



Health Data Access Body-NL





Ministerie van Volksgezondheid,
Welzijn en Sport

Expertgroep HDAB-NL Datakwaliteit

30 september 2025

Samen gezond, fit en veerkrachtig

Programma sessie datakwaliteit



13.30 Start plenair programma:

- Welkom, introducties
- Datakwaliteit EHDS, HDAB-NL, WP9 en QUANTUM - Ellen Leenarts (20 min)
- Metadastandaarden DCAT en HealthDCAT – Hannah Neikes (15 min)
- Notitie *databeschikbaarheid is nog geen datageschiktheid. Over het belang van domeinkennis bij de implementatie van een Nederlandse Health Data Access Body* – Prof. dr. Robert Verheij (10 min)

14.30 Denk mee: Wat is uw tip? - Menti

15.00 Praat mee

1. Het verhogen van datakwaliteit in de huidige praktijk
2. Gebruik van raamwerken voor verbetering datakwaliteit
3. Metadata voor verbetering datakwaliteit

15.20 Wrap up

15.30 Borrel

Voorstelrondje

- Stel jezelf voor in 1 minuut:
- Naam,
- Functie,
- Organisatie

Of ga naar www.menti.com code 5632 2880





Ministerie van Volksgezondheid,
Welzijn en Sport

WP9 datakwaliteit

Publieke samenvatting van 'milestone' 30
september 2025

Samen gezond, fit en veerkrachtig

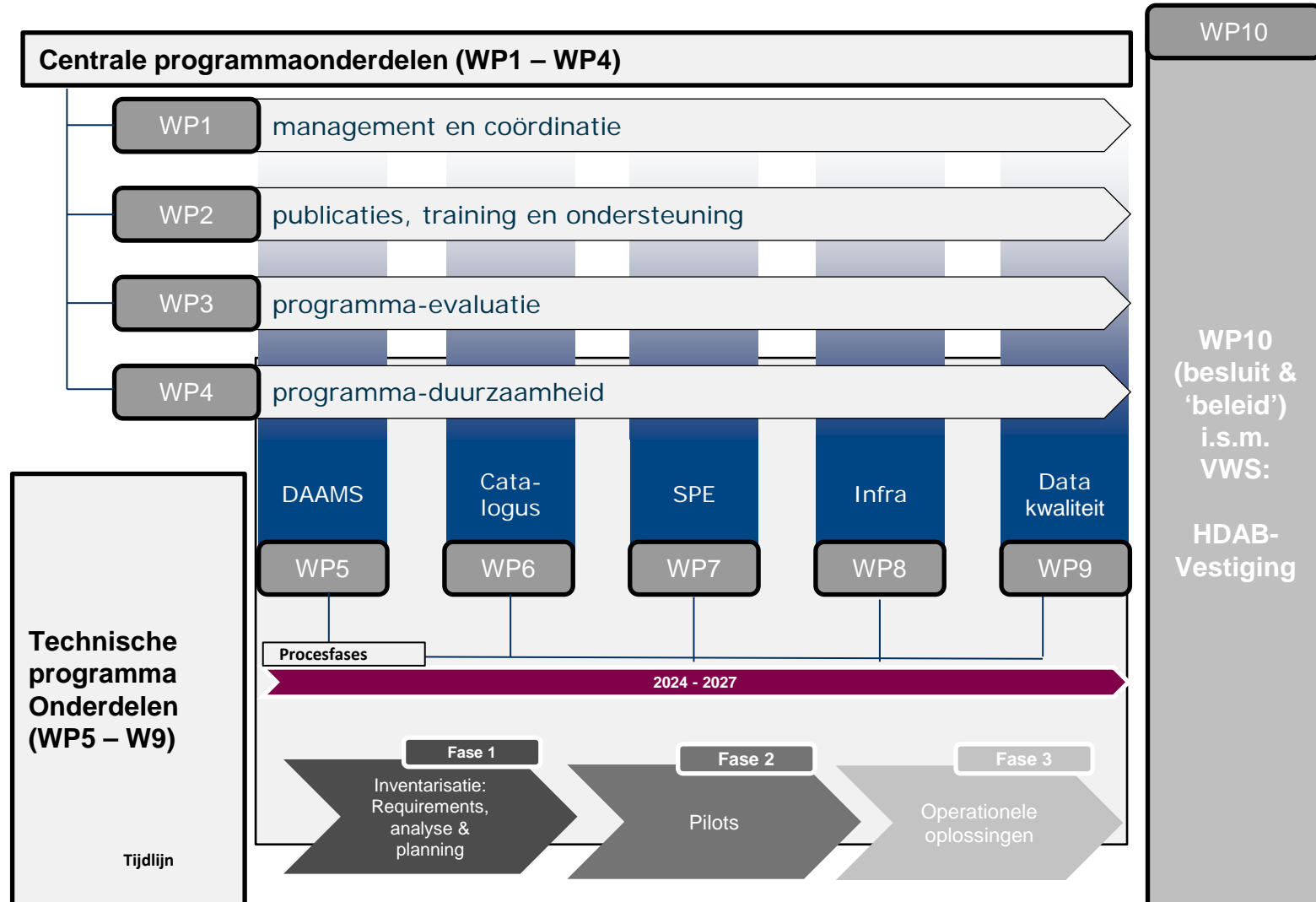
Programma HDAB-NL (2024 – 2027)

Het HDAB-NL programma bestaat uit:

- 4 centrale (c.q. horizontale) programmaonderdelen (WP 1 t/m 4)
- 5 technische onderdelen (WP 5 t/m 9)
- 1 afrondend c.q. beleid & besluitvormend onderdeel o.l.v. VWS (WP 10)

De technische werkpakketten volgen een 3-fasen aanpak:

1. Inventariseren (ophalen eisen en wensen, divergeren)
[jan. 2024 – maart 2025]
2. Vertalen naar concrete beproevingen (convergeren, pilots)
[april 2025 – aug. 2026]
3. Realiseren van operationele oplossingen
[aug. 2026 – nov. 2027]



MS30 werkpakket 9: Verbeteren van datakwaliteit en bruikbaarheid label en volwassenheid

- Eerste milestone
- Datum opgeleverd: 30 juli 2025
- Publicatie: intern, niet publiek

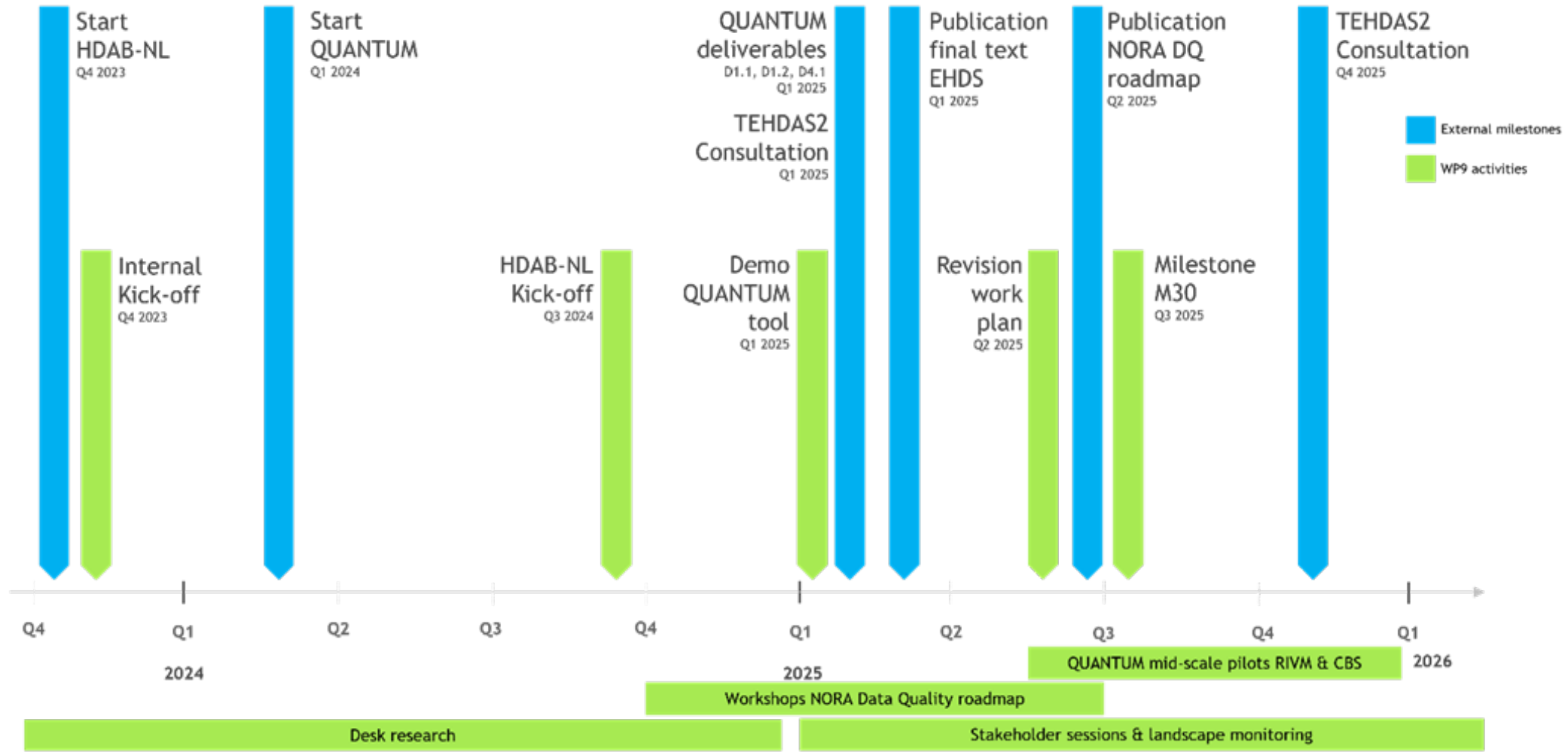
- Verslag over activiteiten vanaf dec 2023 – juni 2025

- **Oorspronkelijk doel werkpakket 9:** ontwikkeling van datakwaliteitsraamwerk, datakwaliteitslabel en volwassenheid, ondersteuning van implementatie met richtlijnen, workshops en webinars.

- Verschuiving in focus door ontwikkeling van Europees breed datakwaliteitsraamwerk, label en volwassenheid binnen het Europese project QUANTUM

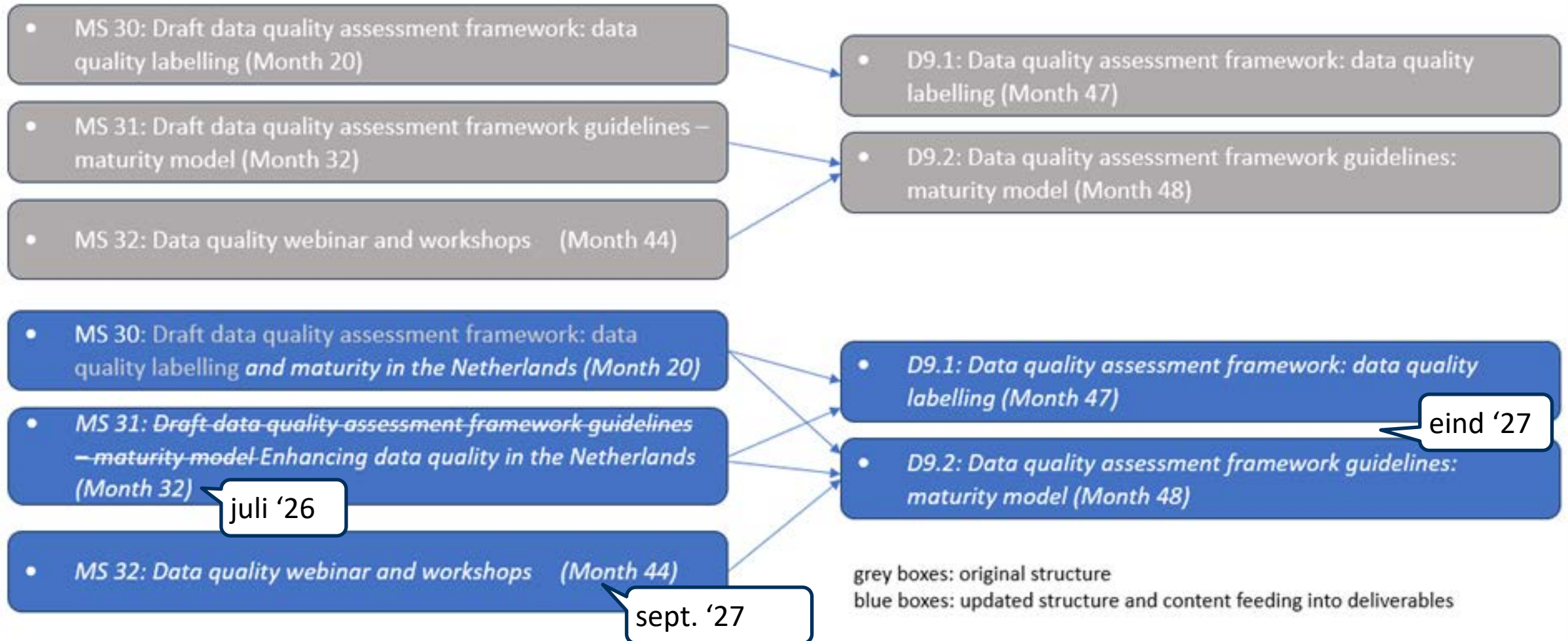
- **Huidige doel werkpakket 9:** Bruikbaarheid van de datakwaliteitsraamwerken (o.a. QUANTUM) met als context de EHDS en de Nederlandse datahouders, gebruikers en toekomstige HDAB.

Timeline context WP 9



Datakwaliteit HDAB-NL – context (externe milestones in blauw). De activiteiten van het werkpakket in groen.

Milestones en deliverables WP9



Original and planned milestones and deliverables of work package 9

Wat kunnen we verwachten?

- **25 maart 2025:** [EHDS](#) van kracht.
 - Gegevenskwaliteit 27 keer genoemd. Relevante artikelen 51 (minimum categoriën), artikel 78
 - Er komt een infrastructuur voor secundair gebruik: HealthData@EU. Dit werkt straks naast het bestaande MyHealth@EU (primair gebruik).
 - De HDABs zijn nationale contactpunten voor secundair gebruik.
 - De HDAB zal vooral infrastructuur maar ook support leveren: metadatacatalogus, beveiligde verwerkingsomgeving, aanvraagstelsel en connectie met Europese HealthData@EU infrastructuur.
 - **26 maart 2027: implementing Act** EHDS voor beschrijving van datasets
- **Eindigt augustus 2026: QUANTUM** project – EU project met als opdracht: ontwikkeling datakwaliteit en bruikbaarheidslabel EHDS, ontwikkeling tool voor datahouders, in kaart brengen implementatie uitdagingen, datakwaliteit volwassenheid bij datahouders, en ontwikkeling van support materialen. Er zijn 2 deliverables gepubliceerd, er komen nog 10 deliverables, er is een eerste testpilot geweest voor de tool.
- **TEHDAS2** EU project – **eindigt december 2026** – bereid richtlijnen voor voor een verbeterde uitwisseling van gezondheidsgegevens in het kader van de EHDS (niet specifiek datakwaliteit) en ondersteunt vooral HDABs, maar ook datahouders en datagebruikers en de EU in de vorm van deliverables waar consultaties aan voorafgaan.
- **Wat betekent de EHDS:**
 - Voor **datahouders**:
 - **begin 2029:** Bied datasetbeschrijvingen aan voor de meeste categorieën en wees bereid om die data beschikbaar te stellen.
 - **begin 2031:** idem voor resterende categorieën
 - Voor **datagebruikers**:
 - **begin 2029:** start indienen van aanvragen voor de meeste categorieën
 - **begin 2031:** idem voor resterende categorieën

Aanpak WP9 afgelopen anderhalf jaar

- **Bestudering verschillende bronnen:**
 - bestaande datakwaliteits raamwerken (zie later in deze presentatie voor meer detail)
 - draft en uiteindelijke EHDS
 - rapporten o.a. TEHDAS1, QUANTUM
- **Participatie in verschillende initiatieven:**
 - werkgroep stappenplan gegevenskwaliteit NORA
 - Delphi onderzoek QUANTUM dimensies
 - Pilot test QUANTUM tool
 - Community of Practice vanuit de EU voor metadatacatalogus en datakwaliteit
 - WP6 metadatacatalogus in HDAB
- **Informatieuitwisseling met verschillende stakeholders** (datagebruikers, datahouders en HDABs) in de vorm van werkbezoeken, participatie en presentaties in events en organisatie van sessies o.a. met WP6 (metadatacatalogus) en QUANTUM.

Datakwaliteit in de EHDS (i)

- 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 gegevensbruikbaarheidslabel: 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))

Datakwaliteit in de EHDS (ii)

- Taken data houder
 - 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))
- Dimensies zijn als volgt:
 - Documentatie: metadata, formats, normen, bron, gegevensmodel
 - Technische kwaliteit: volledigheid, nauwkeurigheid, validiteit, consistentie
 - Kwaliteitsbeheer: mate van procesrijpheid, audits, bias-analyse
 - Dekking: periode, populatie, representativiteit, verblijfsduur in dataset
 - Toegang & levering: tijd tot opname in dataset, tijd tot verstrekking na goedkeuring
 - Gegevenswijzigingen: samenvoegingen, toevoegingen, koppelingen met andere datasets

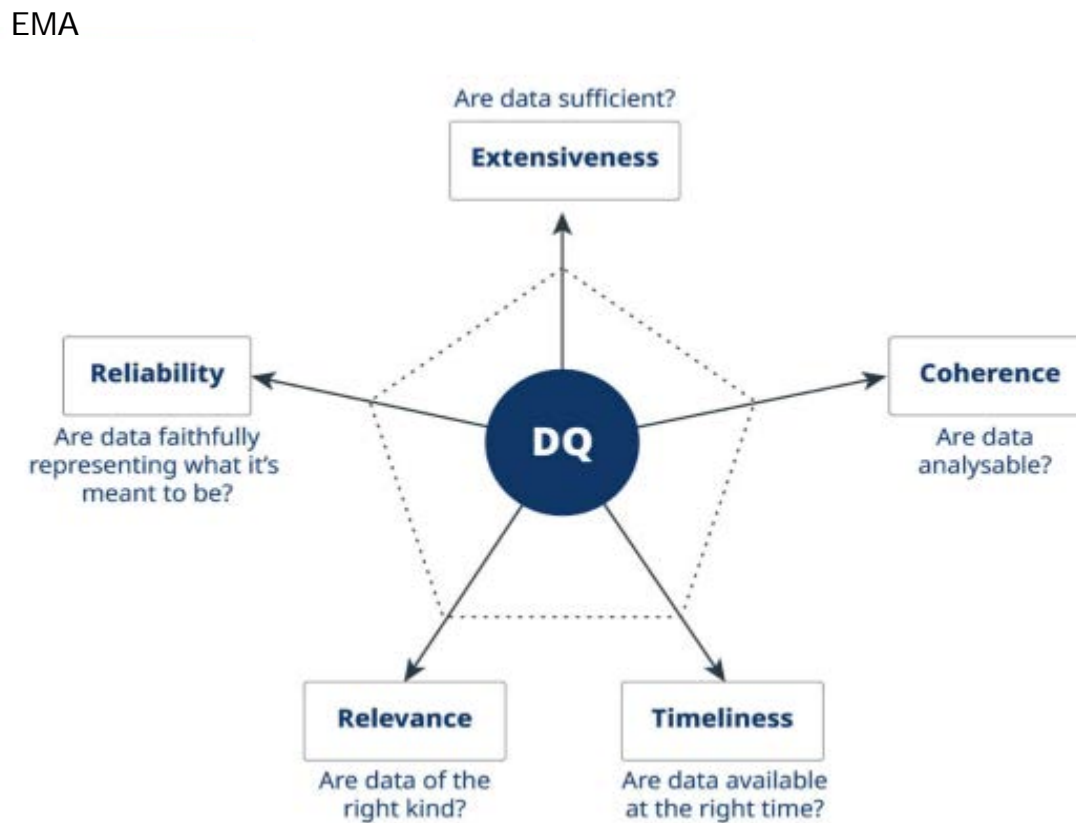
Analyse verschillende raamwerken en principes van gegevenskwaliteit en volwassenheid

Het werkpakket begon met een analyse van de definities en reikwijdte van datakwaliteitsdimensies in verschillende bestaande raamwerken, waarbij vooral werd gekeken naar de overeenkomsten en verschillen tussen deze dimensies en hun bijbehorende indicatoren.

- **TEHDAS1:** [TEHDAS' proposals for data quality and utility in EHDS - deliverable](#)
- **QUANTUM:** [twee deliverables raamwerk \(label & volwassenheid\)](#)
- **NORA online:** [Raamwerk gegevenskwaliteit](#) en [Stappenplan Gegevenskwaliteit](#)
- **DAMA international en DAMA NL:** internationale datamanagement organisatie en [Nederlandse werkgroep datakwaliteit](#)
- **EMA (European Medicines Agency):** [data quality framework EMA](#)
- **Health Data Research UK** [Data Utility Evaluation - HDR UK](#)
- **Data quality at a Ministry**
- **EOSC taskforce FAIR metrics and data quality** [Towards a Data Quality Framework for EOSC](#)
- **ALCOA+ + principes** [guideline](#)
- **Metadata kwaliteit (EU)** [data.europa.eu methodology](#)
- **NICTIZ Nationale bibliotheek en terminologie server**
- **CoreTrustSeal** certificeringseisen data repositories [CoreTrustSeal – Core Trustworthy Data Repositories](#)
- **HealthDCAT** specifieke extensie voor gezondheidsdata voor de metadata-standaard DCAT-AP [HealthDCAT AP | European Health Information Portal](#)

Bestaande raamwerken datakwaliteit - dimensies

TEHDAS1	NORA	DAMA	QUANTUM		EMA
Dimension	Juistheid	Dimension of Quality	Category	Key dimension	
Relevance	Compleetheid	Validity	Access & provision	Accessibility	
Accuracy and Reliability	Validiteit	Completeness			
Coherence	Consistentie	Consistency	Coverage	Population coverage	
Coverage	Actualiteit	Integrity	Data documentation	Population representativity	
Completeness	Precisie	Timeliness		Compliance	
Timeliness	Traceerbaarheid	Currency	Technical quality	Data Provenance	
	Begrijpelijkheid	Reasonableness		Metadata scope	
		Uniqueness/ Deduplication		Accuracy	
		Accuracy		Coherence	
				Completeness	
				Consistency	
				Precision	
				Validity	



Analyse van de datakwaliteitsraamwerken

- De raamwerken hebben elk een verschillende dimensies consistentie en coherentie. Daarnaast worden in verschillende raamwerken dezelfde dimensies met verschillende definities genoemd.
- De exacte invulling en meetmethoden voor dimensies verschillen per raamwerk, waardoor het in kaart brengen van deze dimensies complex is en diepgaande kennis vereist.
- De vergelijking van deze raamwerken op het niveau van de dimensies is nog niet af en wordt vervolgd.

QUANTUM labelling Tool

100%

1. Access and provision

1.1. Accessibility

Definition: Accessibility refers to the dataset being accompanied by clear and transparent access and usage conditions.
Dimension Weight: 9.95 out of 100.

Metric #1

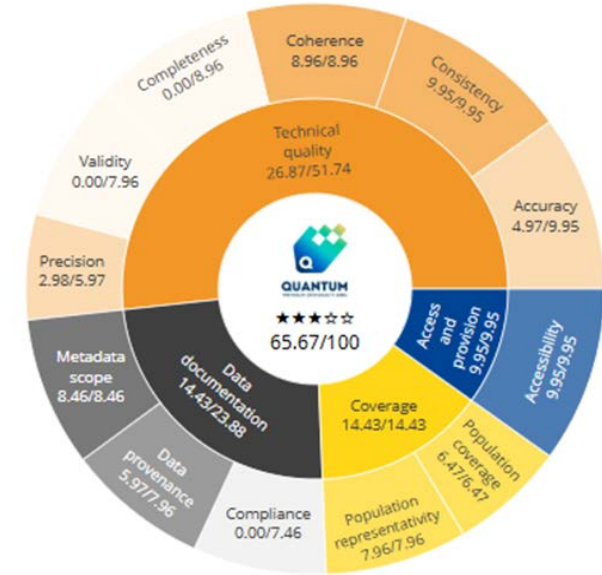
Definition
Availability of a data access & usage policy at the time of release of the dataset

Recommended Measurement Approach
https://www.w3.org/TR/vocab-dcat/#Propertydistribution_access_rights

Weight
50%

Comprehensive policy available

Metric #2



Quality & Utility Assessment score



Total Q&U Score: 65.67 / 100

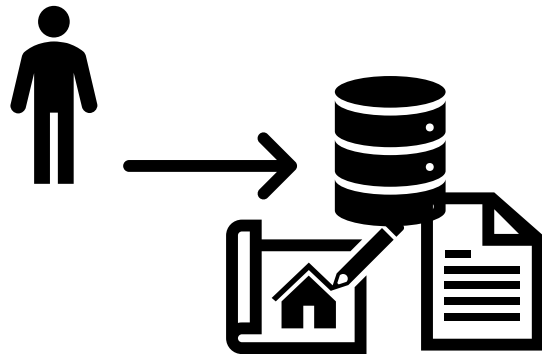
Maturity score: 42.0% (21/50)

Overige bevindingen

- Verwacht wordt dat gezondheidsdatacatalogi zullen werken met een metadataschema als DCAT uitgebreid met HealthDCAT-AP, waarbij de data-kwaliteitsdimensies van QUANTUM nog niet zijn opgenomen; het wordt sterk aanbevolen om deze wel toe te voegen en te integreren in het schema en een gecontroleerde vocabulaire te ontwikkelen.
- Inzichten uit de EOSC Task Force onderstrepen dat het moeilijk is om absolute datakwaliteit te definiëren, omdat de benodigde kwaliteitsdimensies en -niveaus afhangen van het specifieke gebruik, de stakeholders en de levenscyclus van de dataset, waardoor één dataset nooit aan alle mogelijke eisen kan voldoen.
- Niet alle onderzochte raamwerken worden actief onderhouden; het is belangrijk om een duidelijke eigenaar aan te wijzen voor het QUANTUM-raamwerk na afloop van het project, zodat het framework kan worden bijgewerkt en relevant blijft, en hierover volgt later een nadere uitwerking door QUANTUM.
- Een terugkerende aanbeveling is om typische gebruiksinformatie, zoals gebruiksinstructies, beperkingen, toegepaste anonimiseringstechnieken en relaties tussen attributen, toe te voegen aan de metadata of beschikbare documentatie zodat gebruikers beter geïnformeerde keuzes kunnen maken. (zie volgende slide)

Fit-for-use (bruikbaar), fit-for-purpose (bruikbaar voor bepaald doel) en ervaren datakwaliteit

datahouder



HDAB



datagebruiker



De data is fit-for-use en fit-for-purpose en wordt beschreven in metadata (HealthDCAT)
Quantum tool voor datakwaliteitslabel – focust op vastlegging technische kwaliteit (dimensies) en minder op overige documentatie, volwassenheid

Voordeel van verbetering metadata en documentatie aan de kant van de datahouder: minder mismatches, minder vragen die vanuit de hdab / datagebruikers.

Gebruiker vind data in de metadatalogus van HDAB. De vraag is, zal de data die gevonden is, bruikbaar zijn: actual utility = fit-for-purpose en hoe wordt de datakwaliteit beoordeeld.

De ervaren datakwaliteit van de data is waarschijnlijk afhankelijk van: volledigheid van metadata, beschikbare aanvullende documentatie (bijv. data dictionary, gebruik standaarden), aanvullende informatie / communicatie met datahouder en de data zelf.

Wat heeft de gebruiker nodig om een goede match te vinden?

Volwassenheid: stappenplan NORA

	Basis – eerste serieuze stappen	Gevorderd - professioneel	Expert - geavanceerd
1. Bepalen context	<ul style="list-style-type: none"> • Start met concreet probleem, behoefte, nulmeting en/of wetgeving • Zicht op de bedrijfscontext • Uitvoerende verantwoordelijkheid is duidelijk • Betrokkenheid van eindverantwoordelijke 	<ul style="list-style-type: none"> • Jaarlijkse verbetercyclus • Focus op dataset zelf • Uitvoerende en eindverantwoordelijke rollen gedefinieerd en geformaliseerd • Competente medewerkers 	<ul style="list-style-type: none"> • Focus op relatie met andere datasets (m.n. consistentie) • Processen en verantwoordelijkheden geformaliseerd • Borging competenties in functiehuis en in de werving
2. Beschrijven gegevens	<ul style="list-style-type: none"> • Eerste vastlegging van metagegevens, minimaal op het niveau van datasets • Eerste aanzet tot kritieke gegevenselementen 	<ul style="list-style-type: none"> • Centrale metagegevens repository, tot op het niveau van gegevenselementen • Begrippen, informatie- en/of gegevensmodel gedefinieerd • Verbinding met organisatiebrede master- en referentiegegevens gedefinieerd 	<ul style="list-style-type: none"> • Data lineage volledig inzichtelijk • Koppeling van metagegevens en modellen • Metagegevens actief en continu verzamelen, inzichtelijk maken en gebruiken om bij te sturen • Harmonisatie met organisatiebrede master- en referentiegegevens
3. Beschrijven regels en eisen	<ul style="list-style-type: none"> • Eerste inventarisatie van regels en eisen • Informele afspraken over kwaliteit 	<ul style="list-style-type: none"> • Centrale vastlegging van regels en eisen • Gestandaardiseerd formaat • Data profiling met professionele tool • Gegevensleveringsovereenkomsten 	<ul style="list-style-type: none"> • Formeel sturen op gegevensleverings-overeenkomsten • Geautomatiseerde controle, bewaking en bijsturing van regels en eisen
4. Meten gegevenskwaliteit	<ul style="list-style-type: none"> • Eerste controles, met focus op validiteit en consistentie binnen dataset 	<ul style="list-style-type: none"> • Geautomatiseerde metingen en duplicaatdetectie • Organisatiebreed kwaliteitsdashboard 	<ul style="list-style-type: none"> • Pro-actief signaleren en oplossen problemen met gegevenskwaliteit (data observability)
5. Afhandelen signalen	<ul style="list-style-type: none"> • Informeel proces voor melden en herstellen • Meldingen via e-mail • Bepalen oorzaken van problemen 	<ul style="list-style-type: none"> • Formeel proces voor melden en herstellen • Standaard terugmeldmechanisme • Bewaken van signaalafhandeling 	<ul style="list-style-type: none"> • Geautomatiseerde workflows • Sturen op kwaliteitseisen • Zicht en sturing op effect van verbetering
6. Implementeren maatregelen	<ul style="list-style-type: none"> • Informeel proces 	<ul style="list-style-type: none"> • Formeel proces, projectmatig • Implementeren maatregelen als formele beheersmaatregelen 	<ul style="list-style-type: none"> • Kwaliteitscyclus op beheersmaatregelen • Gegevenskwaliteit by design

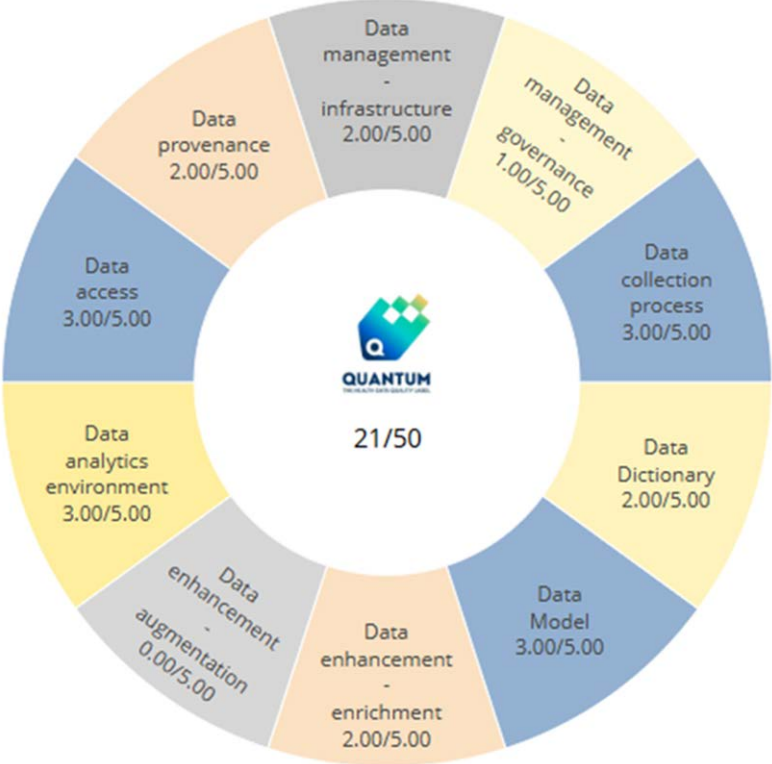
Quantum maturity model (i)

DIMENSION	DEFINITION	Level 1	Level 2	Level 3	Level 4	Level 5
		INITIAL: A NEW OR UNDOCUMENTED PROCESS, UNPREDICTABLE, POORLY CONTROLLED AND REACTIVE	EMERGING (OR DOCUMENTED, STRUCTURED): THE PROCESS IS DOCUMENTED SUFFICIENTLY SO REPEATING THE SAME STEPS MAY BE ATTEMPTED	DEFINED: THE PROCESS IS DEFINED AND CONFIRMED AS A STANDARD PROCESS	MANAGED: THE PROCESS IS QUANTITATIVELY AND PROACTIVELY MANAGED IN ACCORDANCE WITH AGREED-UPON METRICS	OPTIMISED: PROCESS MANAGEMENT INCLUDES DELIBERATE PROCESS OPTIMISATION
DATA COLLECTION PROCESS	The processes relating to the capture of data from multiple sources for secondary purposes	No data collection or capture process for secondary purposes	Ad hoc processes exist for data collection	Defined processes exist for some data collection	Defined processes exist for all data collection and are used for secondary purposes	Processes are automated for all data collection, focus on continuous process improvement
DATA MANAGEMENT – GOVERNANCE	The level of maturity of the data management process	No documented data management governance	A documented data management plan covering collection, auditing, and management is available for the dataset	Evidence that the data management plan has been implemented is available	Demonstrated compliance with the data management plan	Externally verified compliance with the data management plan
DATA MANAGEMENT - INFRASTRUCTURE	The level of implementation and development of the data holder's data management infrastructure	No data management infrastructure	An emerging data management infrastructure, some validation and verification	Data management infrastructure defined and confirmed as a standard process	Data management infrastructure, with partially automated, verified and validated [real time] data management	A robust and comprehensive data management infrastructure, with fully automated, verified and validated real time data management
DATA PROVENANCE	Clear description of the lineage of the dataset providing transparent and comprehensive documentation of the data pipeline to get the dataset	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
DATA ACCESS	How well defined and implemented are data access processes, from a legal, ethical and technical perspective	No data access processes or procedures	Have the processes and procedures but don't respond in timely and consistent manner	Have the processes and procedures, and respond in timely and consistent manner	Data access system that covers both technical and policy areas, in accordance with agreed metrics	A comprehensive data access system that covers technical, ethical and policy areas (allowable uses, API documentation, access and approvals) compliant with EU policy

Quantum maturity model (ii)

DIMENSION	DEFINITION	Level 1	Level 2	Level 3	Level 4	Level 5
		INITIAL: A NEW OR UNDOCUMENTED PROCESS, UNPREDICTABLE, POORLY CONTROLLED AND REACTIVE	EMERGING (OR DOCUMENTED, STRUCTURED): THE PROCESS IS DOCUMENTED SUFFICIENTLY SO REPEATING THE SAME STEPS MAY BE ATTEMPTED	DEFINED: THE PROCESS IS DEFINED AND CONFIRMED AS A STANDARD PROCESS	MANAGED: THE PROCESS IS QUANTITATIVELY AND PROACTIVELY MANAGED IN ACCORDANCE WITH AGREED-UPON METRICS	OPTIMISED: PROCESS MANAGEMENT INCLUDES DELIBERATE PROCESS OPTIMISATION
DATA ANALYTICS ENVIRONMENT	Analytical services, tooling and access to [secure] data environments	No data environment available	Requested analysis can be undertaken by internal teams and provided back in anonymized format to data requestors	The dataset can be used in a secure data environment (SDE)	The dataset can be used in and SDE and other data and tools can be brought in as required	The dataset can be used in federated organised environment
DATA ENHANCEMENT - AUGMENTATION	The application of various techniques to make data more useable for specific purposes	No data augmentation	Some techniques to make data more useable for specific purposes	Defined techniques to make data more useable for specific purposes	Managed techniques to make data more useable for specific purposes	Comprehensive application of various techniques and mapping to data model i.e. OMOP to make more useable for specific purposes
DATA ENHANCEMENT – ENRICHMENT	Data sources enriched for example with annotations, image labels, phenomes, derivations, NLP-derived data labels	The data has no additional derived fields, or enriched data	The data include additional derived fields, or enriched data	The data include additional derived fields, or enriched data used by other available data sources	The derived fields or enriched data were generated from, or used by, a peer reviewed algorithm	The data includes derived fields or enriched data from an [inter]national report
DATA MODEL	Availability of clear, documented data model that provides structure and standardization	There is no data model	Known and accepted data model but some key fields uncoded or free text	Key fields codified using a local standard and updated over time	Key fields codified using a national or international standard and updated	Data model conforms to [inter] national standard and key fields codified using a national / international standard
DATA DICTIONARY	Provided documented data dictionary and terminologies	No data dictionary	Data definitions available	Definitions compiled into local data dictionary which is available online	Dictionary relates to national definitions	Dictionary is based on international standards and includes mapping

QUANTUM tool maturity score



Overige bevindingen volwassenheid datakwaliteit

- Het verdient de voorkeur om gebruik te maken van volwassen, door de gemeenschap gedragen en onderhouden raamwerken voor datakwaliteit, maar niet alle besproken frameworks zijn even bekend, breed toegepast of volledig ontwikkeld.
- Hoe hoger de datakwaliteitsvolwassenheid, hoe beter de kwaliteit van afzonderlijke datasets; verschillende frameworks beoordelen de volwassenheid echter op uiteenlopende niveaus, zoals het niveau van de dataset (bijvoorbeeld QUANTUM en EMA), de organisatie (zoals DAMA en NORA) of de community of discipline (zoals de EOSC Task Force).
- De verbetering van datakwaliteit hangt ook samen met de kwaliteit van metadata, data repositories, federatieve datasystemen, effectieve samenwerkingsomgevingen en gebruik en ontwikkeling van standaarden voor interoperabiliteit van infrastructuren en datasets. Samen zorgen deze voor consistentie, harmonisatie en betere bruikbaarheid van data.

MS 30 samenvatting

Het werkpakket heeft **twee (vastgelegde) hoofdactiviteiten**:

- Het ontwikkelen en promoten van een raamwerk voor datakwaliteit en ontwikkeling van een datakwaliteitslabel (taak 9.1)
- Het ontwikkelen van richtlijnen en een volwassenheidsmodel voor het verbeteren van datakwaliteit (taak 9.2).

Tussen december 2023 en juli 2025 bestond de voornaamste activiteit uit literatuuronderzoek naar bestaande raamwerken voor datakwaliteit, met name in de context van gezondheidsdata.

In het verslag zijn een aantal **relevante raamwerken en richtlijnen voor datakwaliteit** beschreven, o.a.:

- Resultaten van de Europese projecten TEHDAS1 en QUANTUM, en organisaties zoals Health Data Research (HDR UK), EOSC FAIR taskforce
- Ontwikkeling van de metadatastandaard DCAT/HealthDCAT (gebruikt in de HDAB-metadatacatalogus),
- Meer commerciële raamwerken van DAMA DMBOK, DAMA-NL, European Medicine Agency (EMA), ALCOA++ principes
- NORA (Nederlandse Referentie Architectuur) voor publieke organisaties in Nederland.

MS 30 samenvatting (ii)

Aangezien het **QUANTUM-project** reeds zowel een **labelingsysteem** als een **volwassenheidsmodel** heeft ontwikkeld binnen de EHDS-context, is besloten om niet nog een raamwerk voor datakwaliteit in Nederland te ontwikkelen, maar om de **focus te verleggen** naar:

- inzicht verkrijgen in **huidige praktijk en behoeften** rondom datakwaliteit bij Nederlandse datahouders en gebruikers van gezondheidsdata
- **de gap** tussen de vereisten van EHDS, toepasbaarheid van QUANTUM label en tool,
- **ontwikkeling van richtlijnen voor verbetering van datakwaliteit** voor secundair gebruik van data via de toekomstige HDAB.

Op dit moment is er beperkt inzicht in welke raamwerken het meest worden toegepast door datahouders en gebruikers. Om dit te achterhalen zijn consultaties met belanghebbenden gestart en in het verslag beschreven.

Vooruitkijkend zal het volgende verslag (juli 2026) voornamelijk gebaseerd zijn op de inzichten voortkomend uit meer stakeholderactiviteiten. De nadruk zal daarbij liggen op een eerste aanzet van richtlijnen voor verbetering van datakwaliteit. Ook zal trainingsmateriaal dat binnen QUANTUM ontwikkeld wordt getoetst worden bij Nederlandse stakeholders.

MS 30 Highlights (NL)

1. De verbetering van datakwaliteit hangt af van verschillende onderling samenhangende, externe factoren die zich niet op het niveau van de dataset of de datakwaliteitsvolwassenheid van de organisatie bevinden, zoals de kwaliteit en interoperabiliteit van standaarden en infrastructuren.
2. Hoogwaardige metadata biedt duidelijke beschrijvingen van de inhoud, herkomst, het formaat en het gebruik van de data, waardoor het voor gebruikers eenvoudiger wordt om datageschiktheid fit-for-purpose te beoordelen.
3. Het NORA stappenplan voor verbetering van datakwaliteit is door de gemeenschap ontwikkeld en richt zich op zeer praktische stappen en organisatorische veranderingen, waardoor het een uitstekende bron is voor uitvoerbare richtlijnen op organisatieniveau.
4. Het QUANTUM raamwerk voor datakwaliteit vormt een waardevolle basis, maar heeft wel enkele beperkingen; het zou baat hebben bij doorlopende onderhoud, evaluatie en updates op basis van feedback van datahouders, datagebruikers en HDAB's.
5. Het belang van bepaalde datakwaliteitsdimensies lijkt afhankelijk van het doel waarvoor de data wordt gebruikt, terwijl het QUANTUM datakwaliteits- en utiliteitslabel vaste wegen van bepaalde dimensies hanteert.
6. Datakwaliteit kan zowel op het niveau van datasets als op organisatieniveau worden verbeterd; het QUANTUM raamwerk richt zich vooral op datakwaliteitsvolwassenheid op datasetniveau en minder op organisatieniveau.
7. Eisen aan fit-for-purpose en fit-for-use hangen af van het perspectief (datahouder, datagebruiker); het QUANTUM-raamwerk richt zich vooral op 'fit-for-use' en verbeterde kwaliteit van data geleverd door de datahouder.

Vooruitblik 2025-2026 WP9

Verzamelen en evalueren van meest geschikte en effectieve datakwaliteitsprincipes en volwassenheidsmodellen en zo invulling te geven aan best practices en richtlijnen voor stakeholders om de datakwaliteit te verbeteren.

- Werkbezoeken specifieke organisaties
- Aansluiten bij bestaande bijeenkomsten van communities zoals Europese COP voor metadatacatalogus en datakwaliteit en de expert community secundair data gebruik
- Organiseren van workshops
- QUANTUM pilot tests van de labeling tool, bestudering deliverables en training materialen

Delen van best practices en aanbevelingen:

- Ontwikkeling richtlijnen om datakwaliteit en bruikbaarheid te verbeteren
 - Ontwikkeling richtlijnen om datakwaliteit volwassenheid te verhogen
 - Organiseren van webinar(s)
-
- Rapportage voortgang in MS 31 juli 2026

HealthDCAT-AP

Het metadataschema voor de beschrijving van gezondheidsdata

- Wat is HealthDCAT-AP?
- Wat kan ik met HealthDCAT-AP beschrijven?
- Hoe draagt metadata, en specifiek HealthDCAT-AP, bij aan datakwaliteit?



Wat is HealthDCAT-AP?

- EHDS Art. 77: '[...] De beschrijving van elke dataset bevat informatie over de **bron, de reikwijdte, de belangrijkste kenmerken en de aard** van de elektronische gezondheidsgegevens in de dataset alsook de **voorwaarden voor het beschikbaar stellen** ervan.'
- Metadata schema voor het beschrijven van datasets in het HealthData@EU catalogus
- Voor het zoeken in de catalogus op basis van vrije tekst of SPARQL query
- Nog niet officieel vastgesteld, draft beschikbaar



Wat is HealthDCAT-AP?

- Uitbreiding van generiek EU metadata schema: DCAT-AP
 - Hierdoor aansluiting aan generieke EU Data Space mogelijk
 - Voegt een aantal nieuwe metadatavelden (*properties*) voor het **gezondheidsdomein** toe
 - Specificeert **waardelijsten** voor een aantal bestaande metadatavelden (*properties*)
 - Specificeert een aantal **aanbevelingen voor het gebruik** van bestaande metadatavelden (*properties*)



Wat kan ik met HealthDCAT-AP beschrijven?

Nieuwe metadatavelden in HealthDCAT-AP

Onder andere:

- **Populatie** in de dataset
 - Het aantal **subjecten** en **records** in de dataset; leeftijdsrange
 - De **soorten persoonlijke data** die per subject verzameld zijn (bijv. leeftijd, geslacht, beroep, EPD)
- **Coding system of datamodel** gebruikt in de dataset
 - Bijv. ICD-10, FHIR
 - Specifieke codes die in de dataset gebruikt worden (bijv. ICD-10 code voor COVID-19 --> machine-leesbare identificatie van ziekte in dataset)

Wat kan ik met HealthDCAT-AP beschrijven?

Nieuwe metadatavelden in HealthDCAT-AP

- **Data sample** ('*sample distribution*'); inzicht in de structuur/opbouw van de dataset zonder data prijs te geven; welke *variabelen* zitten in de dataset?
 - Bijv. door aanleveren synthetische data of codeboek
 - Gebruiker kan inzien of een dataset geschikt is (en dus: of het nut heeft een aanvraag te starten)
 - Opzetten en testen van analysescripts voordat de data beschikbaar is
- **Analytics**: visualisaties/dashboards voor eenvoudige bevraging van de data zonder de onderliggende, ruwe data prijs te geven

Wat kan ik met HealthDCAT-AP beschrijven?

Waardelijsten

- **Categorisering** van de dataset gebaseerd op:
 - Een of meerdere van de **minimum categorieën** voor secundair datagebruik volgens de EHDS (uit **art. 51**; bijv. biobank data, genomics data, EPD data)
 - **Gezondheidsthema** van de dataset – Wikidata (bijv. COVID-19, borstkanker)

Hoe te gebruiken?

- Invoeren metadata via de **HealthDCAT-AP editor**
- Converteert metadata naar RDF format
- Valideert of metadata conform HealthDCAT-AP specificatie is
- Zelf uitproberen?
De HealthDCAT-AP editor van de HD@EU portal is [hier](#) te vinden

Dataset Description Assistant

Home > Dataset Description Assistant

Context

Non-Public: Personal health data not openly accessible, available only through a Health Data Access Body under EHDS rules.
Restricted: Non-personal health data accessible only under specific conditions (e.g. licence, registration, agreement).
Public: Non-personal health data openly accessible to everyone, without restrictions (eg : open data).

Access Level ⓘ

Non-Public ▾

Dataset Default Language

Dutch ▾

Save

Clear Content

Validate

Export ▾

Dataset 13/20

Dataset Distribution 2/2

Sample

Analytics

Dataset: A conceptual entity that represents the information published.

Mandatory **M** 13/20

Recommended **R** 0/35

Optional **O** 3/21

Title *

ⓘ A name given to the Dataset. View more ▾

My dataset

English ▾

+ Add another Title

Description *

ⓘ A free-text information about the Dataset. View more ▾

Some description of my dataset.

English ▾

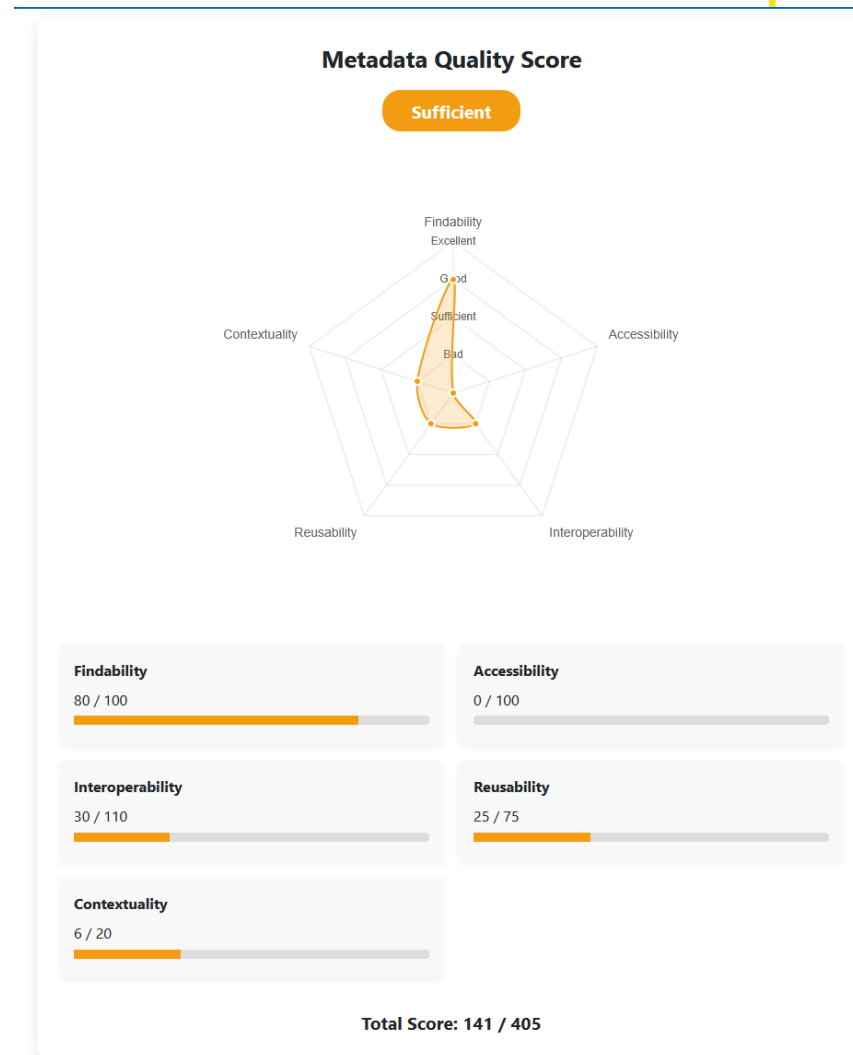
Hoe draagt metadata, en specifiek HealthDCAT-AP, bij aan datakwaliteit?

- Rechtstreeks uit Quantum dimensies:
 - Metadata scope:
 - *Existence of comprehensive standardised metadata*
 - *Existence of an exhaustive data dictionary at variable level*
- Overlap metadata velden met andere Quantum dimensies:
 - *Accessibility, data provenance*



Hoe draagt metadata, en specifiek HealthDCAT-AP, bij aan datakwaliteit?

- Rechtstreeks uit Quantum dimensies:
 - Metadata scope:
 - ***Existence of comprehensive standardised metadata***
 - ***Existence of an exhaustive data dictionary at variable level***
- Overlap metadata velden met andere Quantum dimensies:
 - ***Accessibility, data provenance***



Beschrijvende, structurele en administratieve metadata

- **Beschrijvend:** Inhoud en context van de data. Ondersteunt het zoeken naar en begrijpen van een dataset.
- **Structureel:** Beschrijft de interne structuur van de data. Helpt o.a. bij data linken en hergebruik in andere systemen.
- **Administratief:** Informatie over organisatie, management, andere administratieve aspecten van de data. Helpt o.a. bij specificeren van toegangs- en gebruiksrechten.
- Eigenschappen van HealthDCAT-AP gecategoriseerd
- Zie volledige publicatie op [Zenodo](#)

Categorizing Health-RI metadata schema version 2 (ref. 5) Dataset class properties

A description of each property can be found in the Health-RI metadata schema documentation (ref. 5).

Descriptive	Structural	Administrative
code values	analytics*	access rights
description	distribution	analytics*
documentation	coding system	applicable legislation
frequency*	conforms to	contact point
geographical coverage	in series	creator
health theme	is referenced by	frequency*
keyword	qualified attribution	issued
language	qualified relation	has version
maximum typical age	source	identifier
minimum typical age	temporal resolution	legal basis
number of records		modification date
number of unique individuals		other identifier
personal data		publisher*
population coverage		purpose*
publisher*		quality annotation
purpose*		release date
sample		retention period
temporal coverage		rights
theme		status
title		version
type		
version notes		

*Dual function and therefore in multiple categories

Hoe draagt metadata, en specifiek HealthDCAT-AP, bij aan datakwaliteit?

- Rechtstreeks uit Quantum dimensies:
 - Metadata scope:
 - *Existence of comprehensive standardised metadata*
 - ***Existence of an exhaustive data dictionary at variable level***
- Overlap metadata velden met andere Quantum dimensies:
 - *Accessibility, data provenance*



Data sample voor LINK-VACC dataset

- LINK-VACC voorbeeld dataset beschreven met HealthDCAT-AP metadata, in HD@EU catalogue sandbox:

[hier](#)

- Bijbehorende data sample (synthetische data), inzicht in alle variabelen :

[hier](#)

CSV Viewer
Parsing RDF from:

[Previous](#) Page 1 of 1000 [Next](#)

"TEST_POSITIVE_TO_COVID_NM"	"TEST_NM"	"DOSE_1_BRAND_CD"	"DOSE_1_DT"	"DOSE_2_BRAND_CD"	"DOSE_2_DT"	"DOSE_3_BRAND_CD"	"DOSE_3_DT"	"DOSES_NM"	"FULLY_VACCINATED_DT"	"FULLY_VACCINATED_BL"
1	"BP"	2022-05-02	"BP"	2021-11-20	NA	NA	2	NA	NA	FALSE
1	"BP"	2022-06-12	"NV"	2022-01-25	NA	NA	2	NA	NA	FALSE
7	"BP"	2021-09-03	"BP"	2022-07-26	NA	NA	NA	NA	NA	FALSE
5	"MD"	2021-05-20	NA	NA	NA	NA	1	NA	NA	FALSE
1	"JJ"	NA	"BP"	2022-10-22	NA	NA	2	NA	NA	FALSE
0	NA	NA	"MD"	2021-02-08	"MD"	2022-02-09	3	2022-02-24	NA	TRUE
0	"NV"	2021-11-03	"BP"	2022-06-19	NA	NA	NA	NA	NA	FALSE
6	"AZ"	2022-02-03	"MD"	2021-02-14	NA	NA	2	NA	NA	FALSE
1	"JJ"	2022-03-22	"BP"	2022-10-13	NA	NA	2	NA	NA	FALSE
6	"JJ"	2021-03-19	"MD"	2021-04-27	NA	NA	2	NA	NA	FALSE
6	"JJ"	2022-12-24	"BP"	2021-09-12	"MD"	2022-07-18	3	NA	NA	FALSE
4	"AZ"	2022-08-31	"MD"	2022-09-26	NA	NA	2	NA	NA	FALSE
2	"BP"	2022-01-11	"MD"	2021-10-19	NA	NA	2	NA	NA	FALSE
1	"JJ"	2021-04-18	"BP"	2021-12-08	NA	NA	2	NA	NA	FALSE
2	"NV"	2022-08-26	"NV"	2021-01-11	NA	NA	2	NA	NA	FALSE

We praten straks graag verder over

Metadata voor verbetering van datakwaliteit

- Wat voor type data heeft u, en is de metadata daarvoor al ergens beschikbaar?
Zo ja, waar? Wat is uw ervaring met het beschikbaar maken van de metadata?
- Welke metadata draagt bij aan de vindbaarheid van data?
- Welke metadata draagt bij aan het verbeteren van datakwaliteit?

Denk mee: www.menti.com code 5632 2880



3 flip-overs – 3 groepen

- Flap 1: Het verhogen van datakwaliteit in de huidige praktijk
- Flap 2: Gebruik van raamwerken voor verbetering datakwaliteit
- Flap 3: Metadata voor verbetering datakwaliteit

Flap 1: Het verhogen van datakwaliteit in de huidige praktijk

- Welke stappen voor verbetering van datakwaliteit voor fit-for-use?
- Welke stappen voor verbetering van datakwaliteit voor fit-for-purpose?
- Welke voordelen heeft een datahouder bij verbeteren van datakwaliteit voor secundair gebruik?

Flap 2: Gebruik van raamwerken voor verbetering datakwaliteit

- Bekend met? Ervaring met?
 - NORA stappenplan gegevenskwaliteit
 - QUANTUM dimensies en volwassenheid
 - Quantum Tool
- Lijkt het relevant? (waarom wel / niet)
- Wat zoek je in een raamwerk gegevenskwaliteit?

Flap 3: Metadata voor verbetering datakwaliteit

- Wat voor type data heeft u, en is de metadata daarvoor al ergens beschikbaar? Zo ja, waar? Wat is uw ervaring met beschikbaar maken van de metadata?
- Welke metadata draagt bij aan de vindbaarheid van data?
- Welke metadata draagt bij aan het verbeteren van datakwaliteit?

Wrap up

- Highlights van de flip-overs?
- www.menti.com code 5632 2880





Bedankt



Metadata

[Expertgroep Datakwaliteit Metadata.pptx](#)