



Health Data Access Body-NL





Webinar HDAB-NL

- Update werkpakket 9.
Datakwaliteit voor secundair gebruik
- 29 juni 2026

Denk mee. Praat mee.



Introductie

Doel van het webinar

- **Betrekken..**
..van stakeholders bij de HDAB en de ontwikkelingen
- **Informeren..**
..van geïnteresseerden over de stand van zaken en uitkomsten
- **Beantwoorden..**
..van vragen die er zijn rondom de (technische) onderdelen van de HDAB

Huisregels

- **Dit webinar wordt opgenomen**
Voor een rustiger beeld en geluid graag camera uit en geluid op mute.
- **Vragen?**
Geef ze aan via een handje of in de chat
- **Slides**
Deze worden na afloop gedeeld
- **Let op!**
Om een breed publiek te informeren kunnen we niet ingaan op alle details



Agenda

1. Achtergrond gegevenskwaliteit HDAB-NL
2. Raamwerken voor gegevenskwaliteit en definities van datakwaliteitsdimensies
3. Hoe maak ik een EHDS-compatibel datakwaliteit en bruikbaarheidslabel. De QUANTUM labelling tool.
4. Vooruitblik
5. Vragen of opmerkingen

Denk mee. Praat mee.



Doel HDAB-NL project

- In het kader van de EHDS moet elk EU-land een **Health Data Access Body (HDAB)** aanwijzen of oprichten. Deze HDAB krijgt vanaf maart 2029 met name de taak om besluiten te nemen over vergunningaanvragen t.b.v. het gebruik van data van bijvoorbeeld onderzoekers en beleidsmakers op basis van een streng EHDS-toetsingskader. Dit komt databeschikbaarheid ten goede, want waar een vergunning wordt verleend, moet data snel beschikbaar worden gemaakt.
- **Het project HDAB-NL** heeft als doel om op een duurzame manier, de (technische) functionaliteiten te ontwikkelen voor een nog aan te wijzen coördinerende Health Data Access Body. En geeft advies over de governance, het besluitvormingsproces en de implementatie.



Achtergrond | Health Data Access Body

EHDS

H2 | Primair gebruik

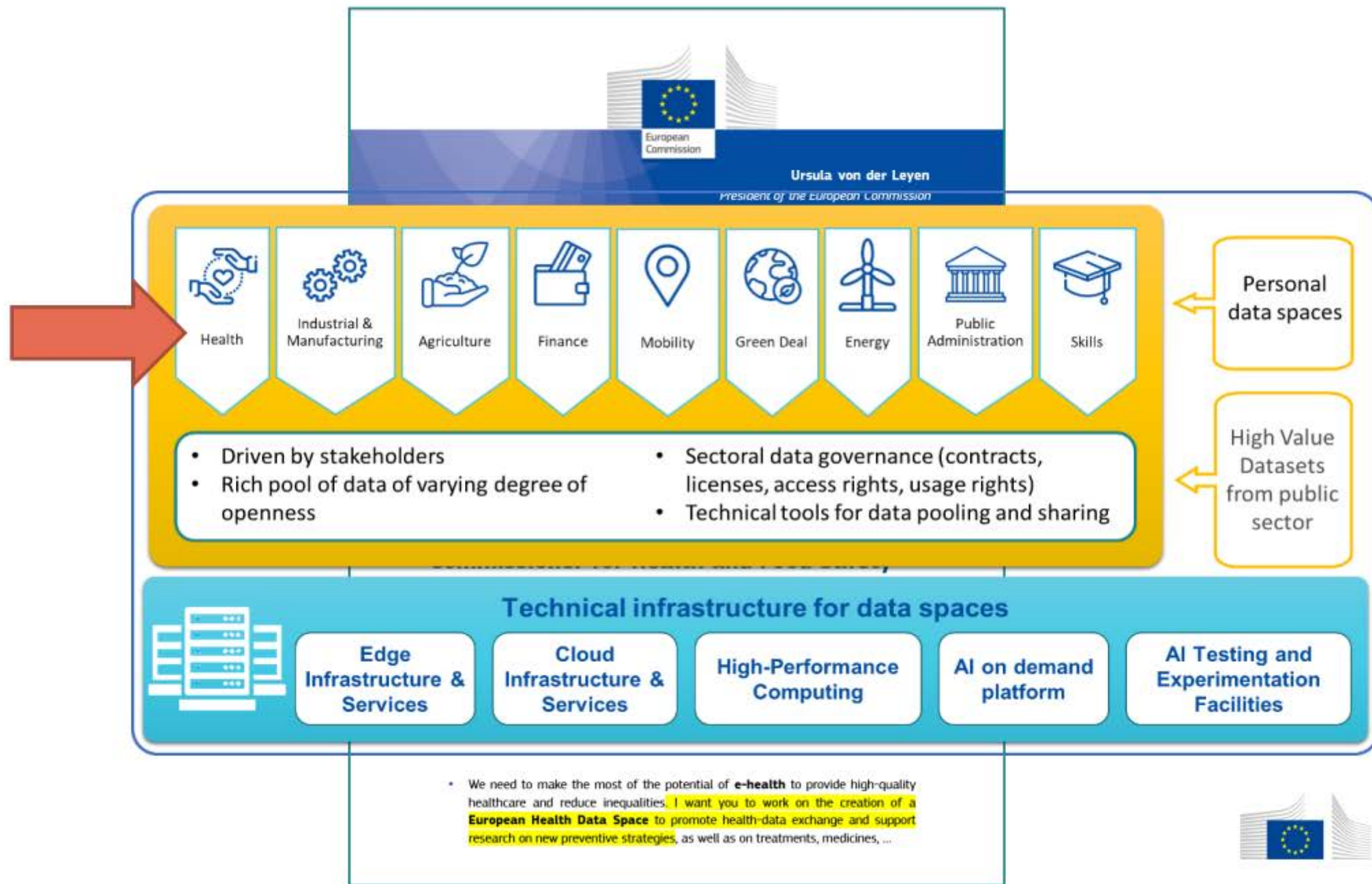
Hoe zorgen we voor de juiste gezondheidsdata voor de zorgverleners?

H3 | EPD-eisen

Hoe zorgen we voor een gezonde en betrouwbare markt van systemen?

H4 | Secundair gebruik

Hoe delen we gegevens voor onderzoek, innovatie en beleid?



Dataset catalogues and quality labels

Dataset descriptions and dataset catalogues

- Data holders will have to provide basic descriptions of their datasets to HDABs => detailed elements to be set out in implementing act;
- HDABs publish catalogues;
- HealthData@EU federates national catalogues.

This implementing act is expected in March '27

Data quality and utility label

- More detailed description of datasets, including 'levels' of label => detailed elements to be set out in implementing act;
- Mandatory if data collection received public funding => NB: funding *for collection* specifically, not for activities that result in incidental collection.
- Otherwise voluntary

Datakwaliteit in de EHDS

- **Definities** onder (art. 2(2))
 - gegevenskwaliteit: de mate waarin de elementen van elektronische gezondheidsgegevens geschikt zijn voor het beoogde primaire gebruik en secundaire gebruik ervan;
 - gegevenskwaliteits- en gegevensbruikbaarheids**label**: een grafisch diagram, met inbegrip van een schaal, waarin de gegevenskwaliteit en de gebruiksvoorwaarden van een dataset worden beschreven
- **Kwaliteitslabel** ondersteunt keuze van geschikte gezondheidsdatasets (recital 85)
- Zonder de beschikbaarheid van gezondheidsdata via de EHDS te belemmeren (recital 85)
- Toevoegen van label niet vrijblijvend voor publiek gefinancierde datasets (art. 78(2))
- Specificaties nog te ontwikkelen door de Europese Commissie (art. 78 (5+6))
- **Taken datahouder:**
 - Primair verantwoordelijk voor toevoegen label aan dataset (art. 78(1))
 - Verstrekken van voldoende documentatie om juistheid te verifiëren (art. 60(4))
- **Taken HDAB:**
 - Supervisie van label, inclusief de mogelijkheid om label in te trekken (art. 57(1d)) & (art. 78(4))
 - Maar ook mogelijkheid tot ondersteuning en samenwerking (art. 57(1d))

Verwachte elementen voor datakwaliteit art, 78 (3)

- a) voor **de documentatie** van de gegevens: metagegevens, ondersteunende documentatie, het gegevenswoordenboek, het gebruikte format en de gebruikte normen, de bron van de gegevens en, indien van toepassing, het gegevensmodel;
- b) voor de beoordeling van de **technische kwaliteit**: de volledigheid, het unieke karakter, de nauwkeurigheid, geldigheid, tijdigheid en consistentie van de gegevens;
- c) voor **processen voor het beheer van de gegevenskwaliteit**: de rijpheid van de processen voor het beheer van de gegevenskwaliteit, met inbegrip van evaluatie- en auditprocessen, en onderzoek naar vertekening (bias)
- d) voor de **beoordeling van de dekking**: de periode, bestreken populatie en, indien van toepassing, representativiteit van de bemonsterde populatie, en de gemiddelde tijdspanne waarin een natuurlijke persoon in een dataset voorkomt;
- e) voor de **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;
- f) voor de **informatie over wijzigingen van de gegevens**: samenvoeging van datasets en toevoeging van gegevens aan een bestaande dataset, met inbegrip van koppelingen met andere datasets.

QUANTUM project

- Het **QUANTUM project** is een Europees project met als doel artikel 78 van de EHDS verder uit te werken. Gestart in 2023, eindigt deze maand.
- Belangrijkste **resultaten**:
 - **raamwerk** waarin keuze gemaakt is voor 12 dimensies met verschillende kwaliteitsindicatoren en weging.
 - **applicatie** waarmee je de datakwaliteit score en visualisatie per dataset gemaakt kan worden (rdf file) en overeenkomend ook voor de score van datakwaliteitvolwassenheid per organisatie
 - QUANTUM raamwerk focus op fit-for-use (potential use) en eerste poging fit-for-purpose (actual use)
- **Werkwijze**: Op basis van eisen vanuit EHDS, literatuuronderzoek, vragenlijsten (Delphi methode), interviews met experts, statistische analyses.
⇒ Technische specificatie (D1.1) Data Quality & Utility label (welke dimensies meest gevraagd/gewild, belang dimensies tov elkaar – weegfactoren, betere scores – sterren visualisatie).
- De resultaten zijn een advies aan de commissie voor uitvoeringshandelingen gegevenskwaliteit (maart '27), MAAR het is niet zeker in welke mate de commissie deze resultaten de uitvoeringshandelingen zullen bepalen. (er is geen TEHDAS consultatie over datakwaliteit)

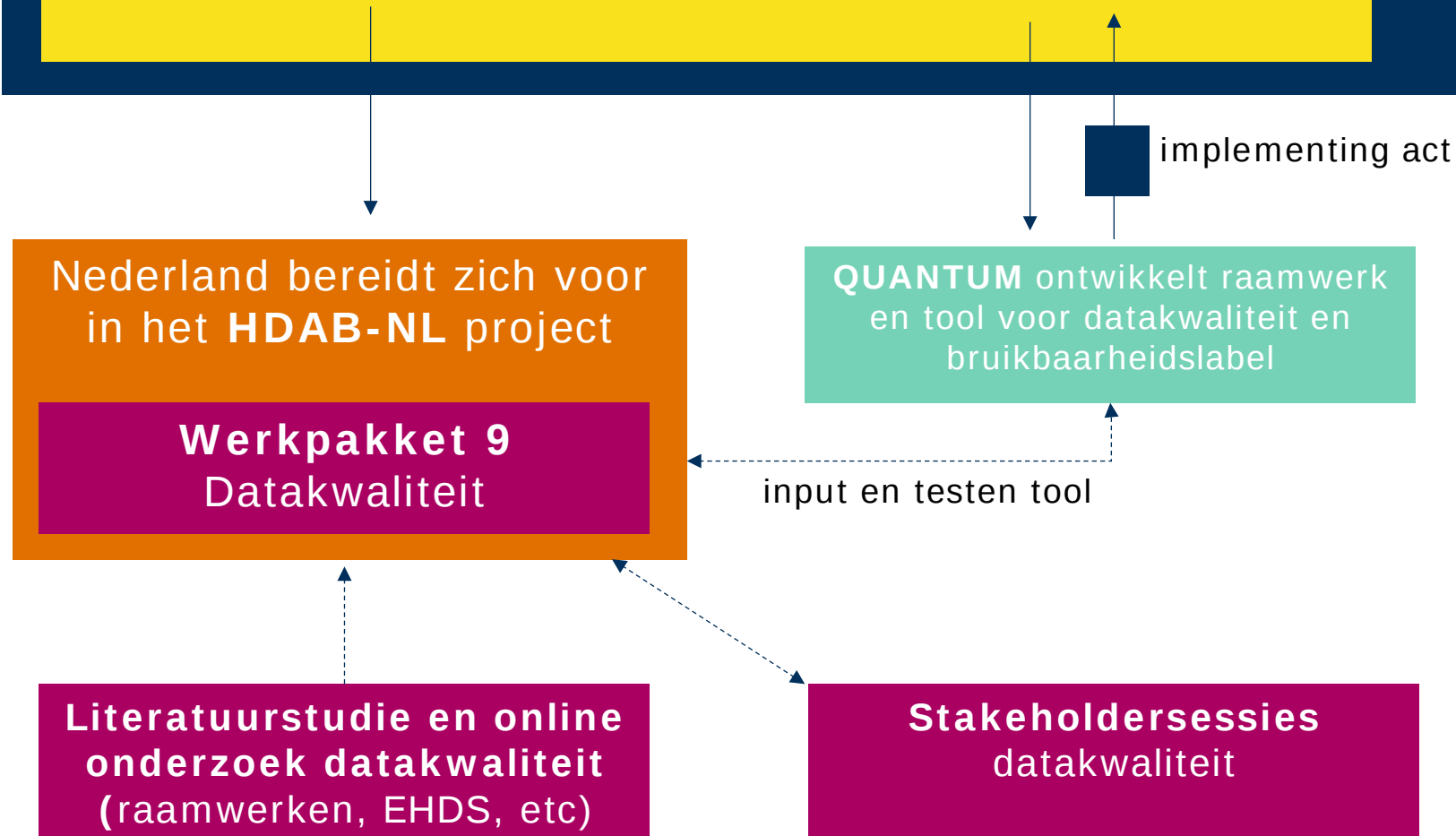
The 12 QUANTUM Data Quality & Utility Dimensions

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



European Health Data Space - EHDS

Europese Data Spaces



Taak WP9: Advisering

- Toekomstige HDAB
- Datahouders
- Datagebruikers

Hoe of waarmee bijv.:

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

Raamwerken voor gegevenskwaliteit en definities van datakwaliteitsdimensies

- QUANTUM is niet het enige datakwaliteitsraamwerk.
- Een vergelijking:
 - EHDS
 - TEHDAS1
 - QUANTUM
 - NORA
 - EMA

Datakwaliteit elementen in de EHDS

- **EHDS artikel 78(3)**

- Geen definities

3. The data quality and utility label shall cover the following elements, where applicable:
 - (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;
 - (b) for assessment of technical quality: completeness, uniqueness, accuracy, validity, timeliness and consistency of the data;
 - (c) for data quality management processes: the level of maturity of the data quality management processes, including review and audit processes, and bias examination;
 - (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;
 - (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;
 - (f) for information on data modifications: merging and adding data to an existing dataset, including links with other datasets.

Eerste opzet EHDS datakwaliteit raamwerk

- **TEHDAS1**
- Deliverable 6.3 Recommendations on a Data Quality Framework for the European Health Data Space for secondary use

Dimension	Definition
Relevance	How well data meets users' needs.
Accuracy and Reliability	How closely data reflects what it was designed to measure and whether this is consistent over time.
Coherence	How consistent is data across data sources and data holders and can be combined and compared.
Coverage	The degree of representativeness in the population to which the dataset refers to, its exposures and events.
Completeness	Level of missingness at variable level.
Timeliness	How up-to-date the information is collected and delivered.

Datakwaliteit in QUANTUM

- QUANTUM
- Deliverable 1.1 Specification of the data sets' quality and utility label

EHDS category	Dimension (by priority)	Dimension definition
Technical quality	Accuracy	Accuracy refers to the degree to which observations correctly describe what it was designed to measure
Technical quality	Precision	Precision refers to the degree of approximation by which data can represent reality
Technical quality	Coherence	Coherence is defined as the dimension that expresses how different parts of the dataset are uniform in their representation and meaning over time, such as formats, semantics (stability of the data models), and methods
Technical quality	Consistency	Consistency refers to the degree to which data has attributes that are plausible and are uniform with other data and over time
Data documentation	Metadata scope	Metadata scope refers to the availability, comprehensiveness, level of detail of metadata and data dictionary that help users understand the data being used
Technical quality	Validity	Validity refers to the degree to which representations of data in a dataset conform to the specification of a data model or data models
Access and provision	Accessibility	Accessibility refers to the dataset being accompanied by clear and transparent access and usage conditions
Data documentation	Compliance	Compliance refers to the degree to which data has attributes that adhere to ethical standards, conventions, protocols or regulations
Coverage	Population representativity	Population representativity refers to the degree to which the data adequately represent the population in question
Coverage	Population coverage	Population coverage refers to the degree to which a dataset includes the potential eligible population
Technical quality	Completeness	Completeness refers to the degree to which all information that could be available is present in a particular dataset
Data documentation	Data provenance	Data provenance means a description of the source of the data, including context, purpose, method and technology of data generation, documenting agents involved in the provenance of data, data validation routines, source data verification, traceability of changes, and quality control of data

Andere datakwaliteit raamwerken

- NORA – Nederlandse Overheid Referentie Architectuur
- Raamwerk gegevenskwaliteit

Juistheid	Thematische juistheid	Classificatie juistheid	Positionele juistheid	Kwantitatieve juistheid		
Compleetheid	Dataset compleetheid	Object compleetheid	Over compleetheid	Historie compleetheid	Metadata compleetheid	Ruimtelijke dekking
Validiteit	Formaat validiteit	Domein validiteit	Geometrische validiteit			
Consistentie	Logische consistentie	Referentiële integriteit	Identificeerbaarheid	Homogeniteit	Topologische consistentie	
Actualiteit	Update frequentie	Versheid				
Precisie	Opslagprecisie	Geometrische precisie	Statistische precisie			
Plausibiliteit	Authenticiteit	Reputatie	Bewijsbaarheid	Waarschijnlijkheid	Represen- tativiteit	
Traceerbaarheid	Herleidbaarheid	Reproduceerbaarheid	Gebruiks inzicht			
Begrijpelijkheid	Leesbaarheid					

Andere datakwaliteit raamwerken

- EMA – European Medicines Agency
HMA - Heads of Medicines Agencies
TEHDAS
- Data Quality Framework for
EU medicines regulation



Dimensions					Attributes			
EHDS	QUANTUM	TEHDAS	EMA	NORA	EHDS	QUANTUM	EMA	NORA
Technical Quality	Technical Quality	Accuracy & Reliability	Reliability	Accuracy	Accuracy	Accuracy	Accuracy	Classification Accuracy
				Precision		Precision	Precision	Thematic accuracy
				Plausibility			Plausibility	Precision
								Authenticity
								Provability
								Reputation
								Probability
							Traceability	
		Coherence	Coherence	Validity	Validity	Validity	Validity	Format validity
								Domain validity
						Coherence	Conformance	
				Consistency	Consistency	Consistency	Format Coherence	Homogeneity
								Logical consistency
							Structural/ Relational Coherence	Referential integrity
							Semantic Coherence	
					Uniqueness		Uniqueness	
		Completeness	Extensiveness	Completeness	Completeness	Completeness	Completeness	Attribute completeness
								Dataset completeness
								History completeness
								Identifiability
								Metadata completeness
								Overcompleteness
		Timeliness	Timeliness	Currentness	Timeliness		Currency	Timeliness
					Time between collection and addition to dataset			Update frequency
Coverage	Coverage	Coverage	Extensiveness		Population coverage	Population Coverage	Coverage	
					Population representativity	Population Representativity	Representativeness	Representativeness
					Period			
					Average timeframe present in dataset			
							Missingness	
Access and provision	Access and provision				Time between collection and addition to dataset			
					Time needed to provide data	Accessibility		
Data documentation	Data documentation	Relevance	Relevance	Understandability	Metadata	Metadata scope		Clarity
					Data dictionary	Metadata scope		
					Formats and standards used			
					Data model			
					Support documentation	Accessibility		
						Compliance		
Data modifications				Traceability	Source of the data	Data provenance	Traceability	Lineage
								Usage
								Reproducibility
					Merging and adding with other datasets	Data provenance		
					Links with other datasets	Data provenance		
Data quality management					Level of maturity of DQ process	Separate maturity label	Separate chapter in the dqf	Separate document

Accuracy = accuracy?

- EMA accuracy:
 - *'the amount of discrepancy between data and reality'*
 - Hoe veel wijk je af?
- QUANTUM accuracy:
 - *'the degree to which observations correctly describe what it was designed to measure'*
 - Hoe vaak meet je goed?
- Een accuracy van 3% is

Dimensions					Attributes			
EHDS	QUANTUM	TEHDAS	EMA	NORA	EHDS	QUANTUM	EMA	NORA
Technical Quality	Technical Quality	Accuracy & Reliability	Reliability	Accuracy	Accuracy	Accuracy	Accuracy	Classification Accuracy
								Thematic accuracy
				Precision		Precision	Precision	Precision
				Plausibility			Plausibility	Authenticity
								Provability
								Reputation
								Probability
							Traceability	
		Coherence	Coherence	Validity	Validity	Validity	Validity	Format validity
								Domain validity
						Coherence	Conformance	
				Consistency	Consistency	Consistency	Format Coherence	Homogeneity
								Logical consistency
							Structural/ Relational Coherence	Referential integrity
							Semantic Coherence	
					Uniqueness		Uniqueness	
		Completeness	Extensiveness	Completeness	Completeness	Completeness	Completeness	Attribute completeness
								Dataset completeness
								History completeness
								Identifiability
								Metadata completeness
								Overcompleteness
		Timeliness	Timeliness	Currentness	Timeliness		Currency	Timeliness
					Time between collection and addition to dataset			Update frequency
Coverage	Coverage	Coverage	Extensiveness		Population coverage	Population Coverage	Coverage	
					Population representativity	Population Representativity	Representativeness	Representativeness
					Period			
					Average timeframe present in dataset			
							Missingness	
Access and provision	Access and provision				Time between collection and addition to dataset			
					Time needed to provide data	Accessibility		
Data documentation	Data documentation	Relevance	Relevance	Understandability	Metadata	Metadata scope		Clarity
					Data dictionary	Metadata scope		
					Formats and standards used			
					Data model			
					Support documentation	Accessibility		
						Compliance		
Data modifications				Traceability	Source of the data	Data provenance	Traceability	Lineage
								Usage
								Reproducibility
					Merging and adding with other datasets	Data provenance		
					Links with other datasets	Data provenance		
Data quality management					Level of maturity of DQ process	Separate maturity label	Separate chapter in the dqf	Separate document

Definities en definities

- Frameworks, Dimensions, Definitions of Aspects, and Assessment Methods for the Appraisal of Quality of Health Data for Secondary Use: Comprehensive Overview of Reviews (JMIR 2024)
- 22 artikelen, 23 dimensies met 62 definities

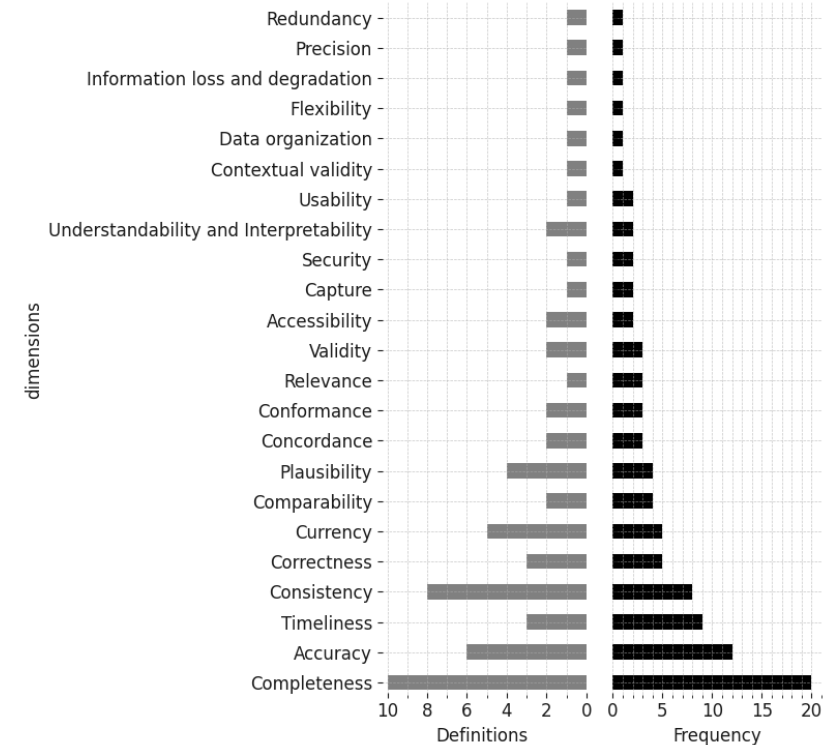


Figure S1. Multimedia Appendix 3. Declerck J, Kalra D, Vander Stichele R, Coorevits P. Frameworks, Dimensions, Definitions of Aspects, and Assessment Methods for the Appraisal of Quality of Health Data for Secondary Use: Comprehensive Overview of Reviews. JMIR Med Inform. 2024 Mar 6;12:e51560. doi: 10.2196/51560. PMID: 38446534; PMCID: PMC10955383.

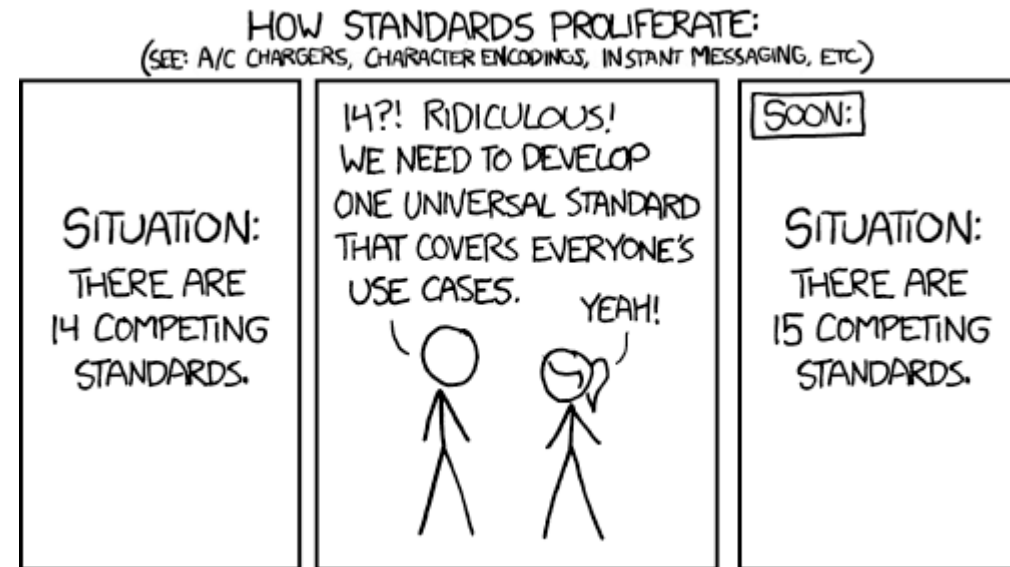
Hoogover gezien..

- Accuracy and precision: Meet de data wat het moet meten op het juiste detailniveau?
 - Validity: Voldoet de data aan standaarden en het datamodel?
 - Consistency: Is de data coherent en consistent?
 - Completeness: Is er genoeg data?
 - Currentness: Is de data actueel genoeg?
 - Coverage: Zit iedereen die geïnccludeerd zou moeten zijn in de dataset?
 - Traceability: Is er documentatie van hoe deze data tot stand is gekomen?
-
- Data documentation: Is er genoeg documentatie over deze dataset?
 - Access and Provision: Kan ik toegang krijgen tot de dataset en staan de voorwaarden beschreven?

is dit ook niet sluitend

Lessons learned

- The devil is in the definitions: leg altijd vast welke definitie je hanteert.
- Het EHDS datakwaliteit- en bruikbaarheidlabel zal niet alle problemen oplossen, maar is een stapje naar gezamenlijke taal.
- Zorg voor mensen met kennis van datakwaliteit.



QUANTUM labelling tool

- Levert na invulling een datakwaliteitslabel op, zowel als RDF als een PDF
- Beschikbaar als webapplicatie en via Dockerinstallatie
- Twee onderdelen
 - o Datakwaliteit: beoordeling van een dataset
 - o Volwassenheid (maturity): beoordeling van procesbeheersing datakwaliteit
- Er is géén verbinding met de data nodig

Welcome to the QUANTUM Labelling Tool

The **QUANTUM Data Quality & Utility Labelling Tool** is a key component of the **European Health Data Space (EHDS)** initiative, designed to address the challenge of ensuring that health datasets are of high quality, accessible, and interoperable across EU member states. As healthcare data plays an increasingly vital role in research, policy, and innovation, this tool provides a standardized mechanism to evaluate and label the **quality, utility, and maturity** of datasets, supporting stakeholders such as healthcare institutions, research organizations, and policymakers.

This tool empowers **data holders** by offering functionalities to:

- Manage **catalogues** of datasets for their labelling.
- Assess datasets quality and utility based on a guided evaluation of the QUANTUM key dimensions (such as **accuracy**, **accessibility**, and **compliance**) using your own practices.
- Generate and visualise **data quality and utility certificates (labels)** and download them in the healthdata@EU standardized **RDF format**.
- Facilitate compliance with **EU-wide quality standards** such as [HealthDCAT-AP](#), [DCAT](#), [DOV](#).
- Provide a transparent and accessible method for **assessing data holders maturity**.

Ultimately, the QUANTUM tool aims to foster **trust in health data** across Europe by promoting reliable, reusable datasets that can be confidently shared within the European Health Data Space.

[Go to DQ&U Dashboard](#)[Go to Maturity Dashboard](#)

Need help? Visit the [Help Center](#) or create a ticket by [clicking here](#)

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 test	Edit	★★★★☆		
CBS-Gezondheid en zorg	Overledenen; doodsoorzaak (uitgebreide lijst), leeftijd, geslacht test	Edit	★★★★☆		
CBS-Gezondheid en zorg	Microdata GGZ DBC trajecten test	Edit	★★★★☆		
Create dataset					

Dataset information

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

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



- 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

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: (Number of individuals in the dataset / Total eligible population) x 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

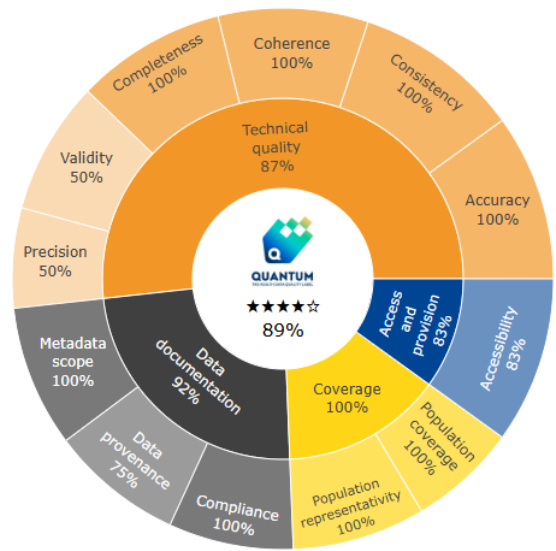
2.2

3. Data documentation

Catalogue	Name
CBS-Gezondheid en zorg	Causes of death Netherlands test

This label shows a completed quality assessment.

100%



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)

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

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

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

7. Data enhancement - augmentation

8. Data enhancement - enrichment

9. Data Model

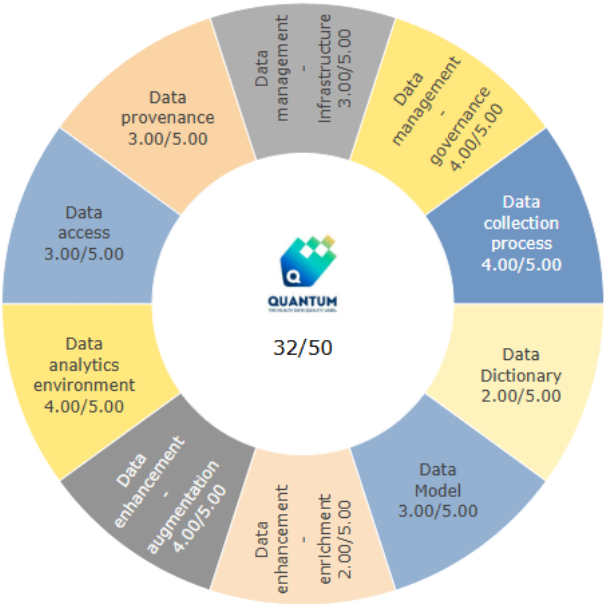
10. Data Dictionary

My organization - CBS test

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.

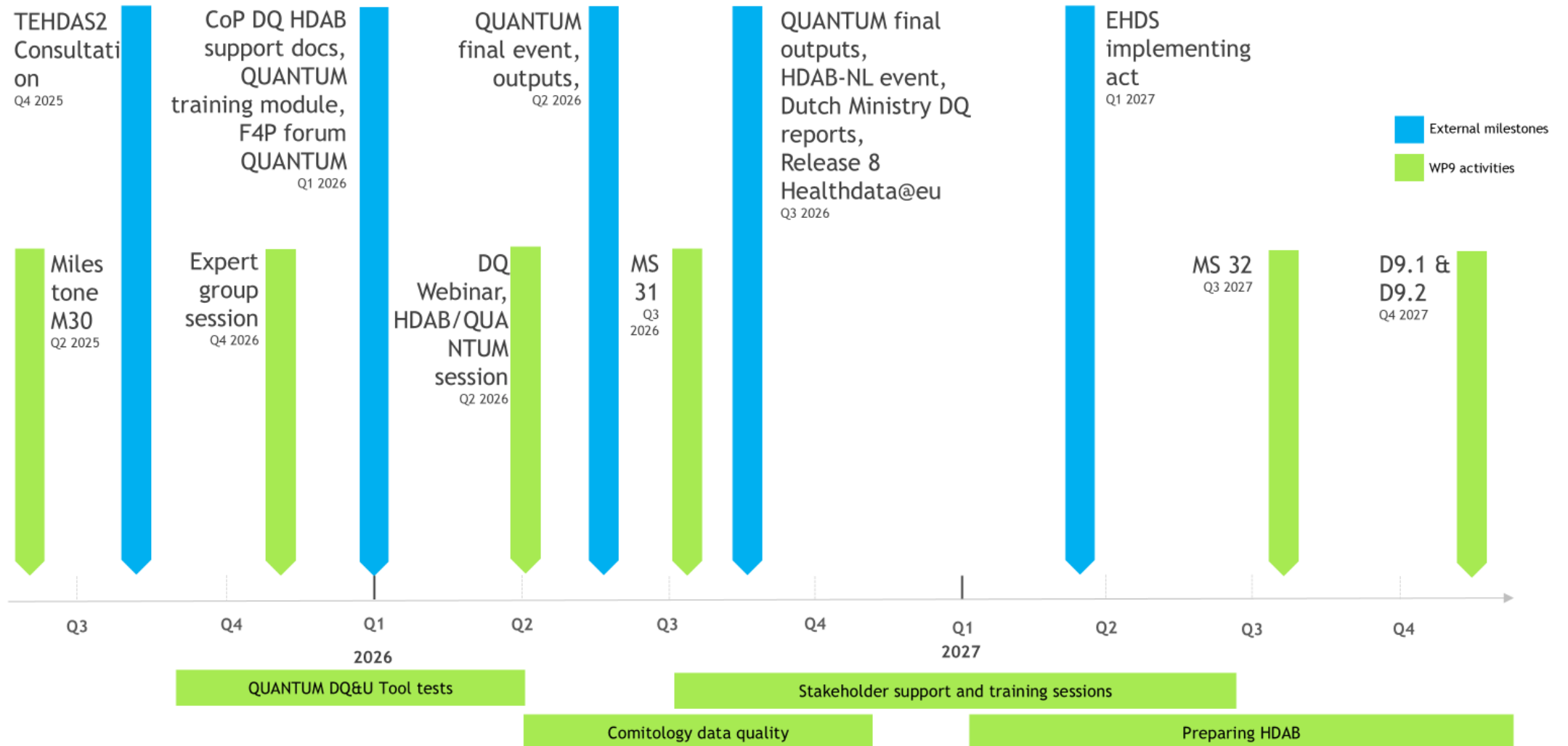
Lessons learned

- Meerdere specialismen nodig om in te kunnen vullen, zowel bij datakwaliteitslabel als bij volwassenheidslabel.
- Aanvullende documentatie gewenst.
- Kan veel handmatig werk zijn.
- Vooral fit-for-use, beperkt fit-for-purpose.
- Aantal sterren zegt niet alles over geschiktheid voor specifiek doel
- Geeft handvaten voor verbetering datakwaliteit

- Zie: <https://zenodo.org/communities/quantumproject>

- NB: Vanuit Europa is nog niet bepaald of dit de standaard wordt.

4. Vooruitblik





Ministerie van Volksgezondheid,
Welzijn en Sport

Vragen of opmerkingen?

Denk mee. Praat mee.

