access bodies shall make information on the conditions under which electronic health data are made available for secondary use publicly available, easily searchable through electronic means and accessible for natural persons. That information shall cover the following: (a) the legal basis under which access to electronic health data is granted to the health data user; the technical and organisational measures taken to protect the rights of natural persons; | This part of Article 58 on page 229 could be part of the catalogue because it probably will be information incorporated in the metadata. That means the HDAB needs to make sure that data holders must fill this in when publishing their metadata. | M |
| REQ-017.14 | **The health data holder shall communicate to the health data access body a description of the dataset it holds in accordance with Article 77. The health data holder shall, at a minimum on an annual basis, check that its dataset description in the national dataset catalogue is accurate and up to date.** | Core requirement for health data holder, but also for HDAB-NL WP6 Catalogue and Metadata as it needs to enable the data holder to do this. | M |
| REQ-017.15 | Where a data quality and utility label accompany the dataset pursuant to Article 78, the health data holder shall provide sufficient documentation to the health data | Data Quality and utility - must be represented in the catalogue - needs to be investigated in relations to HealthDCAT-AP and how exactly the accuracy should be sufficiently documented. WP9 within the HDAB-NL | M |

| | | | |
|---|---|---|---|
| | access body for that body to verify the accuracy of the label. | program is leading for this and in close collaboration with the Quantum project. | |
| REQ-017.16 | Article 77<br>Health data access bodies shall, through a publicly available and standardised<br>machine-readable dataset catalogue, provide a description in the form of metadata of the available datasets and their characteristics. | So, the catalogue must be standardised and machine readable. | M |
| REQ-017.17 | The description of each dataset shall include information concerning the source, scope, main characteristics, and nature of the electronic health data in the dataset and the conditions for making those data available. | Metadata requirement | M |
| REQ-017.18 | The dataset descriptions in the national dataset catalogue shall be available in at<br>least one official language of the Union. | So, this is about the descriptions, not the catalogue | M |
| REQ-017.19 | The dataset catalogue for Union<br>institutions, bodies, offices and agencies provided by the Union health data access<br>service shall be available in all official languages of the Union. | | M |
| REQ-017.20 | The dataset catalogue shall be made available to single information points<br>established or designated under Article 8 of Regulation (EU) 2022/868. | Connectivity - need to check this article | M |
| REQ-017.21 | By ... [two years from the date of entry into force of this Regulation], the<br>Commission shall, by means of implementing acts, set out the minimum elements<br>health data holders are to provide for datasets and the characteristics of those<br>elements. Those implementing acts shall be adopted in accordance with the<br>examination procedure referred to in Article 98(2). | | M |
| REQ-017.22 | Datasets made available through health data access bodies may have a Union data quality and utility label applied by the health data holders. | So, the health data holders must be the ones describing the quality and utility of their data - but for them it's not mandatory to apply these labels, and it must be possible in the catalogue. | M |

| | | | |
|---|---|---|---|
| REQ-017.23 | The data quality and utility label shall cover the following elements, where applicable:<br>(a) for data documentation: metadata, support documentation, the data dictionary, the format and standards used, the source of the data and, where applicable, the data model;<br>(b) for assessment of technical quality: completeness, uniqueness, accuracy, validity, timeliness and consistency of the data;<br>(c) for data quality management processes: the level of maturity of the data quality management processes, including review and audit processes, and bias examination;<br>(d) for assessment of coverage: the period, population coverage and, where applicable, representativity of the population sampled, and the average timeframe in which a natural person appears in a dataset;<br>(e) for information on access and provision: the time between the collection of the electronic health data and their addition to the dataset and the time needed to provide electronic health data following the issuing of a data permit or a health data request approval;<br>(f) for information on data modifications: merging and adding data to an existing dataset, including links with other datasets. | Metadata requirements | M |
| REQ-017.24 | 4. Where a health data access body has reason to believe that a data quality and utility label might be inaccurate, it shall assess whether the dataset covered by the label meets the quality requirements forming part of the elements of the data quality and utility label as referred to in paragraph 3 and, in the event the dataset does not meet the quality requirements, shall revoke the label. | It must be possible for the HDAB to **revoke** data quality labels, this means authorised HDAB staff users must be able to transparently adjust registration of quality and utility, of course assuming this must also include communication to the data holder. | M |
| REQ-017.25 | The Commission is empowered to adopt delegated acts in accordance with Article 97 to amend this Regulation by modifying, adding or removing elements to be covered by the data quality and utility label provided for in paragraph 3. | Meaning version management of data and utility labels must be possible (this means metadata model maintenance but may also be catalogue UI development). | M |
| REQ-017.26 | By ... [two years from the date of entry into force of this Regulation], the Commission shall, by means of implementing acts, set out the visual characteristics and technical specifications of the data quality and utility label, based on the elements referred to in paragraph 3 of this Article. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 98(2) of this Regulation. Those implementing acts shall take into account the requirements in Article 10 of Regulation (EU) 2024/1689 and any adopted common specifications or harmonised standards supporting those requirements, where applicable. | This means iterative development considering newly published implementing acts and keeping in close cooperation with the TEHDAS2 project to see them coming and think along with them. Including incorporating the results from the Quantum project. | M |
| REQ-017.27 | The Commission shall establish an EU dataset catalogue connecting the national dataset catalogues established by the health data access bodies in each Member State as well as the dataset catalogues of authorised participants in HealthData@EU. | National catalogue needs to connect to the EU one and authorised participants will have their own dataset catalogue to connect | M |

| | | | |
|---|---|---|---|
| REQ-017.28 | 2. The EU dataset catalogue, the national dataset catalogues and the dataset catalogues of authorised participants in HealthData@EU shall be made publicly available. | Anyone must be able to find the catalogue. | M |
| REQ-017.29 | Article 80<br>Minimum specifications for datasets of high impact<br>The Commission may, by means of implementing acts, determine the minimum specifications for datasets of high impact for secondary use, taking into account existing Union infrastructures, standards, guidelines and recommendations. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 98(2). | Specific metadata scheme and maybe other actions to take for datasets of high impact, to be further investigated. | M |
| | **Other rules and legislation** | | |
| | From Direct Grant: system will be piloted based on systems available at the consortium partners | The appointed HDAB in the end will be able to make final decisions on the systems to use. Those decisions cannot be made within the HDAB-NL project. | M |
| | The catalogue must be maintained and further developed in compliance with the implementing acts that will be published in relation to the EHDS. | | M |
| | The catalogue must be in line with current legislation, like at this moment we can see the following relevant laws and regulations: EHDS, GDPR, data act, data governance act, NIS2, eIDAS, and in NL: the UAVG, WGBO, de algemene wet bestuursrecht, de wet hergebruik overheidsinformatie. | | M |

# 14 Business and application requirement tables

To start, we created an overview of the most basic features a user would expect from a catalogue and grouped them by type of user: Data user, data holder, system manager (a role that would be expected from the HDAB). This we have processed further into more detailed requirements which can be found in the appendix. Drawing user flows and discussing with stakeholders have lead to requirements being added in the table. With further analysis of common understanding of EHDS, TEHDAS guidelines, insights from EHDS CoP, etc., we expect to be updating these further during our program and revisit the MoSCoW classification added to these requirements.

| As a .... | I want to .... | In order to be able to ... |
|---|---|---|
| **Data user** | | |
| data user | search for a dataset | find suitable data for my research |
| data user | have info of a dataset sent to the request application automatically | avoid providing information manually in the request tool that is already known in the dataset catalogue. |
| data user | save the work I have done | continue the next day without starting all over again |
| data user | get in contact with the owner of a dataset | ask additional questions to determine of the dataset is suitable for me |
| data user | be able to do a feasibility study | to determine if there is enough suitable data for my research |

| As a .... | I want to .... | In order to be able to ... |
|---|---|---|
| data user | see what qualifications I need, to use a dataset (e.g. my identity as a researcher from a certain type of organisation, or any qualifications the EHDS could require or allow to be relevant). | get the required qualifications or dismiss the dataset |
| data user | sort search results | get a better overview |
| data user | be alerted if new datasets appear that match a previous search | use the most recent data possible for my research |
| data user | request a selected dataset | use the dataset in my research |
| data user | copy the identifier and title of a dataset or distribution to the clipboard | paste it in another application without typing it over |
| data user | export the metadata in a user-friendly format | use the information in another application |
| data user | be able to change the number of hits on a page | to get a better and faster insight of the total results list |
| data user | Use basic search features without logging in | find out if the catalogue works for me |
| data user | See some quality indication of a dataset | to determine if it is useful to request the dataset |
| data user | Be able to request a specific distribution of a dataset | get exactly the data I want |
| data user | Be able to see all the metadata of a dataset or distribution | get a complete overview of the metadata registration of a dataset |
| data user | See associated datasets of a specific dataset | get motivated to look outside of the box |
| data user | Download public datasets via reference from catalogue | use them in my research |
| data user | have a log of my previous search actions | go back to a previously found dataset that I forgot to save |
| **Data holder** | | |
| data holder | be able to automate the metadata onboarding process | be predictable and be up to date regarding this process |
| data holder | be able to update and add metadata for single datasets | have a lean and efficient onboarding process |
| data holder | be able to add metadata in some manual procedure | skip investments in automated processes |
| data holder | get feedback if the metadata is not correct or complete | correct or complete the metadata registration |
| data holder | be able to validate the metadata that I have created | have as little (avoidable) incidents as possible |

| As a .... | I want to .... | In order to be able to ... |
|---|---|---|
| data holder | get the metadata for a single dataset using an API | automate compare and other processes |
| data holder | be able to bulk upload metadata sets | have a lean and efficient onboarding process |
| data holder | be unable to delete metadata | metadata does not get lost by mistake |
| **System manager** | | |
| system manager | have insight in the processes | respond to incidents |
| system manager | create reports | provide transparency to controllers |
| system manager | be able to control the filter criteria in the front end | provide a better user experience to data users |
| system manager | have logging in place when exceptions occur | give specific information to a problem-solving party of what happened |
| system manager | logging when metadata is uploaded | monitor who is uploading what |
| system manager | have information of contact persons of data holders | have direct contact to a data holder if something goes wrong |
| system manager | have a solid metadata registration process for data holders | guarantee content quality and stability |
| system manager | know to what 'schema' version the supplied metadata should validate | guarantee content quality |
| system manager | have a validation process in place | guarantee content quality |
| system manager | have an automated archive function in place | have a fallback if something goes wrong |
| Metadata specialist | maintain metadata models | allow users to upload and/or find datasets according to specific metadata models |

# 15 Conclusions

Currently, we are delivering the requirements as we see them based on current knowledge, aware that we need to iteratively update them based on new developments and insights. Therefore, this is definitively not a static set of requirements. We must also anticipate that the availability of tools and applications for these capabilities will expand in the coming years, as Health is the first European data space being developed. We will need to regularly analyse available systems and their maturity to make sure we keep making efficient choices.

We have described the expected user flows to support in the chapters above. The appendices contain the more detailed application requirements, metadata standards, and other elements we have identified thus far. As you can see in the requirements tables, we have colour coded different aspects of the RS37 requirement categories. This will allow us as a next step to easily find the requirements belonging to certain categories and refer to specific norms and best practices that already exist.

The tables in the appendices contain

- Functional requirements as gathered from stakeholders to make a catalogue that would really enable users to find the datasets that they need to find
- Requirements directly from the updated EHDS text that require member states to realise a national catalogue for the EHDS.
- Colour coded references to the RS37 framework which is based on Open DEI and helps us to make sure we cover requirements that each national infrastructure should adhere to.

The requirements will need to be revised based on the ongoing deliveries of documentation from TEHDAS2 and HealthData@EU. Besides that, we are still learning about how the ecosystem should cater for the business capabilities that will be needed.

As a further addition, we're starting to investigate the documentation of DSSC (Data Spaces Support Centre), that could play a big part if their open-source components can be applied in our infrastructure, ensuring interoperability with other data spaces.

# 16 References

Besides from references made in the text, some general references can be found here.

| **Description** | **Where to find** |
|---|---|
| Report from the Healthdata@eu pilot on the landscape analysis of available metadata catalogues and the metadata standards in use | https://ehds2pilot.eu/wp-content/uploads/2024/04/HealthData@EU-Pilot_MS6.1_FIN.pdf |
| R&S37 framework for requirements | https://digilab.overheid.nl/puzzels/ |

# 17 Appendix – Explanation of terms used.

Some important terms used:

- **(N)HDC**: (National) Health Data Catalogue. This is a national catalogue in which health datasets can be found as required by the EHDS.
- **DAAMS**: Data Access Application Management System (covering requests for permits or data requests, evaluating those requests, a fee system for the requests, etc.).
- **SPE**: Secure Processing Environment (concerns the process of accessing research data in a secure environment to be able to do research on the data).
- **MoSCoW:** Method to prioritise requirements for a specified phase within the project. Mind that this is not a static categorization: a requirement can be a "Won't have" at some stage of the project and become a "Must have" later.
  - M: Must have (a solution cannot go in production without this, in this phase of the project)
  - S: Should have (users will seriously miss this if the solution goes into production without this in this phase of the project)
  - C: Could have (for a significant part of the users this will be a very useful addition, but the solution can go into production just fine without this in this phase of the project)
  - W: Won't have for now: Useful requirement, but not feasible or important enough in this phase of the project to delay the solution to go into production.

The table below is not a final one, but copied from the TEHDAS2 project, deliverable Milestone 5.3 Technical specification for a National Metadata catalogue. We include it here to try and align on the terms we use, as we should make sure our understanding of terms is the same. Therefore, keep in mind that this is from a TEHDAS2 deliverable which is still in public consultation, and in the CoP General Assembly it has been mentioned to start further efforts to align on the words used for a common understanding. A further iteration of this document will contain a further iteration of the table from the TEHDAS document.

This list can therefore also mention words not (yet) used in this deliverable, but it still helps to have alignment on these terms when discussing topics that address also other capabilities of the HDAB infrastructure.

| Name | Description |
|---|---|
| Access | Processing by a data user of data that has been provided by a data holder, in accordance with specific technical, legal, or organisational requirements, without necessarily implying the transmission or downloading of such data.<br><br>Ref. DGA COM (2020) 767 final |
| Aggregated data | Data combined or collected in summary or other form such that the data cannot be identified with any individual. |

| | |
|---|---|
| | Ref. Legal Information Institute, Cornell University |
| Anonymisation | The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject. Removing personally identifiable information, to definitively not allow the identification of the data subjects. The methods used to anonymise the data are context dependent.<br><br>Ref.<br><br>https:// eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN |
| Anonymous data | Information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. Anonymisation is permanent and irreversible. Principles of data protection do not apply to anonymous information.<br><br>Ref. GDPR Recital 26 |
| Capability | The preferred way to describe some responsibility for an entity that could be solved by a manual process or a technical solution since this should be up to the member state. |
| Data | Any digital representation of acts, facts or information and any compilation of such acts, facts or information,<br><br>Ref. DGA COM (2020) 767 FINAL |
| Data catalogue | Organized inventory of data assets in the organization. It uses metadata to help organizations manage their data. It also helps data professionals collect, organize, access, and enrich metadata to support data discovery and governance.<br><br>Ref. Oracle |
| Data concerning health | Personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status<br><br>Ref. Article 4(15) of the GDPR and EHDS article 33. |
| Data Governance Act (DGA) | Legislative proposal of the European Commission that aims to create a framework which will facilitate data-sharing.<br><br>Ref. Proposal for a regulation of the European Parliament and of the Council on European data governance on European data governance, (Data Governance Act) COM (2020) 767 final |
| Data holder | Legal person or data subject who, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal or non-personal data under its control. |

| | |
|---|---|
| | Ref. DGA COM (2020) 767 FINAL |
| Data hub | Collection of data from multiple sources organized for distribution, sharing, and often sub setting and sharing.<br><br>Ref. Wikipedia |
| Data model | A data model is an abstract framework that organises the data managed by a computer system. It illustrates the entities involved, their properties, and the relationships between them.<br><br>Ref.<br><br>Data model - Wikipedia |
| Data processing | Any operation or set of operations which is performed on personal data or on sets of personal data, whether by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction<br><br>Ref. Article 4(2) of the GDPR |
| Data quality | Comprehensive view of usefulness of data to support decision making. Data quality is defined as “fitness for use” for users’ needs. The OECD views quality in terms of seven dimensions: relevance, accuracy, credibility, timeliness, accessibility, interpretability and coherence. Cost-efficiency is included in the Quality Framework as the eight items.<br><br>Ref. HIMSS Dictionary of Healthcare Information Technology, OECD |
| Data repository | "The data repository is a large database infrastructure, several databases, that collect, manage, and store data sets for data analysis, sharing and reporting.<br><br>Ref. Digital Guardian |
| Data request | Discovery procedure in which the requesting party asks another person for specified information or requests the production of documents.<br><br>Ref. Law Insider |
| Data request management system | System via which a party applying for a data permit or otherwise requesting data submits their data permit application or data request.<br><br>Ref. Act on the Secondary Use of Health and Social Data. Finland. 552/2019 |
| Dataset | Any organised collection of data.<br><br>Ref. OECD |
| "Dataset catalogue" in EHDS, NHDC, HDC | ‘dataset catalogue’ is a synonym to “metadata catalogue” and means a collection of datasets descriptions, which is arranged in a systematic manner and consists of a |

| | |
|---|---|
| | user-oriented public part, where information concerning individual dataset parameters is accessible by electronic means through an online portal<br><br>Ref.<br><br>https://www.europarl.europa.eu/doceo/document/TA-9-2024-0331_EN.docx<br><br>NHDC: National Health Data Catalogue<br><br>HDC: Health Data Catalogue |
| Data sharing | Provision by a data holder of data to a data user for the purpose of joint or individual use of the shared data, based on voluntary agreements, directly or through an intermediary.<br><br>Ref. OECD |
| Data source | Specific data set, metadata set, database or metadata repository from where data or metadata are available. |
| Data subject | Natural person whose personal data is processed by a data controller or processor.<br><br>Ref. GDPR |
| Data user | Natural or legal person who has lawful access to certain personal or non-personal data and is authorised to use that data for commercial or non-commercial purpose.<br><br>Ref. DGA COM (2020) 767 FINAL |
| European Health Data Space (EHDS) | The creation of a European Data Space is one of the priorities of the Commission 2019-2025, including the health sector. In the Communication on “A European strategy for data” it states that the Commission will support the establishment of nine common European data spaces with one of them being the Common European health data space, which is essential for advances in preventing, detecting and curing diseases as well as for informed, evidence-based decisions to improve the accessibility, effectiveness and sustainability of the healthcare systems.<br><br>Ref. European Commission |
| Functional and non-functional requirements | Functional requirements specify what a system should do and are typically phrased as "the system must do [requirement]." In contrast, non-functional requirements describe the system's qualities or characteristics and are often expressed as "the system shall be [requirement].<br><br>Ref.<br><br>Functional requirement - Wikipedia<br><br>https://en.wikipedia.org/wiki/Non-functional_requirement |
| FAIR principles | "The FAIR principles are guidelines aimed at enhancing the management and sharing of scientific data, promoting its utility and transparency. |

| | |
|---|---|
| | They stand for:<br><br>Findable: Data should be easy for humans and computers to locate, with richly described searchable metadata.<br><br>Accessible: Data should be retrievable through a clear protocol, ideally open and free, while adhering to legal and ethical requirements.<br><br>Interoperable: Data should be in formats that allow integration with other data sources and usability across various tools and platforms.<br><br>Reusable: Data should be well-documented and described to support replication and further research, including information on provenance and usage licenses.<br><br>The goal is to maximize the value and impact of scientific data by making it more accessible, usable, and transparent. |
| GDPR | General Data Protection Regulation (GDPR) is a regulation in EU law strengthening and harmonising EU/EEA procedures concerning the collection, storage, processing, access, use, transfer and erasure of personal data.<br><br>Ref. The General Data Protection Regulation (EU) 2016/679 (GDPR) |
| Health Data Access Body (HDAB) | An entity designated by an EU Member State responsible for managing and overseeing access to health data within the framework of the European Health Data Space (EHDS). The HDAB's role includes ensuring that health data is accessed and shared in compliance with legal and ethical standards, facilitating the secondary use of health data for purposes such as research, policymaking, and innovation. Additionally, the HDAB manages the national metadata catalogue, supports interoperability, and ensures that data sharing aligns with EU and national regulations.<br><br>Ref. |
| Health | State of complete physical, mental and social well-being and not merely the absence of disease or infirmity.<br><br>Ref. WHO |
| Healthdata@EU Central platform | The central EU Metadata catalogue for health |
| HealthDCAT-AP | An extension of the DCAT-AP (Data Catalogue Vocabulary Application Profile for Data Portals in Europe) standard, specifically tailored for the health sector. It provides guidelines for describing and sharing metadata related to health data in a standardised way across the EU. By using HealthDCAT-AP, health data catalogues can ensure interoperability and harmonisation, allowing health data to be easily discovered, accessed, and used across different systems and national borders. The |

| | |
|---|---|
| | standard supports the European Health Data Space (EHDS) by facilitating the secondary use of health data while adhering to legal and technical requirements |
| Joint action (JA) | Joint action constitutes an operational action by the member states.<br><br>Ref. European Commission |
| Metadata | Data collected on any activity of a natural or legal person for the purposes of the provision of a data sharing service, including the date, time and geolocation data, duration of activity, connections to other natural or legal persons established by the person who uses the service. |
| Metadata catalogue | definition 1. Metadata catalogues describe the available data collections in a repository or hub<br><br>Ref. Glossary for HealthyCloud)<br><br>definition 2. A metadata catalogue is an organized collection of metadata designed to support data discovery, management, and interoperability by providing a structured repository of information about datasets, resources, and their attributes."<br><br>Ref. Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1-9. DOI: 10.1038/sdata.2016.18 |
| Metadata record | the information about a dataset, stored in the catalogue |
| Metadata repository | A place to store Metadata |
| Minimum viable product (MVP) | The simplest version of a product that includes just enough features to satisfy early adopters and gather feedback for further development. The idea is to launch quickly with a basic set of functionalities that address the core problem or need, allowing an organisation to test its assumptions, learn from user behaviour, and make informed improvements |
| National Node | National Node (NN) is an organisational entity, often linked to a national institution or governmental unit that functions as a national liaison and brings together relevant national stakeholders in the country in a systematic way. The relevant stakeholders may include, for example, the national statistical office, the national public health institutes, representatives from ministries of health, research and/or science, and others. In addition, the NN may function as a discussion and advisory forum in matters of health data and information both for national and international matters<br><br>Ref. INFACT JA on Health information |
| Non-personal data | Data other than personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679.<br><br>Ref. DGA COM (2020) 767 FINAL Open data |

| | |
|---|---|
| Open data | Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.<br><br>Ref. Open Data Handbook |
| Personal data | Any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.<br><br>Ref. Article 4(1) of the GDPR |
| Primary use of data | Use of personal health information by the organisation or entity that produced or acquired these data in the process of providing real-time, direct care of an individual.<br><br>Ref. C. Safran et al. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007 Jan-Feb. 14(1): 1–9. Doi: 10.1197/jamia.M2273 |
| Processing | Processing covers a wide range of operations performed on personal data, including by manual or automated means. It includes the collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction of personal data.<br><br>Ref. European Commission |
| Re-use | "Use by natural or legal persons of data held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the data were produced, except for the exchange of data between public sector bodies purely in pursuit of their public tasks.<br><br>Ref. DGA COM (2020) 767 FINAL |
| Secondary purpose of personal data | Processing of personal data for a purpose other than the primary purpose referred to in primary purpose of personal data.<br><br>Ref. Act on the Secondary Use of Health and Social Data. Finland. 552/2019 |
| Secondary source of data | Organisation or individual other than those responsible for the collection and aggregation of data from their initial source. Secondary sources may redistribute information received from the primary source either in their initial form or after some transformation including further aggregation, reclassification or other manipulation such as seasonal adjustment.<br><br>Ref. OECD |

| | |
|---|---|
| Secondary use of data | Secondary use of data Secondary use of data occurs when data is used for a purpose different from the purpose for which the data was initially collected. (Note: secondary use of data is not the same as re-use of data.)<br><br>Ref. Code of Practice on Secondary Use of Medical Data in Scientific Research Projects - 27 Aug 2014 Final Draft. Innovative Medicines Initiative |
| Systems architecture | Systems architecture is a number of architectural views, used to describe a software system.<br><br>Ref. Wikipedia. |
| User story | User story is an informal, natural language description of features of a software system. They are written from the perspective of an end user or user of a system,<br><br>Ref. Wikipedia. |

# 18 Appendix – list of tables / figures

## 18.1 Table 1: Application Requirements with link to RS037, following from business requirements and user flows and EHDS text of Q3 2024

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| **REQ-001** | **Search functionality** | | | | | |
| REQ-001.1 | Search based on keywords | The user can use keywords and datasets that have that keyword in their title or description are found | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-001.2 | Search based on free text | Multiple search terms must lead to further limited search results. E.g. search for cohort - you see multiple cohorts. Search for cohort AUMC - you see only cohort AUMC and not LUMC or others. | M | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.3 | Filters for standardised metadata elements | Filters must be available to enable the user to find the correct datasets. In the filters the dropdown lists must be filled with the metadata properties. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-001.4 | Query metadata by record-level - assuming open data (other requirements are in Data discovery part) | Search dataset(s) using record-level search functionalities to find relevant data (collections). Assuming we're considering open data only. | S | Interoperabiliteit | Dataservices | Orchestratie |
| REQ-001.5 | Search the data catalogue utilizing ontology-based search using a hierarchical structure, such as SNOMED CT, to allow for more efficient and accurate searches - for a more Google-like experience in searching. | This includes the ability to search for a concept and retrieve all its descendants in the hierarchy without needing to know every possible variant or term. This will enable users to easily retrieve all relevant information related to a concept, improving the overall usability and accessibility of the data catalogue. - To be clarified: does enhancing the search algorithm cover this or do we also need to adjust metadata properties? - can we link the terminology server from DHD/Nictiz for this? | S | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-001.6 | The search will return a list of datasets that match the search | User should be able to view results of the search. | M | Interoperabiliteit | Dataservices | Notificaties |
| REQ-001.6A | Search results can be viewed in a list view | The user should be able to see the results in the form of a list or similar efficient way (as opposed to solely a tile view that allows for less information on the screen) - see also REQ-001.6D. | S | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-001.6B | Search results can be sorted | Sorting options should include the most often used metadata fields, be based on best practices and further to be optimised in collaboration with stakeholders. | M | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.6C | Search results can be generated with a provenance or ontology relation schema | Presenting the search results as a UML like view with all the relations based upon their liked properties. | C | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-001.6D | Search results can be exported | Results can be exported in different formats: csv, Json, xml, pdf, etc. | S | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.7 | As a data requester, I want to explore the catalogue to find a dataset of interest accessing a limited set of variables | This is the requirement for the limited variable set. Future version, we want to use expression logic (AND, OR ...) | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-001.7B | Advanced search - I want to search by expression logic (AND, OR, NOT, XOR, ...) | To have more control over the exact search that is being done (via an advanced search button for instance) | S | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.8 | Data Search | An interface for data consumers to describe the existing data he/she needs and find it in the node registries. In general, the data described is expected to be a cohort of patients with a specific inclusion criterion, such as a given diagnosis or certain type of intervention (codified using a standard encoding system), other characteristics of interest (age, sex, etc.), etc. | C | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.9 | Data search broadcast | A service to send the searches among the EHDS nodes. | M | Interoperabiliteit | Dataservices | Notificaties |
| REQ-001.9B | Data search broadcast transparency | Desirably, the broadcast should be done in a manner that is transparent to the data consumer doing the search. | S | Interoperabiliteit | Dataservices | Notificaties |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-001.10 | The results of the data search should provide an informative summary regarding the number of records found, high-level quality measures and other information useful to the data consumer to decide the potential feasibility of the further actions. | So the user must be able to see easily how many results were found including relevant information, to be further determined based on best practices and stakeholder feedback. | S | Interoperabiliteit | Dataservices | Notificaties |
| REQ-001.11 | Search based on concept browsers, topic or keyword search, actual data scanners or info about publications based on archived data. | | S | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.12 | If I have a user profile (so I am logged in) I can view/access/execute/repeat stored search/query | If a user has an account and wishes to keep the searches performed before, the user should be able to view, access, repeat/execute them | C | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-001.13 | Search results show numbers to know how many hits you have | | M | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.14 | Language: use of multilingualism in the search filter | The user searching using a certain language will find matching datasets even if the data and/or metadata is in another language. | S | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.15 | Use of AI to assist the searching | | C | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.16 | Dataset search based upon Provenance/Lineage relations | The user should be enabled to search further for related datasets to a found dataset. E.g. Which datasets are (indirectly) used to produce a dataset? Which datasets are using this dataset (data linkage?) | C | Interoperabiliteit | Dataservices | Zoeken |
| REQ-001.17 | Show sample data | To determine for a dataset, found by a metadata search, is indeed relevant and useful, a sample/synthetic dataset, when available, could be shown. like header and first 10 records. | C | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| **REQ-002** | **Data presentation** | | | | | |
| REQ-002.1 | Structured representation of the metadata | Metadata must be represented structurally, so that humans and computers are able to interpret and compare datasets by their metadata. | M | Governance | Operationeel | Annoteren |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-002.2 | Representation based on the role of the user | Based on user roles, users can have options in the system. E.g. Data holders must be able to maintain their metadata (either via the catalogue or in some other way, this is subject of investigation), data users need to be able to search, admins must be able to maintain the catalogue technically, HDAB needs to monitor the processes, etc. These authorization levels and role-based functionalities need to be clearly visually represented. | M | Betrouwbaarheid | Toegang | Autorisatie patronen |
| REQ-002.3 | Data quality and utility label | Information about the quality of a dataset, based on the quality labels as supplied in the Quantum project | M | Datawaarde | Metadata | Classificatie |
| **REQ-003** | **Data discovery/ Search deeper** | **Enabling the user to further investigate datasets** | | | | |
| REQ-003.1 | National catalogues | Central data catalogue should seamlessly comprise information from national data catalogues where applicable | S | Betrouwbaarheid | Toegang | Technische_interoperabiliteit |
| REQ-003.2 | Synthetic datasets | Availability of synthetic datasets to help users to determine the value of the dataset or even be sufficient to base their (statistical) research on. | S | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-003.3 | Assessment of usability | Data catalogue should offer characterisation of datasets for requesters to allow assessing the usability of the data for their purposes; | M | Datawaarde | Metadata | Classificatie |
| REQ-003.4 | Specify type of access request | Option for data requesters to specify their type of access request (research, healthcare, policy development, other - still to be defined based on use case input) | W | Datawaarde | Metadata | Classificatie |
| REQ-003.5 | Identify relevant data subjects | Possibility for requesters to identify relevant data subjects based on genotype, phenotype and/or other features such as ethnic origin | C | Interoperabiliteit | Dataservices | Zoeken |
| REQ-003.6 | Specify the area of research | For subject level data discovery, the requester must also specify the area of research based on controlled vocabulary (e.g. disease specific, legal basis) | C | Datawaarde | Metadata | Classificatie |
| REQ-003.7 | Terms of use | Offer suitable “terms of use” that need to be confirmed before the requester can start the search on a subject-level (i.e. data on individuals). Mandatory IF this option is provided. Coud be outside of catalogue work package, needs to be investigated together with DAAMS and SPE work packages. | M | Governance | Operationeel | Dataleveringsovereenkomsten |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-003.8 | Limit data search | Data management system that allows to limit data searches to relevant data only (tagging of data for use restrictions: offer only datasets for search that can be processed under the legal basis of the requester [if applicable], by the category of requester for the purposes of the project) | C | Datawaarde | Metadata | Classificatie |
| REQ-003.9 | Discovery options while preventing bypassing focussed search | Data discovery tool that allows subject-level searches for genotype and/or phenotype without releasing information that can be used to identify subjects and/or reconstruct the data by single search or combination of multiple searches by built in protective features | C | Governance | Operationeel | Gegevensbescherming_Security |
| REQ-003.10 | Logging of searches | Logging of subject level searches | M | Datawaarde | Metadata | Logging |
| REQ-003.11 | Download synthetic datasets | Refer the user to the possibility to download / view synthetic datasets of relevant data collections through the download links or SPE environment noted in the metadata | S | Datawaarde | Metadata | Vindbaarheid_technisch_en_discovery_Directory |
| REQ-003.12 | As a data provider, I want to make my data discoverable, by pushing the metadata of my collections into the (Federated) Catalogue. | Provide support in onboarding metadata onto the catalogue so that data holders are enabled to share their metadata. Provide transparent options to do so with corresponding documentation and make sure these are findable from the catalogue itself. It's not yet clear in how far search processes will be federated. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| **REQ-004** | **Data use** | | | | | |
| REQ-004.1 | Versioning | Possibility of metadata versioning in case of deletes or rectification or updates with additional metadata points | M | Interoperabiliteit | Dataservices | Versiebeheer |
| REQ-004.2 | Archiving | Possibility to archive processed metadata sets and allow referencing for publications. | S | Interoperabiliteit | Dataservices | Versiebeheer |
| REQ-004.3 | Publication search | Automated search for publications based on mandatory reference to 1+MG / sharing search URLs/strings with someone else | C | Datawaarde | Metadata | Vindbaarheid_technisch_en_discovery_Directory |
| REQ-004.4 | Conditional use of data | As a data requester, I want to know conditions for data use and data access instructions, directly from the catalogue. | S | Governance | Operationeel | Dataleveringsovereenkomsten |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-004.5 | Adhere to extended metadata requirements, e.g. DCAT NL + Health DCAT + HRI Core + Petal + Prov | Besides HealthDCAT-AP there are additional NL specific requirements for metadata, e.g. DCAT NL, or information specific to funders and information specific to certain health domain that we would like to include to make the catalogue more valuable. DCAT-NL would be mandatory make the catalogue interoperable within the Netherlands. But is not an EHDS obligation. | S | Interoperabiliteit | Modellen | Translatie |
| REQ-004.6 | Selection and request | The catalogue must enable the user to select an existing dataset and navigate to a request functionality to request it. | M | Betrouwbaarheid | Toegang | Technische_interoperabiliteit |
| REQ-004.7 | Ability to select one or more metadata sets and store them in a 'shopping basket' | | S | Betrouwbaarheid | Toegang | Technische_interoperabiliteit |
| REQ-004.8 | Ability to link with an external workflow system | | | Betrouwbaarheid | Toegang | Technische_interoperabiliteit |
| **REQ-006** | **Metadata management and Data integration** | | | | | |
| REQ-006.1 | Registration and management of the metadata records for a dataset | How to register and after that manage the metadata records for a dataset must be handled based on best practices, automated as much as possible for scalability reasons. | M | Interoperabiliteit | Dataservices | Versiebeheer |
| REQ-006.2 | Management of multiple versions of a dataset's metadata records | It is to be further investigated how to handle this exactly, based on best practices and stakeholder input, developing iteratively if not available out of the box. Also, the guidelines from TEHDAS2 may provide insights on how to handle this. | M | Interoperabiliteit | Dataservices | Versiebeheer |
| REQ-006.2A | Manage updated version of metadata records continuously whenever underlying data changes | e.g. metadata age or gender distribution changes if data set is changing. Needs to become clear how this is to be handled. | M | Interoperabiliteit | Dataservices | Versiebeheer |
| REQ-006.3 | Support for consuming metadata from a FAIR Data Point (Harvesting) | As FAIR Datapoint are used by a number of parties in health-related research it would be beneficial to be able to harvest them directly into the catalogue. | S | Betrouwbaarheid | Toegang | Automatisering beheer koppelingen |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-016.3.B | Support for harvesting protocol | Best practices for harvesting must be adhered to like the OGC CSW 2.0.2 Open Standard - pyCWS.org - or OAI-PMH etc. Choices and/or prioritization needed, based on best practices and the stakeholder collaboration in the ecosystem. | M | Betrouwbaarheid | Toegang | Automatisering beheer koppelingen |
| REQ-006.4 | Support for consuming metadata from a DCAT v2 source | As DCAT v3 is very new, data holders may still have only DCAT v2 available while needing to comply to the EHDS. To support the ecosystem the catalogue should have backward compatibility with DCAT v2 (though some information loss might result). | S | Betrouwbaarheid | Toegang | Automatisering beheer koppelingen |
| REQ-006.5 | Support for consuming metadata from a DCAT v3 source | DCAT v3 is the new standard the catalogue must be able to support sources on DCAT v3. | M | Betrouwbaarheid | Toegang | Automatisering beheer koppelingen |
| REQ-006.6 | Support for metadata pulling into the catalogue | Either push or pull must be possible to make the catalogue function. Enabling them both would be a better solution. This needs to be further investigated based on best practices and needs in this ecosystem. | M/S | Betrouwbaarheid | Toegang | Automatisering beheer koppelingen |
| REQ-006.7 | Support for metadata pushing into the catalogue | Either push or pull must be possible to make the catalogue function. Enabling them both would be a better solution. This needs to be further investigated based on best practices and needs in this ecosystem. | M/S | Betrouwbaarheid | Toegang | Automatisering beheer koppelingen |
| REQ-006.8 | The metadata holder and catalogue of origin are linked to the data set | Metadata holder must be linked by the metadata set, and the catalogue from which we harvest, so that we can show the proper acknowledgements and have traceability. This needs to be possible in the metadata model. | M | Datawaarde | Metadata | Classificatie |
| REQ-006.9 | Data validation and quality checks on manually entered metadata | Validations against minimal metadata is the least, then checking also against indication of more extended metadata models that are supported by the catalogue and making transparent to which of those the metadata complies. (e.g. DCAT-AP v3, DCAT-NL, HealthDCAT-AP, Oncology specific metadata model). | M | Datawaarde | Metadata | Classificatie |
| REQ-006.10 | Data validation and quality checks on automated metadata entry | A procedure must be put in place that can check metadata set and communicate errors to the source of the metadata | M | Datawaarde | Metadata | Classificatie |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-006.11 | Draft / staging of a metadata set | It should be possible to describe a dataset but not yet publish the metadata in the catalogue, e.g. because the user still needs to collect parts of the metadata, cannot enter all details in one go, or because the user filling in part of the details is not the person with the mandate to publish the metadata on behalf of the data holder organisation. Mind that this also makes it necessary to have a user account in the catalogue to return to. | S | Datawaarde | Metadata | Classificatie |
| REQ-006.12 | Review of a metadata set by a second person before publishing | It would be helpful to enable users to request a review by someone else on a metadata set before publishing it. | C | Datawaarde | Metadata | Classificatie |
| REQ-006.13 | Support for import of metadata (might be linked with REQ-006.6) into the metadata catalogue | Data holders that cannot realise a connection to the catalogue must be able to do an import of metadata on the catalogue | S | Governance | Operationeel | Service_levels |
| REQ-006.14 | Support for export of metadata from the metadata catalogue | A metadata export may be requested by users who want to be able to process the metadata in other systems. And it can be needed to be able to comply to an exit-strategy (a loosely coupled module can be replaced without loss of any data), although in those cases harvesting anew would solve all automatically harvested metadata. Also, the data holder may request this to confirm what | S | Governance | Operationeel | Service_levels |
| **REQ-007** | **IAM / IAAI** | | | | | |
| REQ-007.1 | Support for SAML | If users have a user account in the catalogue and DAAMS they should easily be able to switch between applications without unnecessary login steps. Mind this doesn't necessarily mean that the user is an identified person, but at the least that the user is identified as that user and can continue actions done before | S | Betrouwbaarheid | Identiteit | Authenticatie |
| REQ-007.2 | Support for OpenID Connect | See above. Best practices for login must be applied, if login is available. Though login in itself is not needed in the most minimal form of catalogue functionality for data users. It is however mandatory for managing previously entered metadata. This could be a natural person, or a system logging in. | S | Betrouwbaarheid | Identiteit | Authenticatie |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-007.3 | Support for group membership-based permissions | Authorization must be based on best practices like this, based on current state of technology. At a minimum for data holder(system)s. | S | Betrouwbaarheid | Identiteit | Autorisatie patronen |
| REQ-007.4 | Support for role-based permissions | Authorization must be based on best practices like this, based on current state of technology. At a minimum for data holder(system)s. | M | Betrouwbaarheid | Identiteit | Autorisatie patronen |
| REQ-007.5 | Permission per metadata dataset | The user should be able to edit a profile so that he can search for metadata set of which the data access conditions meet the profile. Mind that the user should NOT be able to give themselves access to restricted information based on non-verified identity details. | S | Betrouwbaarheid | Identiteit | Autorisatie patronen |
| REQ-007.6 | Support for e-IDAS | It is to be expected that eIDAS, EU login or an equivalent will be needed at the least for access to SPEs. For a seamless experience this might mean users would want to login with that in the catalogue as well, though directly navigating to catalogue should not require personal identification from the user. Best practices need to be followed here and the GDPR considered. This could mean logging into the catalogue based on a created account (using common MFA options), and only when trying to navigate to an SPE a higher-level identification would be requested. | W | Betrouwbaarheid | Identiteit | Autorisatie patronen |
| **REQ-008** | **Portal management functionalities** | | | | | |
| REQ-008.1 | Administrative role that enables management of the portal | There should be admin users who are the only ones that can do maintenance of the portal options itself | M | Betrouwbaarheid | Identiteit | Autorisatie patronen |
| REQ-008.2 | Management of metadata templates | Manage preferably automated harvesting of metadata templates and being able to validate against it, blocking or warning. Make sure that users with permissions for it can manage templates | S | Interoperabiliteit | Dataservices | Orchestratie |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-008.3 | Management of metadata | Manage metadata that was manually entered (automatically harvested metadata must be maintained at the source and updated via automated harvesting) | S | Interoperabiliteit | Dataservices | Orchestratie |
| REQ-008.3A | Metadata entry interface | Be able to deliver metadata through a user interface that holds the metadata schema (all required / recommended / optional fields) | S | Interoperabiliteit | Dataservices | Orchestratie |
| REQ-008.4 | Management of search engine indexing | It must be possible to have insight in and manage the sources that are being harvested | M | Interoperabiliteit | Dataservices | Orchestratie |
| REQ-008.5 | Management of search engine optimization (SEO) | It must be possible to optimise the search process | S | Interoperabiliteit | Dataservices | Orchestratie |
| REQ-008.6 | Usage statistics, including search statistics | We need to be able to see usages statistics: who logs in, who searches what, which datasets are popular, which datasets are never used, ...<br><br>E.g. We are interested in:<br>A. How many datasets are in it; possibly broken down into<br>-- Connection over time<br>-- Connections per participant/ organization / region<br>B. How many data-seeking users; possibly broken down into<br>-- Over time<br>-- By organization/ Region<br>- -Active users performing concrete searches<br>-- What are they looking for (possibly dividing into categories)<br>-- How many Find dates (put on list) (plotted in time)<br>-- How many mishits (search result = 0) (in time/ what (categories) searched for?)<br>-- How much do they request data (future with request tool) (plotted in time) | M | Datawaarde | Verantwoording | Logging |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-008.7 | Adding new users, deprecating users no longer working at their organisation | It must be possible to do common user management for data holders, data users, HDAB users and maintenance parties. | M | Betrouwbaarheid | Identiteit | Automatisering beheer koppelingen |
| REQ-008.8 | Sufficient help functionalities for users on data holder side | Support must be available for the intended user groups to perform their tasks related to the catalogue. | M | Governance | Operationeel | Webtoegankelijkheid |
| REQ-008.9 | Accessibility: EN 301 549/WCAG 2.1 | Make sure users have access to the catalogue according to best practices, standards, current state of technology. | S | Governance | Operationeel | Webtoegankelijkheid |
| **REQ-009** | **Data holder Dashboard functionalities** | | | | | |
| REQ-009.1 | Per metadata set insight in search and how the metadata set is found | Per metadata set we want to see how often it turned up in searches, and how often it was further looked at and maybe some additional info like from which countries was it searched the most | S | Datawaarde | Verantwoording | Logging |
| REQ-009.2 | Access to the management functions for the dataset' metadata | See management of metadata | M | Betrouwbaarheid | Identiteit | Autorisatie patronen |
| REQ-009.3 | Insight in requests for the dataset's of the data holder | Per data holder see how often the datasets were requested | S | Datawaarde | Verantwoording | Logging |
| **REQ-010** | **User Dashboard functionalities** | | | | | |
| REQ-010.1 | Insight and access to their (performed) searches for datasets | Users must be able to see earlier searches | S | Governance | Operationeel | Webtoegankelijkheid |
| REQ-010.2 | Insight in the requests for datasets initiated by the user | Users must be able to see requests and status of those | C | Governance | Operationeel | Webtoegankelijkheid |
| **REQ-011** | **Metadata template management** | | | | | |
| REQ-011.1 | Ability to register templates | Be able to create templates as well as import templates from an external library | C | Interoperabiliteit | Modellen | Historie_en_Tijdreizen |
| REQ-011.2 | Management of templates, including version management | During pilot year it will already be necessary to deal with updates to HealthDCAT-AP | M | Interoperabiliteit | Dataservices | Versiebeheer |
| REQ-011.3 | Ability to activate / deactivate specific versions of a template | To allow data holders some time to adjust to new versions of metadata models it is needed to keep several versions of metadata models active and later on deactivate them at least for new metadata entries. | S | Interoperabiliteit | Dataservices | Versiebeheer |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-011.4 | Ontology & thesaurus management | Use search features from thesauri and ontologies | W | Interoperabiliteit | Modellen | Translatie |
| REQ-011.5 | Ontology & thesaurus mapping functionality | Use mapping functions from thesauri and ontologies | W | Interoperabiliteit | Modellen | Translatie |
| REQ-011.6 | Support for external DCAT application profiles | This would extend the catalogue so that it can also show other domain's DCAT-AP metadata, e.g. Geo. | W | Interoperabiliteit | Modellen | Translatie |
| **REQ-012** | **Explore the catalogue** | | | | | |
| REQ-012.1 | Anonymous users can explore the catalogue and retrieve basic information | There is a part of the catalogue that's open to anonymous users | M | Governance | Operationeel | Webtoegankelijkheid |
| REQ-012.2 | Other functionalities such as searching for collections fulfilling specific criteria will be limited to registered users only. | | C | Governance | Operationeel | Webtoegankelijkheid |
| REQ-012.3 | An anonymous user can access the federated catalogue through the dashboard, which will display the collections registered (e.g. in EUCAIM for images). | | C | Interoperabiliteit | Dataservices | Zoeken |
| REQ-012.4 | An anonymous user will be able to retrieve basic aggregated information of a specific collection (data model to be defined in the Hyper ontology). | | M | Interoperabiliteit | Dataservices | Zoeken |
| REQ-012.5 | An anonymous user can retrieve the access condition for a specific collection. | | M | Interoperabiliteit | Dataservices | Zoeken |
| REQ-012.6 | When finding a dataset based on HealthDCAT properties in the EU catalogue, the user can navigate to the more local (national or even more local) catalogue to gain more information about the specific dataset | | S | Interoperabiliteit | Dataservices | Zoeken |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-012.7 | A user can choose to search in specific member states preferably, or to include specific more local catalogues | | C | Interoperabiliteit | Dataservices | Zoeken |
| REQ-012.8 | A user can choose to search also in external catalogues outside of EHDS grounds | | W | Interoperabiliteit | Dataservices | Zoeken |
| **REQ-013** | **Register collections into the catalogue** | | | | | |
| REQ-013.1 | A data provider would like to register a collection related to the EUCAIM project in the Federated Catalogue. The data provider should prepare the metadata of the collection following EUCAIM's specification. | Show collections, which is in DCAT likely to be delivered as Dataset Series, liked to datasets with Dataset in Series. | S | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-013.2 | A data provider is authorised to register collections in the Federated Catalogue. The authorisation requires the submission of an application and the signature of a Service Level Agreement. This will be implemented in the resource management section. | This is a requirement in current legislation and doesn't one on one translate to the EHDS situation or HDAB pilots. | W | Governance | Operationeel | Service_levels |
| REQ-013.3 | A data provider service registers the collection in the Federated Catalogue. | This is a requirement to make sure collections related to international projects can be made available in the catalogue. Will need to be investigated further to see what the exact approach needs to be. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| **REQ-014** | **Language support** | | | | | |
| REQ-014.1 | The software, user interface and documentation are available in both the Dutch and the English language | In the Netherlands we need to provide Dutch in general, but because there's a large group of international researchers English will also be a must have in the end. To start with a pilot, one language would suffice. | S | Governance | Operationeel | Webtoegankelijkheid |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-014.2 | If searching on Dutch terms, you should also find datasets with English metadata values, if the Dutch ones are not available in Dutch for a dataset | Search should be smart enough to search in more than one language. To be further investigated, also based on how this is solved on a European level. | S | Governance | Operationeel | Webtoegankelijkheid |
| REQ-014.3 | If in Dutch mode, metadata values are shown in Dutch, if in English mode, metadata values are shown in English. If one of those is not available, the one that is there must be shown. E.g. In Dutch mode you see English values if there are no Dutch values in the metadata of a dataset that is part of the results. | This makes sure that a search will yield results if there are any, even if they are not in a language the user understands. Because of this, the user will at least know that there is something. | S | Governance | Operationeel | Webtoegankelijkheid |
| REQ-014.4 | Support other member states' languages so that the researcher can read the metadata - at the least show the English metadata if Dutch is not available, if user has Dutch as a language selected in the UI | | W | Governance | Operationeel | Webtoegankelijkheid |
| REQ-014.5 | By default, use system or browser language from language settings, for setting the language in the catalogue. | Based on best practices. | W | Governance | Operationeel | Webtoegankelijkheid |
| REQ-014.6 | The user must be able to choose whether to use translation of metadata values to your native language, or to English, or to disable it. | | W | Governance | Operationeel | Webtoegankelijkheid |
| **REQ-015** | **Hosting environment** | | | | | |
| REQ-015.1 | The software can be hosted by the HDAB-NL itself, or a hosting party of HDAB-NL choice | As the HDAB-NL has not been appointed yet, we will make no choice in this matter. | W | Governance | Bestuurlijk | Eigenaarschap |
| REQ-015.2 | Malware protection | Hosting partners have measures in place to scan for malware URLs - this is essential to our reputation as a safe catalogue to search for data | S | Governance | Operationeel | Gegevensbescherming_Security |
| **REQ-016** | **Metadata and Interoperability** | | | | | |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-016.1 | Standardised metadata (FAIR) | The fields of the health DCAT (and other) metadata standards must be standardised (in understandable language), so every field of the metadata has the same meaning and can be used to write metadata by every organisation in the Netherlands. This is needed for any domain specific metadata. | S | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-016.2 | Flexibility in adding, using and referring to new and different metadata standards | Reference to the list of standards | S | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-016.3 | Unity of language / ontologies, standards for the data itself: Support for external Ontologies like SNOMED and ICD-10 | Use the expertise of medical specialist to establish well-defined ontologies and unity of language. In metadata state on which ontology the dataset is based.<br>This requirement is about referring to ontologies in the metadata. When metadata records contain a SNOMED link or other referrals to external ontologies, they must be supported throughout the metadata catalogue chain | C | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-016.4 | Deepen (Health) DCAT); Controlled vocabularies | Health specific dropdown lists of predetermined themes, categories, keywords, etc are needed to prevent proliferation. This can be added via URLs. | S | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-016.5 | It must be possible to have metadata entries that describe non-clinical data (which does not have information such as age, population coverage, number of individuals, which are mandatory in the HealthDCAT-AP) | To support valuable parts of health-related research that do not have actual patients or population to describe in the metadata. To investigate further and add examples. | S | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-016.6 | Descriptive metadata | Need for information in the metadata about; 1) context of data collection, 2) context where data is collected, 3) insight in data lineage, 4) information about data holders, 5) locations of data (maps) | S | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-016.7 | Ability to indicate connections between different datasets / collections / catalogues etc. | This will prevent a lot of duplicates in results found while the user remains unaware of it. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-016.8 | Additional metadata to describe in more detail what is not covered in HealthDCAT-AP (domain-specific metadata) | There may be several properties that stakeholders find necessary for a well working catalogue which are not covered in the HealthDCAT-AP specification. For example: cohort studies, care process, synthetic data, domain-specific metadata. We would want to extend our catalogue to contain those properties as well. While keeping in touch with the HealthDCAT-AP team to give our feedback. | W | Datawaarde | Metadata | Classificatie |
| REQ-016.9 | Metadata on data quality, including quality of the metadata itself | Data quality values may be automatically derived in the catalogue | C | Datawaarde | Metadata | Classificatie |
| REQ-016.10 | Support full HealthDCAT-AP | HealthDCAT-AP is not yet final, but we do need to support it so we're working with the currently available specs. We cannot keep changing the models constantly so a release timeline will be needed to make clear when a next version of the specs will be adopted. | M | Interoperabiliteit | Modellen | Translatie |
| REQ-016.11 | Support full DCAT-AP including other extensions | Because of combining datasets that are also outside of health. Include the properties that are common among different domains of DCAT, but that are not included in the core DCAT. | W | Interoperabiliteit | Modellen | Translatie |
| REQ-016.12 | Support for a REST-API , support of endpoints of other communication protocols, such as OAI-PMH | Enable machine-machine communication to Create, Read, Update and Delete (CRUD) metadata records | S | Betrouwbaarheid | Toegang | Technische_interoperabiliteit |
| REQ-016.14 | Sustainable metadata schema | To keep repetitive workload of users/data holders/ researchers low. + backward compatibility | S | Betrouwbaarheid | Toegang | Automatisering beheer koppelingen |
| REQ-006.17 | Support for Lineage / Provenance metadata | W3C Prov. standard support for mutations on the (meta)dataset | M | Interoperabiliteit | Traceerbaarheid | Lineage_Provenance |
| **From EHDS text to basic requirements - based on new text as known in November 2024:** https://www.europarl.europa.eu/meetdocs/2024_2029/plmrep/COMMITTEES/ENVI/DV/2024/12-04/2022_0140COR01_EN.pdf | | | | | | |
| REQ-017.1 | Data holders as specified by EHDS must be enabled to share their metadata with the HDAB catalogue, by themselves, **or** by | See chapter I Article 2 definition of data holder (page 119). Makes clear that this can also be done by intermediate entities. So, these must also be identified as stakeholders for the catalogue. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| | intermediate entities appointed by national law. | | | | | |
| REQ-017.2 | Extra data holders specified in national law must be enabled to share their metadata with the HDAB catalogue. | Dependent on additional national law. To be further investigated later. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-017.3 | Data holders must be enabled to share their datasets that are described by the minimum categories of electronic health data. | See chapter IV article 51 minimum categories of health data (page 208-209). As HealthDCAT-AP is still in development, this must also be revisited later and will need to be iteratively developed. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-017.4 | Data holders must be enabled to share their datasets of extra categories appointed by national law. | Dependent on additional national law. To be further investigated later. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-017.5 | Data holders must be able to correct, annotate or enrich data based on a permit depending on additional national rules | Dependent on additional national law. To be further investigated later. Not sure if the full scope of this statement affects the catalogue itself. But at the least the metadata itself in the catalogue needs to be managed in these cases. | M | Governance | Operationeel | Annoteren |
| REQ-017.6 | Data holders must be able to take measures and additional safeguards for categories of data under paragraph 1 points (f), (g), (i) and (q) in correspondence with additional national law | Dependent on additional national law. To be further investigated later. | M | Betrouwbaarheid | Toegang | Baseline Informatieveiligheid Overheid |
| REQ-017.7 | Data holders must be enabled to share metadata making clear extra protections for electronic health data protected by intellectual property rights, trade secrets or covered by the regulatory data protection right | At the latest following a request received from the health data access body. Needs to be investigated further how to make this sufficiently clear in the metadata. | M | Governance | Operationeel | Annoteren |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-017.8 | **HDAB shall make public through electronic means a national dataset catalogue that includes details about the source and nature of electronic health data, in accordance with Articles 77, 78 and 80, and the conditions for making electronic health data available;** | Core requirement for HDAB-NL WP6 Catalogue and Metadata | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-017.9 | (vii) information, at a minimum on an easily accessible website or web portal, on the connection to HealthData@EU of national contact points for secondary use of a third country, or of a system established at international level by an international organisation, as soon as the third country or the international organisation becomes an authorised participant in HealthData@EU. | Not sure if this applies to catalogue itself, so included in this list at least till we're sure it's not picked up elsewhere. Concerns transparency about the third countries that can become part of the ecosystem. | S | Governance | Operationeel | Webtoegankelijkheid |
| REQ-017.10 | The national dataset catalogue referred to in point (j)(i) of this paragraph shall also be made available to single information points under Article 8 of Regulation (EU) 2022/868. | Not sure yet what Article 8 of Regulation (EU) 2022/868 says. To be further investigated. | M | Datawaarde | Metadata | Vindbaarheid_technisch_en_discovery_Directory |
| REQ-017.11 | cooperate with all relevant stakeholders, including patient organisations, representatives of natural persons, health professionals, researchers, and ethics committees, where applicable in accordance with Union or national law; | This is required of the HDAB, so what the HDAB-NL program must start to do in relations to piloting the catalogue to give the HDAB a good basis. | M | Governance | Bestuurlijk | Eigenaarschap |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-017.12 | cooperate with other national competent bodies, including the national competent authorities supervising data altruism organisations under Regulation (EU) 2022/868, the competent authorities under Regulation (EU) 2023/2854 and the national competent authorities under Regulations (EU) 2017/745, (EU) 2017/746 and (EU) 2024/1689, where relevant. | These organisations may be relevant for sharing metadata in the catalogue, or in using it. | M | Governance | Bestuurlijk | Eigenaarschap |
| REQ-017.13 | Health data access bodies shall make information on the conditions under which electronic health data are made available for secondary use publicly available, easily searchable through electronic means and accessible for natural persons. That information shall cover the following: (a) the legal basis under which access to electronic health data is granted to the health data user; the technical and organisational measures taken to protect the rights of natural persons; | This part of Article 58 on page 229 could be part of the catalogue because it probably will be information incorporated in the metadata. That means the HDAB needs to make sure that data holders must fill this in when publishing their metadata. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-017.14 | **The health data holder shall communicate to the health data access body a description of the dataset it holds in accordance with Article 77. The health data holder shall, at a minimum on an annual basis, check that its dataset description in the national dataset catalogue is accurate and up to date.** | Core requirement for health data holder, but also for HDAB-NL WP6 Catalogue and Metadata as it needs to enable the data holder to do this. | M | Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus |
| REQ-017.15 | Where a data quality and utility label accompany the dataset pursuant to Article 78, the health data holder shall provide sufficient documentation to the health data access body for that body to verify the accuracy of the label. | Data Quality and utility - must be represented in the catalogue - needs to be investigated in relations to HealthDCAT-AP and how exactly the accuracy should be sufficiently documented. WP9 within the HDAB-NL program is leading for this and in close collaboration with the Quantum project. | M | Datawaarde | Metadata | Classificatie |
| REQ-017.16 | Article 77<br>Health data access bodies shall, through a publicly available and standardised machine-readable dataset catalogue, provide a description in the form of metadata of the available datasets and their characteristics. | So, the catalogue must be standardised and machine readable. | M | Betrouwbaarheid | Toegang | Technische_interoperabiliteit |
| REQ-017.17 | The description of each dataset shall include information concerning the source, scope, main characteristics, and nature of the electronic health data in the dataset and the conditions for making those data available. | Metadata requirement | M | Datawaarde | Metadata | Classificatie |
| REQ-017.18 | The dataset descriptions in the national dataset catalogue shall be available in at | So, this is about the descriptions, not the catalogue | M | Governance | Operationeel | Webtoegankelijkheid |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| | least one official language of the Union. | | | | | |
| REQ-017.19 | The dataset catalogue for Union institutions, bodies, offices and agencies provided by the Union health data access service shall be available in all official languages of the Union. | | n.a. | | | |
| REQ-017.20 | The dataset catalogue shall be made available to single information points established or designated under Article 8 of Regulation (EU) 2022/868. | Connectivity - need to check this article | M | Betrouwbaarheid | Toegang | Technische_interoperabiliteit |
| REQ-017.21 | By ... [two years from the date of entry into force of this Regulation], the Commission shall, by means of implementing acts, set out the minimum elements health data holders are to provide for datasets and the characteristics of those elements. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 98(2). | | M | Datawaarde | Metadata | Classificatie |
| REQ-017.22 | Datasets made available through health data access bodies may have a Union data quality and utility label applied by the health data holders. | So, the health data holders must be the ones describing the quality and utility of their data - but for them it's not mandatory to apply these labels, and it must be possible in the catalogue. | M | Datawaarde | Metadata | Classificatie |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-017.23 | The data quality and utility label shall cover the following elements, where applicable:<br>(a) for data documentation: metadata, support documentation, the data dictionary,<br>the format and standards used, the source of the data and, where applicable,<br>the data model;<br>(b) for assessment of technical quality: completeness, uniqueness, accuracy, validity, timeliness and consistency of the data;<br>(c) for data quality management processes: the level of maturity of the data quality management processes, including review and audit processes, and bias examination;<br>(d) for assessment of coverage: the period, population coverage and, where<br>applicable, representativity of the population sampled, and the average<br>timeframe in which a natural person appears in a dataset;<br>(e) for information on access and provision: the time between the collection of the<br>electronic health data and their addition to the dataset and the time needed to<br>provide electronic health data following the issuing of a data | Metadata requirements | M | Datawaarde | Metadata | Classificatie |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| | permit or a health data request approval; (f) for information on data modifications: merging and adding data to an existing dataset, including links with other datasets. | | | | | |
| REQ-017.24 | 4. Where a health data access body has reason to believe that a data quality and utility label might be inaccurate, it shall assess whether the dataset covered by the label meets the quality requirements forming part of the elements of the data | It must be possible for the HDAB to **revoke** data quality labels, this means authorised HDAB staff users must be able to transparently adjust registration of quality and utility, of course assuming this must also include communication to the data holder. | M | Datawaarde | Metadata | Classificatie |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| | quality and utility label as referred to in paragraph 3 and, in the event the dataset does not meet the quality requirements, shall revoke the label. | | | | | |
| REQ-017.25 | The Commission is empowered to adopt delegated acts in accordance with Article 97 to amend this Regulation by modifying, adding or removing elements to be covered by the data quality and utility label provided for in paragraph 3. | Meaning version management of data and utility labels must be possible (this means metadata model maintenance but may also be catalogue UI development). | M | Interoperabiliteit | Dataservices | Versiebeheer |
| REQ-017.26 | By ... [two years from the date of entry into force of this Regulation], the Commission shall, by means of implementing acts, set out the visual characteristics and technical specifications of the data quality and utility label, based on the elements referred to in paragraph 3 of this Article. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 98(2) of this Regulation. Those implementing acts shall take into account the requirements in Article 10 of Regulation (EU) 2024/1689 and any adopted common specifications or harmonised standards supporting those requirements, where applicable. | This means iterative development taking into account newly published implementing acts and keeping in close cooperation with the TEHDAS2 project to see them coming and think along with them. Including incorporating the results from the Quantum project. | M | Interoperabiliteit | Dataservices | Versiebeheer |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| REQ-017.27 | The Commission shall establish an EU dataset catalogue connecting the national dataset catalogues established by the health data access bodies in each Member State as well as the dataset catalogues of authorised participants in HealthData@EU. | National catalogue needs to connect to the EU one and authorised participants will have their own dataset catalogue to connect | M | Betrouwbaarheid | Toegang | Technische_interoperabiliteit |
| REQ-017.28 | 2. The EU dataset catalogue, the national dataset catalogues and the dataset catalogues of authorised participants in HealthData@EU shall be made publicly available. | Anyone must be able to find the catalogue. | M | Governance | Operationeel | Webtoegankelijkheid |
| REQ-017.29 | Article 80<br>Minimum specifications for datasets of high impact<br>The Commission may, by means of implementing acts, determine the minimum specifications for datasets of high impact for secondary use, taking into account existing Union infrastructures, standards, guidelines and recommendations. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 98(2). | Specific metadata scheme and maybe other actions to take for datasets of high impact, to be further investigated. | M | Datawaarde | Metadata | Classificatie |
| **Other legislation** | | | | | | |
| | From Direct Grant: system will be piloted based on systems available at the consortium partners | The appointed HDAB in the end will be able to make final decisions on the systems to use. Those decisions cannot be made within the HDAB-NL project. | M | Governance | Bestuurlijk | Eigenaarschap |

| Requirement ID | Features | Additional description | MoSCoW for Piloting | RS37 Match - 1e ring: Kernwaarden | RS37 Match - 2e ring: Functie | RS37 Match - 3e ring: Vereiste en Specificaties |
|---|---|---|---|---|---|---|
| | The catalogue must be maintained and further developed in compliance with the implementing acts that will be published in relation to the EHDS. | | M | Betrouwbaarheid | Toegang | Policies |
| | The catalogue must be in line with current legislation, like at this moment we can see the following relevant laws and regulations: EHDS, GDPR, data act, data governance act, NIS2, eIDAS, and in NL: the UAVG, WGBO, de algemene wet bestuursrecht, de wet hergebruik overheidsinformatie. | | M | Betrouwbaarheid | Toegang | Policies |
| **Other** | | | | | | |
| | The catalogue must be maintained and developed to stay in line with future legislation. | | M | Betrouwbaarheid | Toegang | Policies |
| | The catalogue must be based on best practices and current state of technology. | | M | Betrouwbaarheid | Toegang | Policies |
| | The catalogue should adhere to RS37 requirements that are applicable to data catalogues, or it must be clearly explained why this is not feasible or desirable | This RS37 framework offers us an overview of requirements that are commonly needed or required for government infrastructures; therefore, we expect that also this national catalogue needs to adhere to the same principles. We will still be further investigating the requirements framework and determine which of the categories apply to the catalogue. | S | Betrouwbaarheid | Toegang | Policies |

## 18.2 Table 3: (Meta)data standards inventory

This inventory is (partly) based on stakeholder input, as stakeholders expressed the importance of considering the standards out there during the HDAB-NL kick-off meeting. It helps us to keep track of the information that should be taken into account and/or displayed in the metadata in the catalogue. How to deal exactly with this and UI aspects of that are still an open discussion. Harmonizing on standards and models used is part of improving data quality and findability of relevant subjects across datasets and is therefore essential for the national data catalogue to make visible and machine readable in a useful way.

| Name | Short description | Metadata standard or data model |
|---|---|---|
| OMOP common data model | Data model used to harmonise data from different sources, like EHRs and claims data. To facilitate research of health conditions. | Data model |
| I2b2 (Informatics for Integrating Biology and the Bedside) | Platform assisting researchers to integrate and analyse clinical and genetic data. | Data model |
| CDISC (Clinical Data Interchange Standards Consortium) ODM (Operational Data Model) & SDTM (Study Data Tabulation Model) | Standard for submission of clinical health data to instances like the FDA. Helps to structure clinical trial data. | Data standard |
| OpenEHR | Standard for exchange of EHR (electronic health record) data. | Data standard |
| SNOMED-CT | Clinical terms and definitions of clinical concepts. | Ontology |
| ICD 10 / ICD 11 | Clinical terms and definitions, especially diseases and conditions. | Ontology |
| LOINC | Terminology standard for health measurements, lab observations, and documents. | Ontology |
| UCUM | Coding system for units of measures. | Classification system |
| IDMP | 5 ISO standards to 'specify the use of standardised definitions for the identification and description of medicinal products for human use.' | Data standard |
| HL7 FHIR | Technical standard for exchange of data in healthcare (on record level, not on dataset level). | Data standard |
| HL7 CDA | HL7 Clinical Document Architecture | Data standard |

| Name | Short description | Metadata standard or data model |
|---|---|---|
| DCM | Detailed Clinical Models. | Data model |
| ZIBs | Zorg Informatie Bouwstenen (NL variant van DCMs) | Data model |
| DCAT-AP v3 | DCAT application profile, to be used in EU data portals. | Metadata standard (EU) |
| DCAT-AP NL | Dutch specification of DCAT-AP v3. | Metadata standard (NL) |
| HealthDCAT-AP | (Draft version!) of health extension DCAT-AP v3, to be used EU level for the EHDS. | Metadata standard for health related datasets and services. |
| DICOM | Medical imaging standard. | Data standard |
| MIABIS | Biobanks standard (BBMRI). | Information model |
| See https://www.iso.org/standard/73172.html | ISO 23081-1:2017 Information and documentation — Records management processes — Metadata for records | |
| DSC (meta)data model | Model to describe datasets from Statistics Netherlands (CBS: Central bureau for statistics, Statistics Netherlands). | |
| ISO910115 | Metadata profile for geo data. | |
| ISO19110 | | |
| ISO19139 | | |
| GEODCAT | | Metadata |
| DDI | This specification is designed to support the discovery of microdata sets and related metadata using RDF technologies in the Web of Linked Data' | Metadata |
| Kwaliteitsstandaarden | Quality standards. | Data standard |
| Declaratiestandaarden | Declaration standards. | Data standard |

| Name | Short description | Metadata standard or data model |
|---|---|---|
| ICPC | International classification of primary care. Classification system for primary care encounters. 4 elements: reason for encounter, diagnosis/health problem, functioning, processes of care. Borrows from ICD-10/11, SNOMED | Classification system |
| TNM | Cancer staging system - beschrijft de grootte van tumor, spreiding naar lymphe nodes en metastases | Classification system / data model |
| ICCR | International collaboration on cancer reporting. Develop internationally standardised and evidenced-based checklists for the pathology reporting of cancer | Data standard |
| DataCite Metadata schema | Identification of a resource. A mapping from Datacite to DCAT is available. | Metadata |
| Thesaurus Zorg en Welzijn | Collection of terms from healthcare, social domain and other aspects of wellbeing (living, education). | Glossary |
| ICD-O-3 | International Classification of Diseases for Oncology | Classification system |
| Dublin Core | Standaard voor Metadata (internationaal) https://www.dublincore.org/specifications/dublin-core/ | Metadata |

# 19 Table 4: Overview of Non-functional requirement categories to keep in mind

This list is meant to be used when evaluating solutions to make sure all aspects that are needed have been considered.

Accessibility
Adaptability
Auditability and control
Availability (see service level agreement)
Backup
Boot up time
Capacity, current and forecast
Certification
Compliance
Configuration management
Conformance
Cost, initial and Life-cycle cost
Data integrity
Data retention
Dependency on other parties
Deployment
Development environment
Disaster recovery
Documentation
Durability
Efficiency (resource consumption for given load)
Effectiveness (resulting performance in relation to effort)
Elasticity
Emotional factors (like fun or absorbing or has "Wow! Factor")
Environmental protection
Escrow
Ethics
Exploitability
Extensibility (adding features, and carry-forward of customizations at next major version upgrade)
Failure management
Fault tolerance (e.g. Operational System Monitoring, Measuring, and Management)
Flexibility (e.g. to deal with future changes in requirements)
Footprint reduction - reduce the exe files size
Integrability (e.g. ability to integrate components)
Internationalization and localization

Interoperability
Legal and licensing issues or patent-infringement-avoidability
Maintainability (e.g. mean time to repair – MTTR)
Management
Memory Optimization
Modifiability
Network topology
Open source
Operability
Performance / response time (performance engineering)
Platform compatibility
Privacy (compliance to privacy laws)
Portability
Quality (e.g. faults discovered, faults delivered, fault removal efficacy)
Readability
Reliability (e.g. mean time between/to failures – MTBF/MTTF)
Reporting
Resilience
Resource constraints (processor speed, memory, disk space, network bandwidth, etc.)
Response time
Reusability
Robustness
Safety or factor of safety
Scalability (horizontal, vertical)
Security (cyber and physical)
Software, tools, standards etc. Compatibility
Stability
Supportability
Testability
Throughput
Transparency
Usability (human factors) by target user community
Volume testing

## 20 Table 5: RS037 requirements framework (In Dutch)

This is a framework used by government in the Netherlands to categorise requirements. The application requirements in table 2 have been marked for the best fitting categorisation from this framework.

| L1 - BY DESIGN - WAT KAN IK? | L2 - WAT ZOEK IK? | L3 - WAT MOET IK? ANTW=VEREISTE EN OF SPECIFICATIE | RS# |
|---|---|---|---|
| Governance | Bestuurlijk | Eigenaarschap | 1 |
| Governance | Bestuurlijk | Financieringsconstructen | 2 |
| Governance | Operationeel | Annoteren | 3 |
| Governance | Operationeel | Dataleveringsovereenkomsten | 4 |
| Governance | Operationeel | Gegevensbescherming_Security | 5 |
| Governance | Operationeel | Service_levels | 6 |
| Governance | Operationeel | Terugmelden_onderzoeken | 7 |
| Governance | Operationeel | Webtoegankelijkheid | 8 |
| Governance | Beheer | Duurzame_toegangkelijkheid | 9 |
| Governance | Beheer | Schaalbaarheid | 10 |
| Vertrouwen | Identiteit | Identificatie | 11 |
| Vertrouwen | Identiteit | Authenticatie | 12 |
| Vertrouwen | Identiteit | Delegatie | 13 |
| Vertrouwen | Identiteit | TLScertificaten_ACME | 14 |
| Vertrouwen | Toegang | Policies | 15 |
| Vertrouwen | Toegang | Baseline Informatieveiligheid Overheid | 16 |
| Vertrouwen | Toegang | Autorisatie patronen | 17 |
| Vertrouwen | Toegang | Anonimiseren_pseudonimiseren | 18 |
| Vertrouwen | Toegang | Dataversleuteling_encryptie | 19 |
| Vertrouwen | Toegang | Hardening | 20 |
| Vertrouwen | Toegang | Technische_interoperabiliteit | 21 |
| Vertrouwen | Toegang | Validators | 22 |
| Vertrouwen | Toegang | Automatisering beheer koppelingen | 23 |
| Datawaarde | Metadata | Classificatie | 24 |
| Datawaarde | Metadata | Vindbaarheid_technisch_en_discovery_Directory | 25 |
| Datawaarde | Verantwoording | Logging | 26 |
| Datawaarde | Publicatie | Vindbaarheid_semantisch_Catalogus | 27 |
| Interoperabiliteit | Modellen | Translatie | 28 |

| | | | |
|---|---|---|---|
| Interoperabiliteit | Modellen | Historie_en_Tijdreizen | 29 |
| Interoperabiliteit | Traceerbaarheid | Lineage_Provenance | 30 |
| Interoperabiliteit | Dataservices | Zoeken | 31 |
| Interoperabiliteit | Dataservices | Orchestratie | 32 |
| Interoperabiliteit | Dataservices | Notificaties | 33 |
| Interoperabiliteit | Dataservices | Homomorfe encryptie | 34 |
| Interoperabiliteit | Dataservices | CAP_theorema | 35 |
| Interoperabiliteit | Dataservices | ACID_Base afwegingen | 36 |
| Interoperabiliteit | Dataservices | Versiebeheer | 37 |