



Ministerie van
Volksgezondheid,
Welzijn en Sport

Samen klaar voor de HDAB:
van weten naar doen

HDAB-event
21 september 2026



Denk en praat mee.





Breakout: Datakwaliteit voor secundair gebruik

Sprekers:

Ellen Leenarts (RIVM / HDAB-NL)

Floor van Oers (CBS / HDAB-NL)

Georganiseerd en ondersteund door WP9-HDAB-NL:

Jeroen Beliën (Health-RI)

Dianne van Essen (ICTU)

Sophie van Wingerden, Stijn Voerman (RIVM)

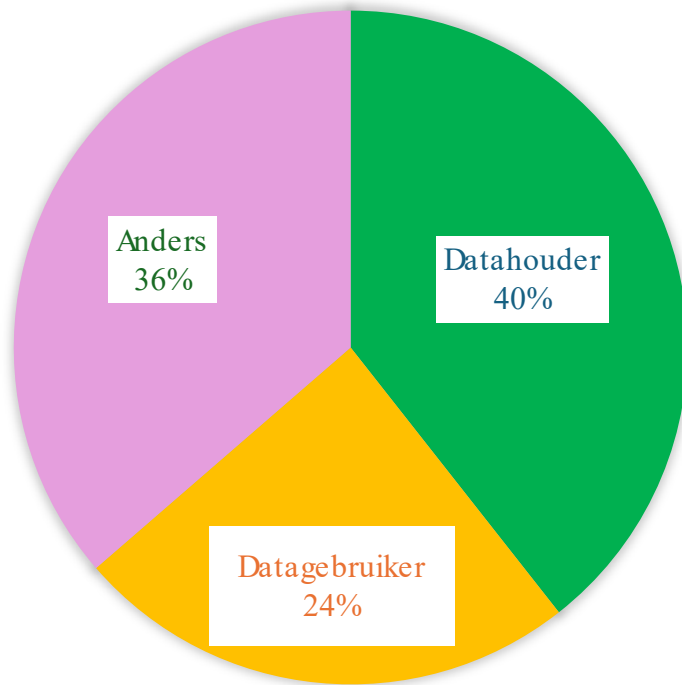


Inhoud

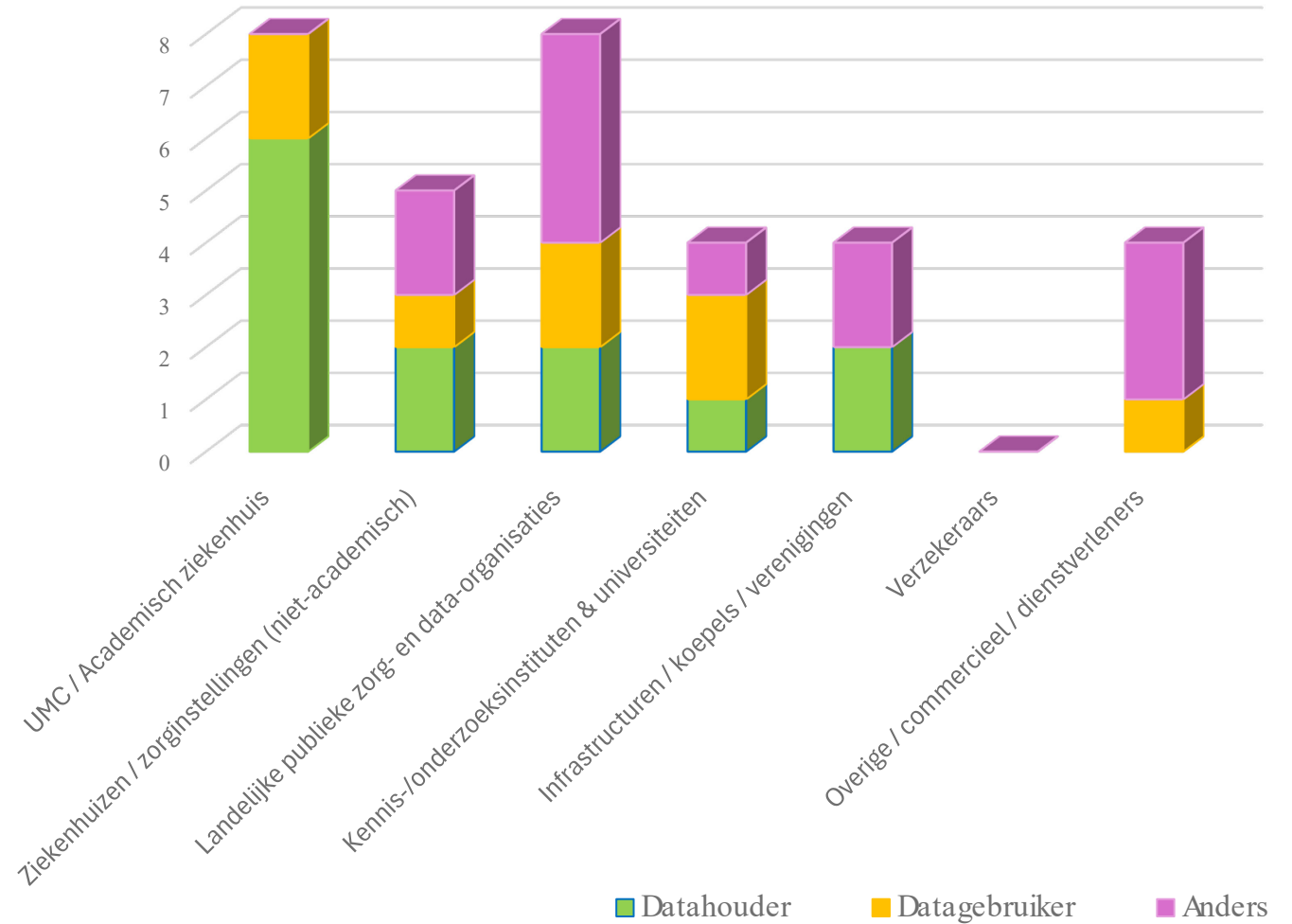
- Wie zijn wij?
- Waar staan we?
- Hoe maak ik een EHDS-compatibel datakwaliteit- en bruikbaarheidslabel. De QUANTUM labelling tool
- Praktijkoefening kwaliteitslabel
- Vooruitblik



Wie zijn jullie?

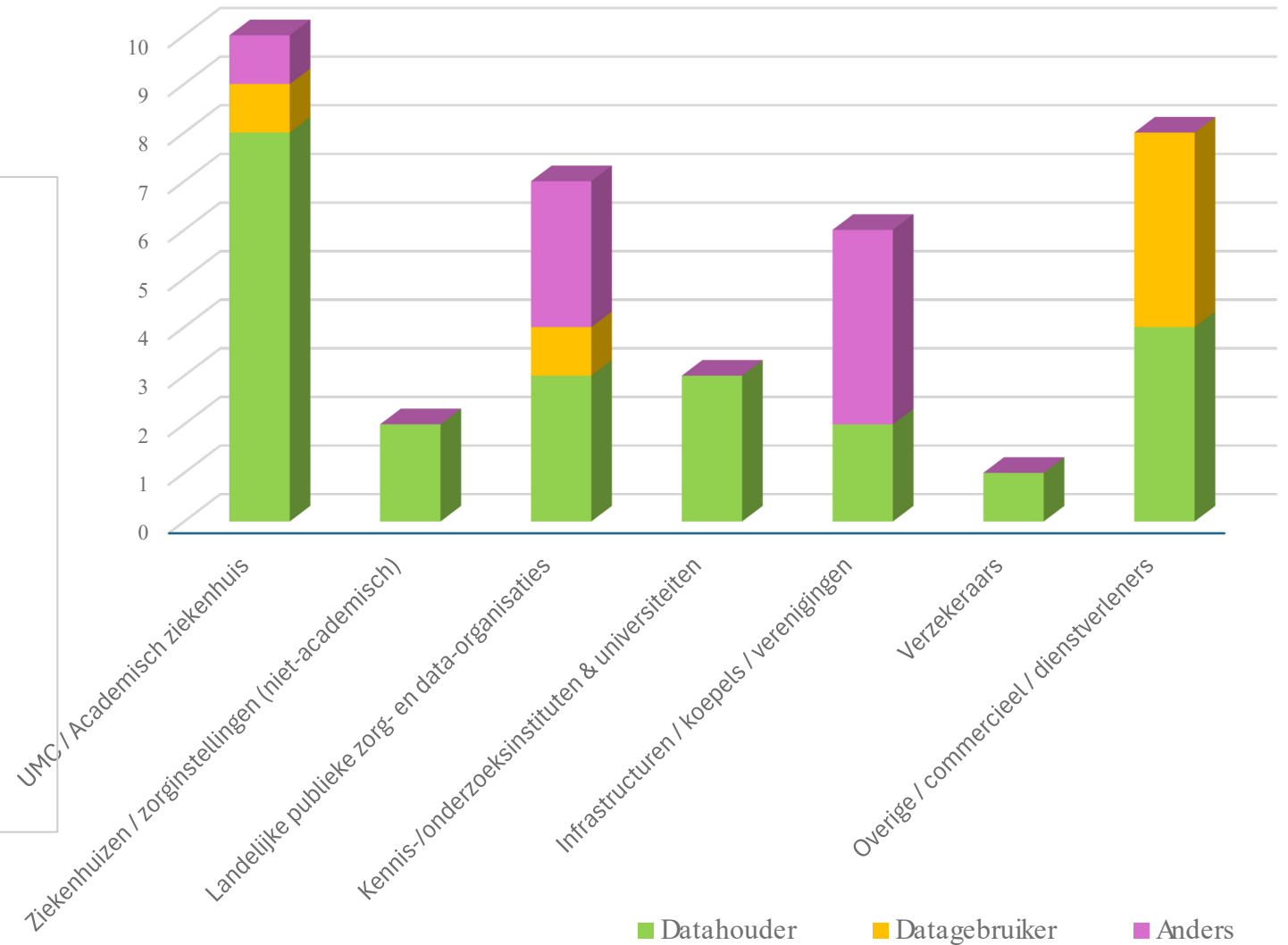
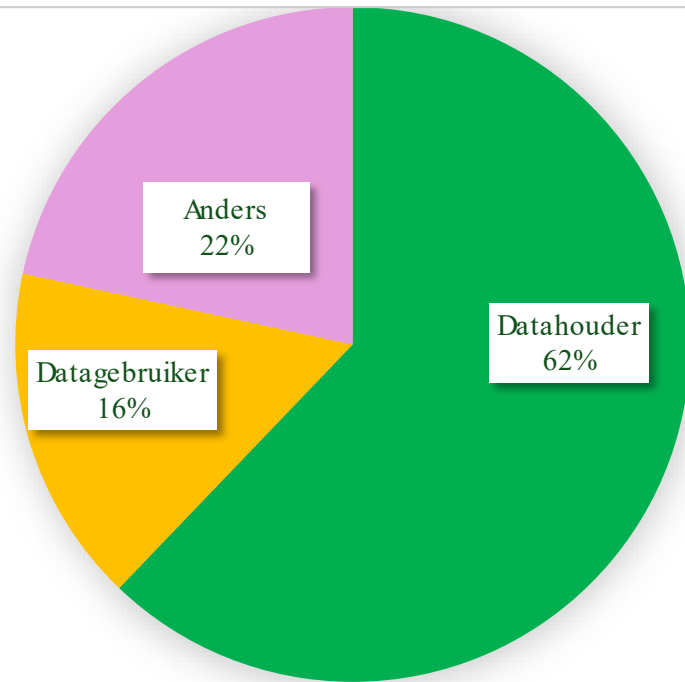


Deelnemers Sessie 1



Wie zijn jullie?

Deelnemers Sessie 2



European Health Data Space - **EHDS**

Europese Data Spaces

implementing act = uitvoeringshandelingen

Nederland bereidt zich voor
in het **HDAB-NL** project

Werkpakket 9 Datakwaliteit

QUANTUM ontwikkelt raamwerk
en tool voor datakwaliteit en
bruikbaarheidslabel

input en testen tool

Taak WP9: Advisering

- Toekomstige HDAB
(wordt onderdeel van de GDA)
- Datahouders
- Datagebruikers

Met:

- Informatiesessies en trainingen
- Overzicht trainingsmaterialen
datakwaliteit EHDS
- Extra materialen:
aanbevelingen, beslisboom,
stappenplan.

**Literatuurstudie en online
onderzoek datakwaliteit**
(raamwerken, EHDS, etc)

Stakeholdersessies
datakwaliteit

Definitie datakwaliteit

Data zijn van **hoge kwaliteit** wanneer ze voldoen aan de **verwachtingen en behoeften** van de **datagebruikers**: de gegevens zijn dan geschikt voor de doeleinden waarvoor zij deze willen gebruiken. Gegevens zijn van lage kwaliteit wanneer ze niet geschikt zijn voor dat doel. **Datakwaliteit is dus contextafhankelijk.**

De EHDS definieert datakwaliteit:

*"de mate waarin de elementen van elektronische gezondheidsgegevens geschikt zijn voor het **beoogde primaire gebruik, en secundaire gebruik** ervan".*

Met andere woorden, datakwaliteit is de mate waarin digitale gezondheidsgegevens zó compleet, juist, actueel en duidelijk is vastgelegd, dat zorgverleners én anderen (zoals onderzoekers of beleidsmakers) deze veilig en goed kunnen gebruiken voor hun doel.



Definitie datakwaliteitslabel

De EHDS definieert het datakwaliteitslabel als volgt:

*“een **grafisch diagram**, met **inbegrip van een schaal**, waarin de gegevenskwaliteit en de gebruiksvoorwaarden van een dataset worden beschreven”*

Met andere woorden, dit kwaliteitslabel is een soort **etiket** bij je dataset. In één duidelijk plaatje laat het twee dingen zien:

- **Hoe goed de gegevens zijn:** zijn ze compleet, recent en betrouwbaar? En is daar documentatie van.
- **Hoe gaat de organisatie met de data om (datavolwassenheid):** hoe goed wordt de data verzamelt, beheert, beveiligt en gebruikt.



Van elementen voor datakwaliteit -> naar label

Samen vormen deze elementen visueel een gestandaardiseerd label dat transparant maakt hoe betrouwbaar, compleet en bruikbaar een dataset is.

Element	Uitleg
Documentatie	Metagegevens, ondersteunende documentatie, het gegevenswoordenboek, het gebruikte format en de gebruikte normen, de bron van de gegevens en, indien van toepassing, het gegevensmodel
Technische kwaliteit	de volledigheid, het unieke karakter, de nauwkeurigheid, geldigheid, tijdigheid en consistentie van de gegevens;
Processen voor het beheer	de rijpheid van de processen voor het beheer van de gegevenskwaliteit, met inbegrip van evaluatie- en auditprocessen, en onderzoek naar vertekening (bias)
De dekking	de periode, bestreken populatie en, indien van toepassing, representativiteit van de bemonsterde populatie, en de gemiddelde tijdsperiode waarin een natuurlijke persoon in een dataset voorkomt
Informatie over toegang en verstrekking	de tijd tussen het verzamelen van de elektronische gezondheidsgegevens en de toevoeging ervan aan de dataset en de tijd die nodig is voor het verstrekken van elektronische gezondheidsgegevens na het verlenen van een gegevensvergunning of een goedkeuring van een verzoek om gezondheidsgegevens
Informatie over wijzigingen van de gegevens	samenvoeging van datasets en toevoeging van gegevens aan een bestaande dataset, met inbegrip van koppelingen met andere datasets.

Geleerde lessen tot nu toe

- **Het QUANTUM project** heeft **mooie resultaten** opgeleverd: **raamwerk** datakwaliteit voor secundair gebruik uitgaande van de EHDS met het datakwaliteit en bruikbaarheids**label**, handleidingen, een **tool** en **trainingsmateriaal**.
- Door de QUANTUM **ontwikkelingen** te **volgen** en de tool te testen weten we goed wat het wel en niet is.
- Belangrijk om aan te sluiten bij **Europese en nationale ontwikkelingen op het gebied van datakwaliteit**, zoals de HealthData@EU infrastructuur, Community of Practice van voortgang verschillende lidstaten op dit onderwerp, HealthDCAT-AP metadatastandaard, TEHDAS2 rapporten, en nationale initiatieven zoals een nieuw rijksbreed datakwaliteitskader, NORA datakwaliteitskader / tool en volwassenheidstabel, Wegiz en de IBDS en FDS rondom datakwaliteit.
- **Samenwerking en afstemming, ook met Nederlandse stakeholders** blijken cruciaal voor een goed beeld van de behoeften vanuit het zeer diverse datalandschap in de zorg.
- Door veel en vaak in gesprek te gaan met verschillende stakeholders hebben we een **concept** overzicht van **praktische stappen** voor datahouders in voorbereiding op datakwaliteit voor secundair gebruik samengesteld.

QUANTUM project

- Europees project om artikel 78 te ondersteunen
- Levert na invulling een datakwaliteitslabel op, zowel als RDF als een PDF
- Twee onderdelen
 - o Datakwaliteit: beoordeling van een dataset
 - o **Volwassenheid (maturity)**: beoordeling van procesbeheersing datakwaliteit
- Er is géén verbinding met de data nodig
- NB: Vanuit Europa is nog niet bepaald of dit de standaard wordt.



Onderdelen datakwaliteit

The 12 QUANTUM Data Quality & Utility Dimensions

Understanding how QUANTUM evaluates data quality and utility in 12 key dimensions

COVERAGE

Is the data representative?

Population Coverage

Dataset includes all potential eligible population members.

Population Representativity

Data adequately represents the target population.

Accessibility

Clear and transparent access and usage conditions.

Data Provenance

Detailed documentation of data generation and validation.

Compliance

Data adheres to ethical standards and regulations.

Accuracy

Data correctly measures what it is designed to measure.

Precision

Data approximates reality with a high degree of accuracy.

Coherence

Data maintains uniformity in representation and meaning over time.

Consistency

Data attributes are plausible and uniform with other data.

Validity

Data representations conform to data model specifications.

Completeness

All potentially available information is present in the dataset.

QUANTUM
THE HEALTH DATA QUALITY LABEL

TECHNICAL QUALITY

Is the data reliable?

DATA DOCUMENTATION

Is the data well documented?

www.quantumproject.eu

Funded by
the European Union

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 European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

Datakwaliteit tool

f.vanoers's DQ&U Dashboard

How to Use the Dashboard?

Dataset & Catalogue Definitions

My data catalogues

Title	Version	Modify	Delete
CBS-Gezondheid en zorg	1.0	Modify	Delete
Create catalogue			

My datasets

		Assessment in progress Assessment done			
Catalogue	Name	Assessment	Label	RDF	Actions
CBS-Gezondheid en zorg	Causes of death Netherlands	Edit	★★★★☆		+
CBS-Gezondheid en zorg	Overledenen: doodsoorzaak (uitgebreide lijst), leeftijd, geslacht	Edit	★★★★☆		+
CBS-Gezondheid en zorg	Microdata GGZ DBC trajecten	Edit	★★★★☆		+
Create dataset					

Datakwaliteit tool

Dataset information

Catalogue CBS-Gezondheid en zorg	Name Causes of death Netherlands
Version 1.0	URI
Description In this dataset the causes of death of inhabitants of the Netherlands are recorded in ICD10	

Quality & Utility Assessment

How to Use the Quality & Utility Assessment?

Evaluate and report the following Data Quality & Utility Dimensions, within the given EHDS categories, to obtain the Dataset Quality & Utility Label

Unfold all

Fold all

To be filled

Filled

100%	
1. Access and provision	▼
2. Coverage	▼
3. Data documentation	▼
4. Technical quality	▼

Save changes

Go to DQ&U Dashboard

Download RDF

Go to DQ&U label



Datakwaliteit tool

Quality & Utility Assessment

[How to Use the Quality & Utility Assessment?](#)

Evaluate and report the following Data Quality & Utility Dimensions, within the given EHDS categories, to obtain the Dataset Quality & Utility Label

Unfold all

Fold all

To be filled

Filled

100%

1. Access and provision

2. Coverage

2.1. Population coverage

Definition: Population coverage refers to the degree to which a dataset includes the potential eligible population.
Dimension Weight: 6.47%.

Metric #1 - Coverage Rate

Definition

Coverage Rate (percentage of the eligible population represented in the dataset)

Recommended Measurement Approach

According to data holder information: $(\text{Number of individuals in the dataset} / \text{Total eligible population}) \times 100\%$

95-100%: Near-universal or universal coverage

<80%: Limited coverage

80-90%: Good coverage

90-95%: Very good coverage

95-100%: Near-universal or universal coverage

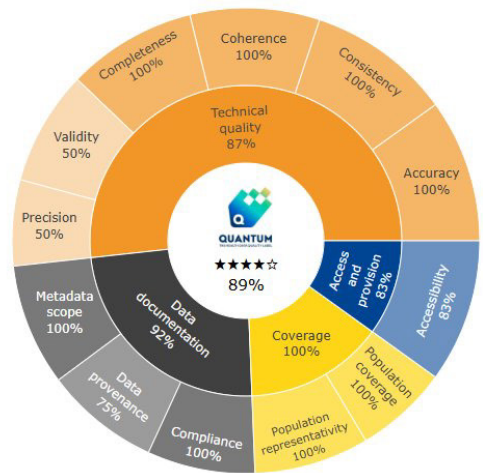
2.2

3. Data documentation

Datakwaliteit tool

Catalogue	Name
CBS-Gezondheid en zorg	Causes of death Netherlands

This label shows a completed quality assessment.



The Maturity label and the DQ&U label are completely independent. Neither is calculated based on the other.

Quality & Utility Assessment score



Total Q&U Score: 89.39 / 100

Maturity score: 64.0% (32/50)





QUANTUM Data Holders' Maturity Framework

A framework to assess readiness for health data reuse at organisational or workflow level, across 10 key dimensions, each assessed through 5 progressive maturity levels.



1. Data collection process

Processes relating to the capture of data from multiple sources for secondary purposes.



2. Data management – governance

Level of maturity of data management processes.



3. Data management – infrastructure

Level of implementation and development of data holder's data management infrastructure.



4. Data provenance

Clear description of dataset lineage providing transparent and comprehensive documentation of the data pipeline to get the dataset.



5. Data access

Definition and implementation of legal, ethical, and technical data access processes.



6. Data analytics environment

Availability of analytical services, tools, and access to secure data environments.



7. Data enhancement – augmentation

Application of techniques to improve data usability for specific purposes.



8. Data enhancement – enrichment

Addition of derived information such as annotations, labels, or NLP-derived data labels.



9. Data model

Availability of clear and documented data models to structure and standardise data.



10. Data dictionary

Availability of documented data dictionary, terminologies..



MATURITY LEVELS

LEVEL 1 INITIAL

No documentation/
processes

LEVEL 2 EMERGING

Basic, partial or ad hoc
processes

LEVEL 3 DEFINED

Standardised
(interoperable) processes

LEVEL 4 MANAGED

Partially automated,
proactively managed processes

LEVEL 5 OPTIMISED

Automated processes, and
continuous improvement

Volwassenheidstool

Evaluate and report the following Maturity Dimensions, to compute the organization maturity score

Unfold all

Fold all

To be filled

Filled

100%

1. Data collection process



2. Data management - governance



3. Data management - infrastructure



4. Data provenance



5. Data access



6. Data analytics environment



7. Data enhancement - augmentation



8. Data enhancement - enrichment



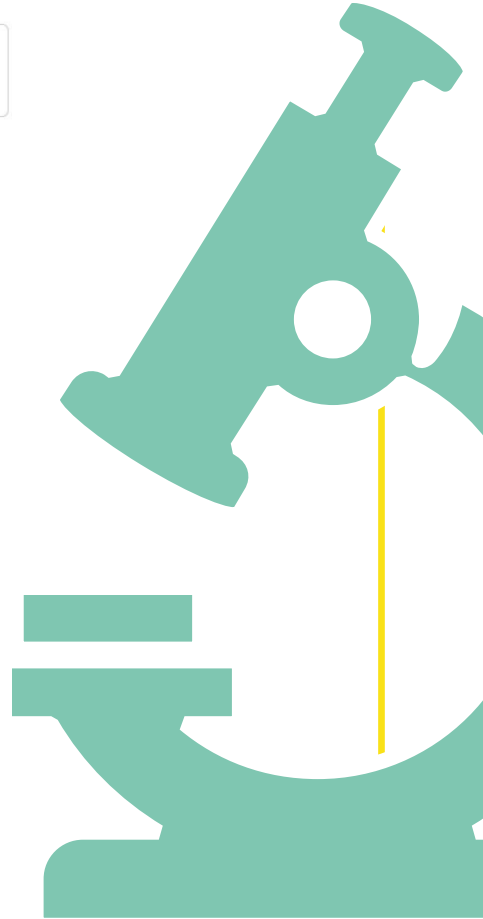
9. Data Model



10. Data Dictionary



Save changes



Volwassenheidstool

Evaluate and report the following Maturity Dimensions, to compute the organization maturity score

Unfold all

Fold all

To be filled

Filled

100%

1. Data collection process

2. Data management - governance

3. Data management - infrastructure

4. Data provenance

7. Data enhancement - augmentation

8. Data enhancement - enrichment

9. Data Model

10. Data Dictionary

Definition: Clear description of source and history of the dataset, providing a "transparent data pipeline"

Source of the dataset and any transformations, rules and exclusions

No documented provenance

Source of the dataset is documented

Source of the dataset and any transformations, rules and exclusions

All original data items listed, all transformations, rules and exclusion listed and impact of these

Ability to view earlier versions, including "raw" or "source" dataset, and review the impact of each stage/step



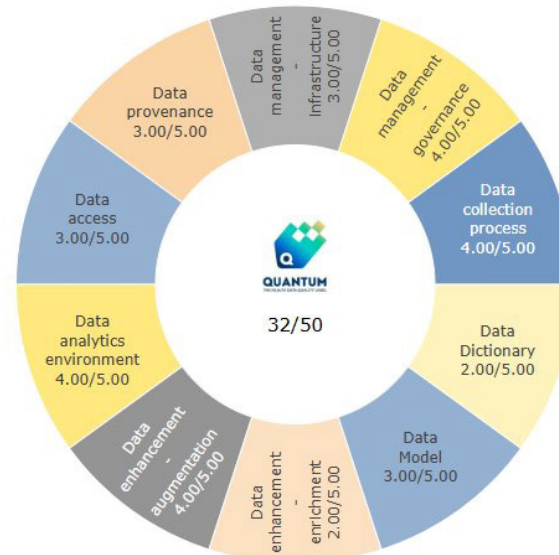
Volwassenheidstool

My organization - CBS

Maturity Score

Maturity is a feature at the data holder level. It refers to the level of automation of data and data quality management procedures.

Current score: **64.0%** = **32** / 50



The colours used in the plot are solely for visual differentiation of the maturity dimensions and do not carry any specific meaning.



Aan de slag!

- Volwassenheidslabel invullen
- Groepen van ca. 5 personen



Volwassenheidslabel invullen

- 5 A4tjes met de onderdelen, met een scoreblad.
- Focus op de drie gemarkeerde vragen.
- Kies één organisatie om de (gemarkeerde vragen) over volwassenheid in te vullen
- Beantwoord de vragen op de flipover



Vragen datahouders

- Wie zou dit volwassenheidslabel moeten en kunnen invullen (welke rol)?
- Is er training nodig om dit volwassenheidslabel te kunnen invullen?
 - Zo ja, op welk gebied?
- Biedt dit label aangrijpingspunten om de datakwaliteit te verbeteren?



Labels interpreteren

- Ga uit van een databehoeftte met meerdere mogelijke datasets.
- 4 labels over datakwaliteit
- Bekijk de labels en lees de toelichting.
- Beantwoord de vragen op de flipover



Vragen datagebruikers

- Is het verschil tussen de 4 datakwaliteitslabels duidelijk?
- Voegen de labels iets toe bij het kiezen voor een dataset?
 - Zo ja: welke onderdelen zijn voor jullie doorslaggevend?
- Is er voor het begrip van de labels training nodig?



Vragen datagebruikers

- Begrijp je de betekenis van dit label?
- Voegt het datavolwassenheidslabel iets toe bij het kiezen voor een dataset?

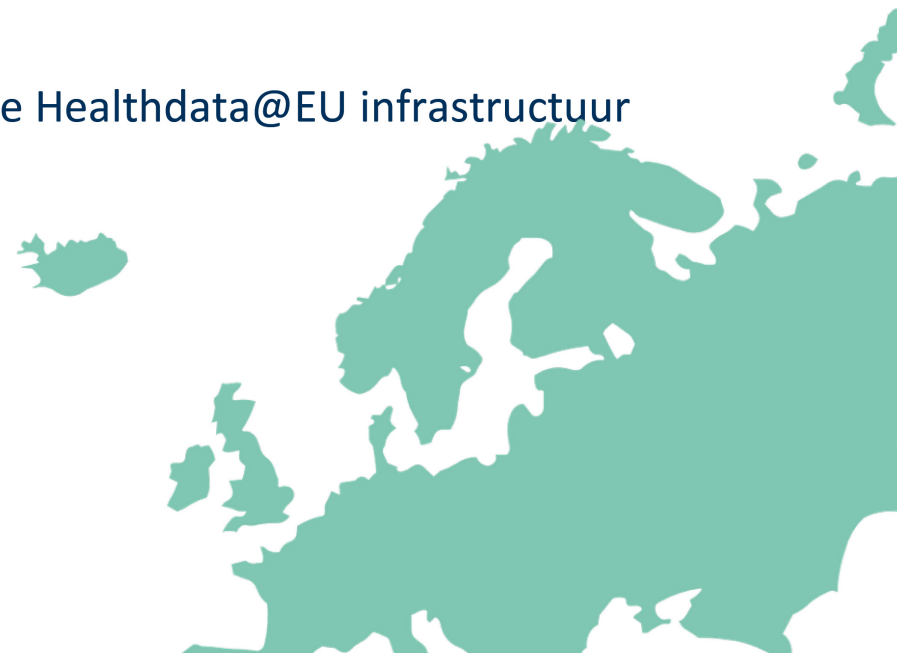


Onze leerpunten uit de testen tot nu toe

- Testen laat zien dat zelfs ervaren datahouders **metrics verschillend interpreteren**, wat kan leiden tot grote verschillen in labelscores. **Heldere definities en praktische voorbeelden zijn nodig** voor een consistente toepassing van datakwaliteitskaders, dimensies en tools.
- Er zijn **trainingsmodules nodig** om ervoor te zorgen dat organisaties en landen datakwaliteit op een consistente wijze interpreteren en beoordelen, zodat resultaten goed vergelijkbaar zijn.
- **“Fit-for-use” versus daadwerkelijke “fit-for-purpose”**. Het datakwaliteit en bruikbaarheidslabel geeft aan of een dataset geschikt is voor gebruik (“fit-for-use”), maar of een dataset daadwerkelijk voldoet aan een specifiek doel (“fit-for-purpose”) hangt af van de behoeften van de gebruiker, waardoor alleen het label onvoldoende is.
- Datahouders kennen hun datasets het beste en zijn daardoor het meest geschikt om te **beoordelen of een dataverzoek passend** is (“fit-for-purpose”) en om belangrijke beperkingen van de data te signaleren, wat essentieel is voor zinvolle en betrouwbare analyses.
- Gebruikersfeedback en extra **informatie-uitwisseling tussen datahouders en gebruikers**, lijken essentieel om de “fit-for-purpose” van datasets te vergroten
- Gebruikersreviews en labelaanpassingen op basis van **gebruikservaringen** kunnen belangrijk zijn voor vertrouwen in het label.

Vooruitblik

- HDAB-NL WP9
 - januari – juni 2027 Ontwikkeling trainingsplan en ondersteunende materialen voor stakeholders
 - aug. 2026 – februari 2027 Stakeholdersessies met beperkt aantal stakeholders
 - april 2027 Publiek webinar over uitkomsten stakeholdersessies en uitvoeringshandelingen
- Europa:
 - maart 2027 publicatie uitvoeringshandelingen
 - januari 2027 integratie van datakwaliteit en bruikbaarheidslabel in de Healthdata@EU infrastructuur





Bedankt

