

Ontwikkeling van een betrouwbaarheidslabel *voor informatieproducten*



NNK- 25 januari 2024 Annemiek Teuben – Coördinator datakwaliteit

ammb.teuben@rotterdam.nl



**Gemeente
Rotterdam**

Annemiek Teuben

Mei 2022 Coördinator datakwaliteit bij Datadiensten

2008-2022 Erasmus Universiteit Rotterdam
Coördinator Onderzoeksinformatiesystemen,
Informatie-analist, Functioneel beheerder

1998-2008 Pink Roccade Implementatie consultant en BI
consultant

1989-1995 Gezondheidswetenschappen, Maastricht
Richtingen: biomedische wetenschappen en klinische
psychologie





Feiten over Rotterdam



- >175 nationaliteiten
- Op één na grootste stad in Nederland (>666.000 inwoners)
- 4 van de 5 hoogste wolkenkrabbers in Rotterdam
- Meeste slachtoffers doodslag of moord (2022)
- Laagste opkomst bij 2^e kamerverkiezingen (2023)
- “Bijstandshoofdstad” (al 12 jaar)
- Rotterdam in de top 10 Lonely Planet (2016)
- >900 applicaties (gemeente)
- 150.000 datasets (schatting) (gemeente)
- Jaarlijkse uitgaven 4 miljard (gemeente)

Steeds meer keurmerken om ons heen....

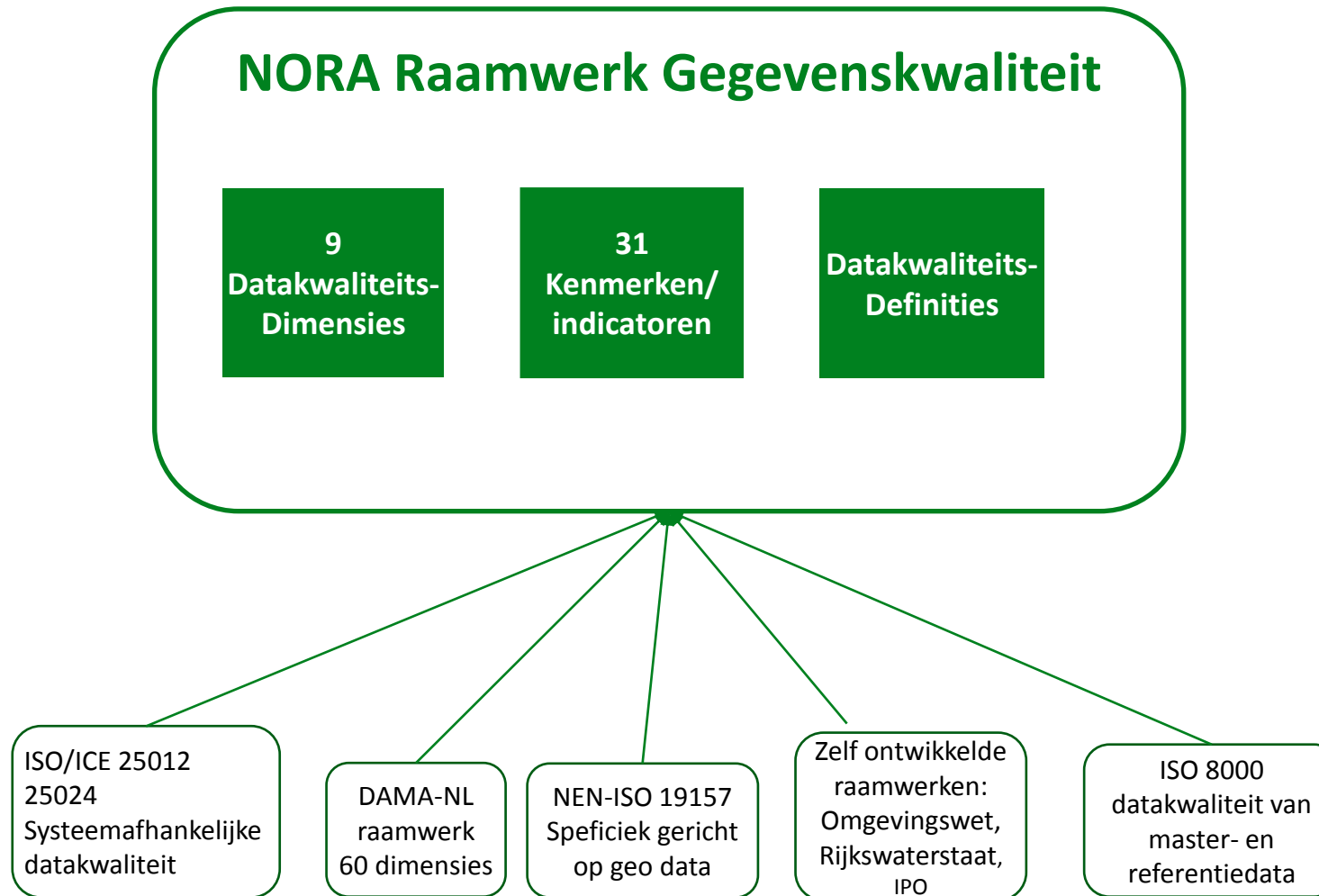


	 Overlast regen	 Hitte	 Droogte	 Overstroming
A	Geen water	< 38°	< 1 m	< 20 cm
B	< 5 cm	38° - 40°	1 - 2 m	20 - 50 cm
C	5 - 10 cm	40° - 42°	2 - 4 m	50 - 200 cm
D	10 - 23 cm	42° - 44°	4 - 8 m	200 - 500 cm
E	> 23 cm	> 44°	> 8 m	> 500 cm
	Verskil maximale waterdiepte en drempelhoogte (uitgaande van een hevige bui van 93 mm/70 min)	De gevoelstemperatuur op een hele warme zomerdag rondom uw woning (peildatum 1 juli 2015)	De diepste grondwaterstand in de zomer rondom uw woning	De hoogte van het water in uw huis als gevolg van een dijkdoorbraak (gebaseerd op kaarten Europese Richtlijn Overstromingsrisico's)



Welk voertuig is beter?





ISO 27001
informatiebeveiliging
(BIV zeker stellen)

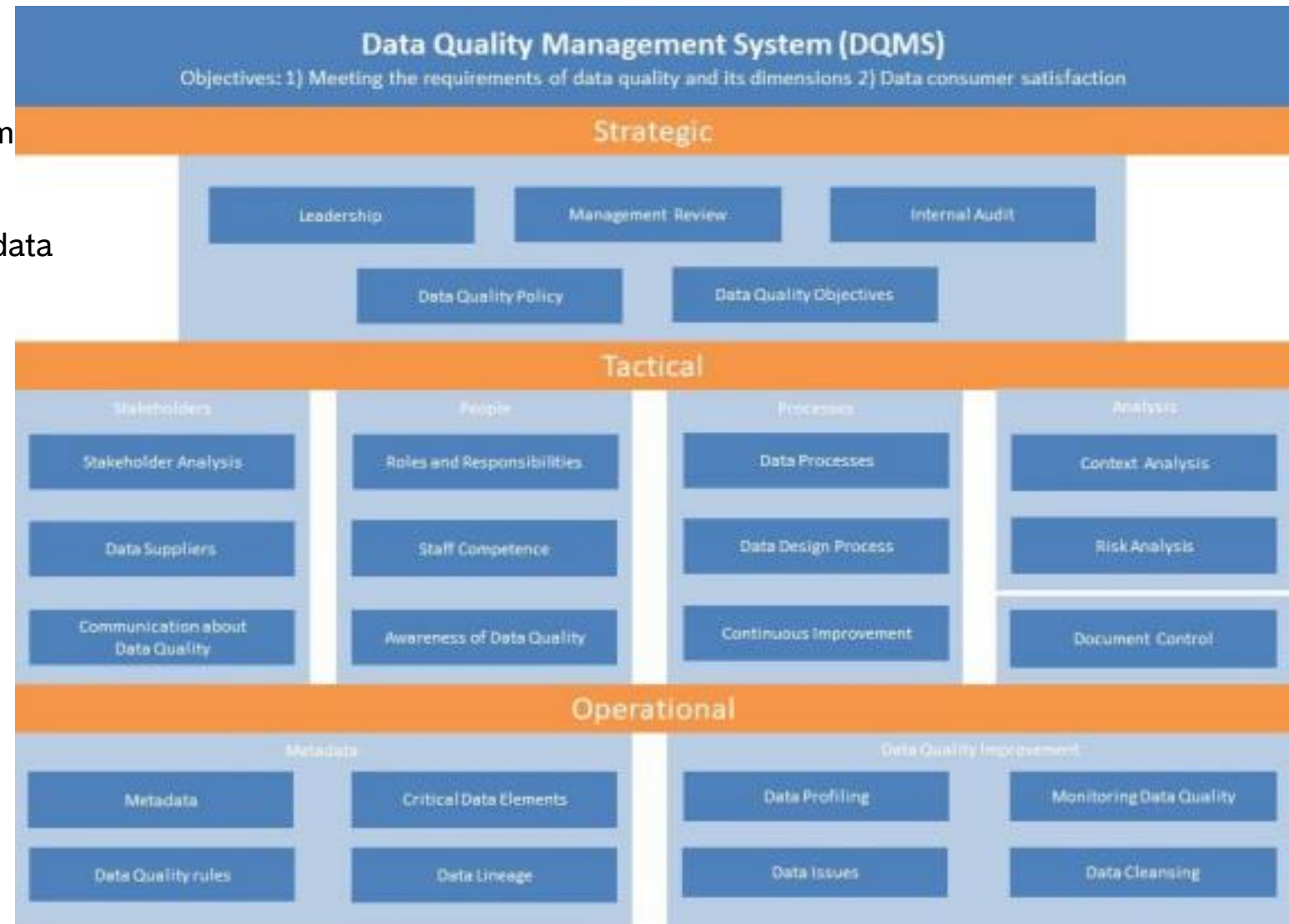
Algoritme keurmerk
bestaat nog niet

* NORA: Nederlandse Overheid Referentie Architectuur

Keurmerken voor kwaliteitsmanagement

ISO 9001 – kwaliteitsmanagementsysteem voor producten en systemen

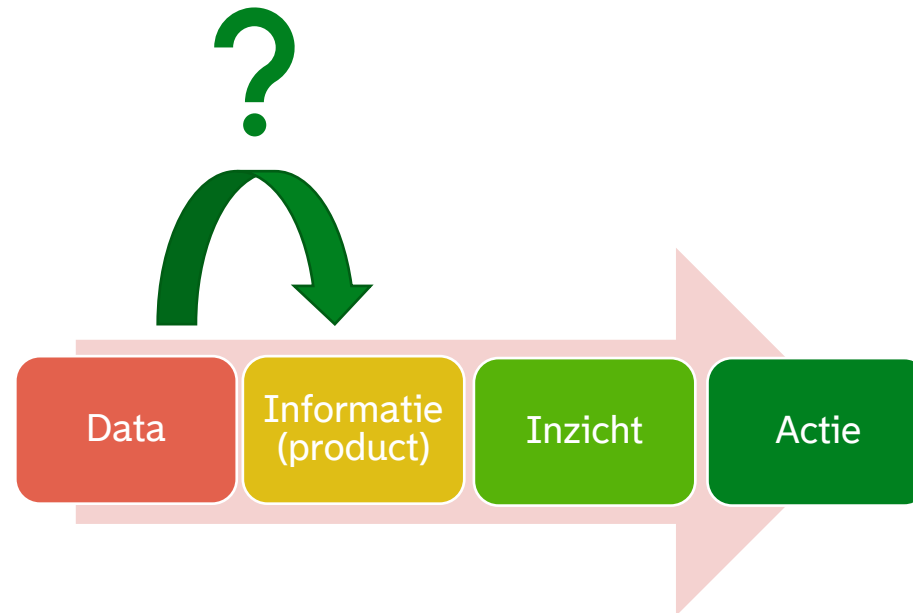
DAMA-NL heeft dit systeem toegepast op data



Ontwikkeling betrouwbaarheidslabel voor informatieproducten



Waarom willen we een label?



Transparantie over de kansen en risico's van een informatieproduct in beeld brengen

Communicatiemiddel om de verantwoordelijkheden op de juiste plek te leggen

Wat is een betrouwbaarheidslabel van een informatieproduct?

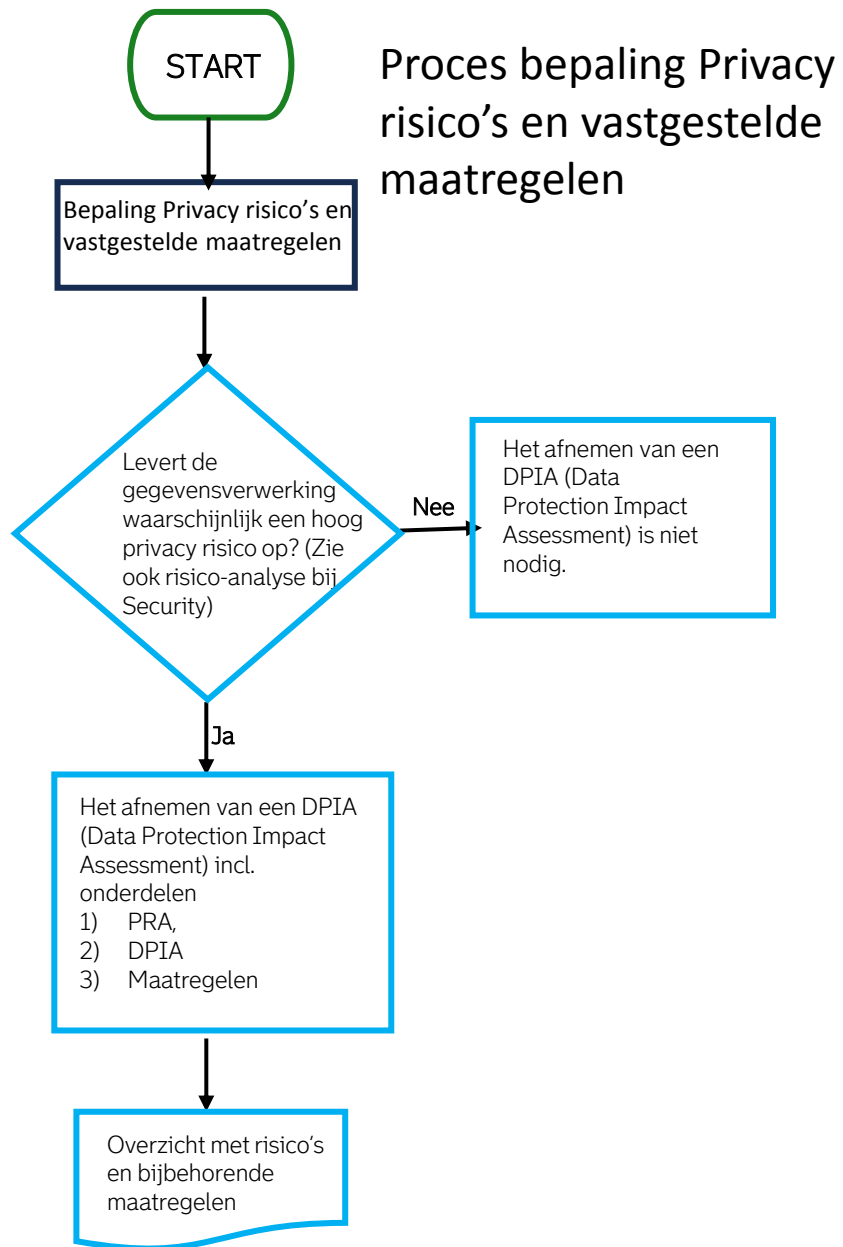
“Een betrouwbaarheidslabel zegt iets over de balans tussen waarde kunnen halen uit data én de risico’s op het gebied van datakwaliteit, veiligheid, privacy, beschikbaarheid, interpretatie, ethiek of reputatie.”

“Een betrouwbaarheidslabel zegt iets over de volwassenheid van een informatieproduct.”

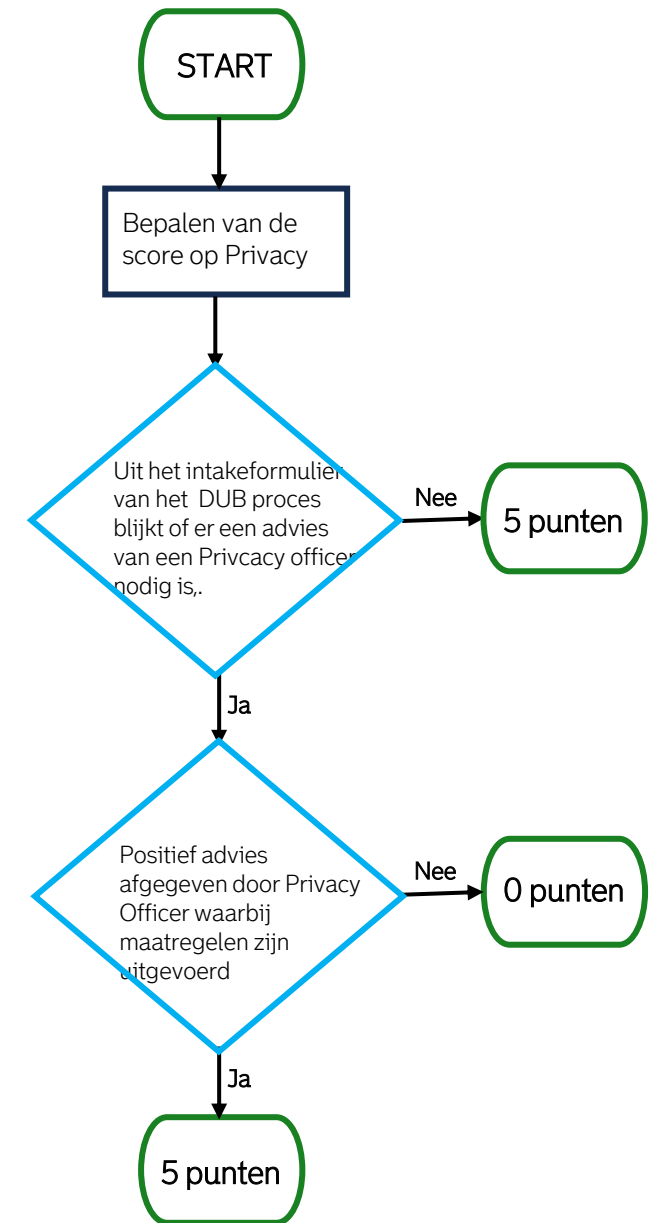
11 criteria voor bepaling betrouwbaarheidslabel

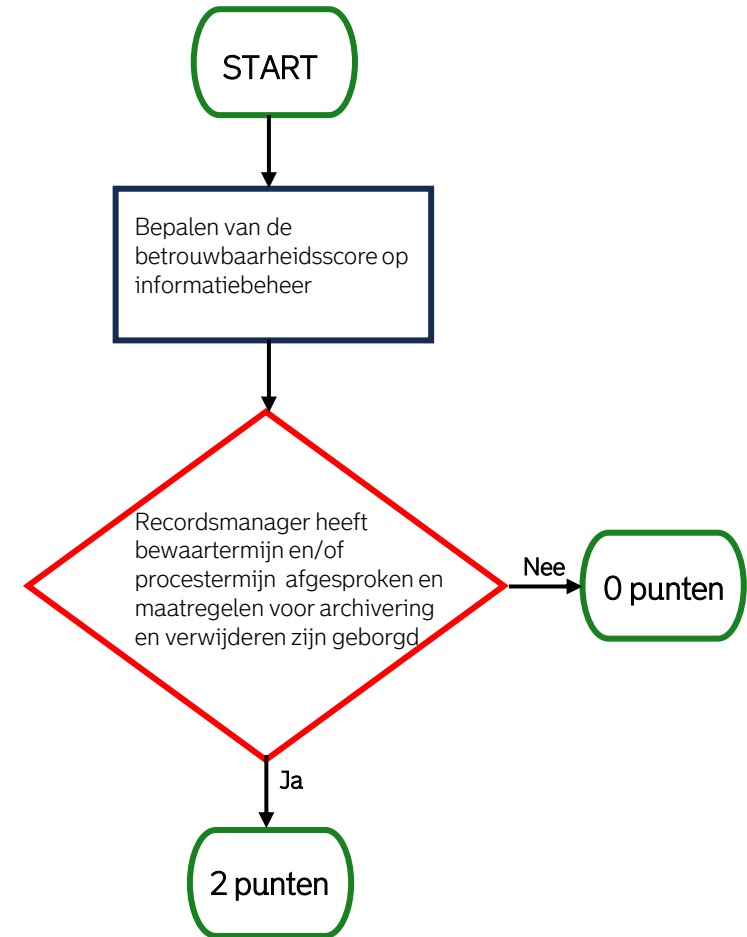
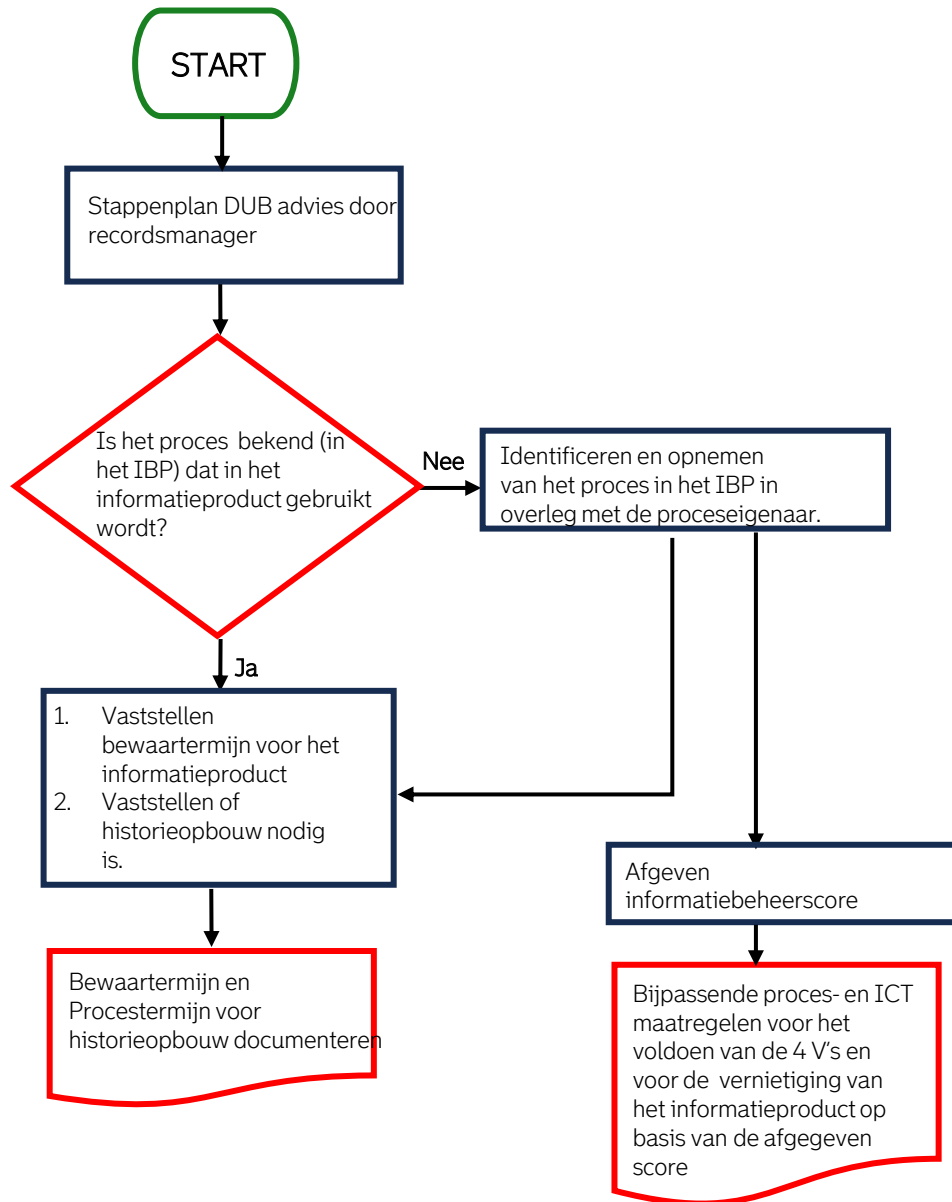
Dimensie	Definitie	
Compliance kenmerken	Privacy (AVG)	Er is een positief advies van privacy officer is gegeven (op basis van een Privacy Risk Assessment.
	Beveiliging	Er is een positief advies afgegeven door de security officer.
	Ethiek	Er is een positief advies afgegeven door de ethiek adviseur.
	Informatiebeheer	De informatie is Vindbaar, Volledig, Veilig en Voldoet aan de Wet & Regelgevingseisen Ook is een bewaartermijn afgesproken in overleg met de recordsmanager.
	Algoritmen	Er is een positief advies afgegeven door de algoritme expert
	Eigenaarschap	De data-eigenaar van de bron(nen) én de eigenaar van het informatieproduct zijn bekend.
Datakwaliteit kenmerken	Datakwaliteit	Er is een datakwaliteitsdashboard ontwikkeld voor het informatieproduct, dat in elk geval de datakwaliteit toont van de drie belangrijkste dimensies te weten volledigheid, juistheid en actualiteit. Voor dit dashboard wordt de data uit het informatieproduct getoond. Daarnaast wordt eventueel aanvullend een dashboard per afzonderlijke databron opgebouwd. De kwaliteitseisen die gesteld worden aan het informatieproduct zijn afgesproken met de eigenaar van het informatieproduct. Afhankelijk van de gemeten datakwaliteit scoort een informatieproduct hoger dan wel lager.
Kenmerken over ontwikkeling product	Definities	Het is duidelijk wat elk gegeven betekent. Er zijn conceptuele en logische modellen. De bedrijfsregels (business logica) zijn gevalideerd en gedocumenteerd.
	Datalineage	Het proces waarmee data vanuit vele systemen (automatisch) wordt opgewerkt tot een tastbaar data product verloopt via real time berichtenverkeer over Data en Analyse platform van Rotterdam (DAP'R). Aanwezigheid van dit berichtenverkeer identificeert de metadata over de databronnen en identificeert ook de transformaties die het heeft doorgemaakt tot het punt van toepassing. Aanwezigheid van deze metadata én de juistheid en actualiteit van deze metadata zijn bepalend voor de score.
	Specificaties	<u>Functioneel ontwerp:</u> Doel, Wettelijke kaders, Doelgroep, Gebruikte bron, locatie en Layout; <u>Technisch ontwerp:</u> Relaties en afhankelijkheden met andere bronnen, Layout, Queries, Measures, Wijze van vernieuwen en distributie;
	Acceptatietest	Acceptatie test is uitgevoerd door een collega data-analist én daarnaast door de afnemer van het product vóór ingebruikname en heeft betrekking op de punten: databron, definities, validiteit, performance en gebruiksgemak.

PRIVACY Stroomschema - Betrouwbaarheidsscore



Vaststellen score





DATAKWALITEIT Stroomschema - Betrouwbaarheidsscore

Aanwezigheid van een datakwaliteitsdashboard voldoet minimaal aan de volgende criteria:

Verplichte dimensies:

- Volledigheid
- Actualiteit
- Juistheid*

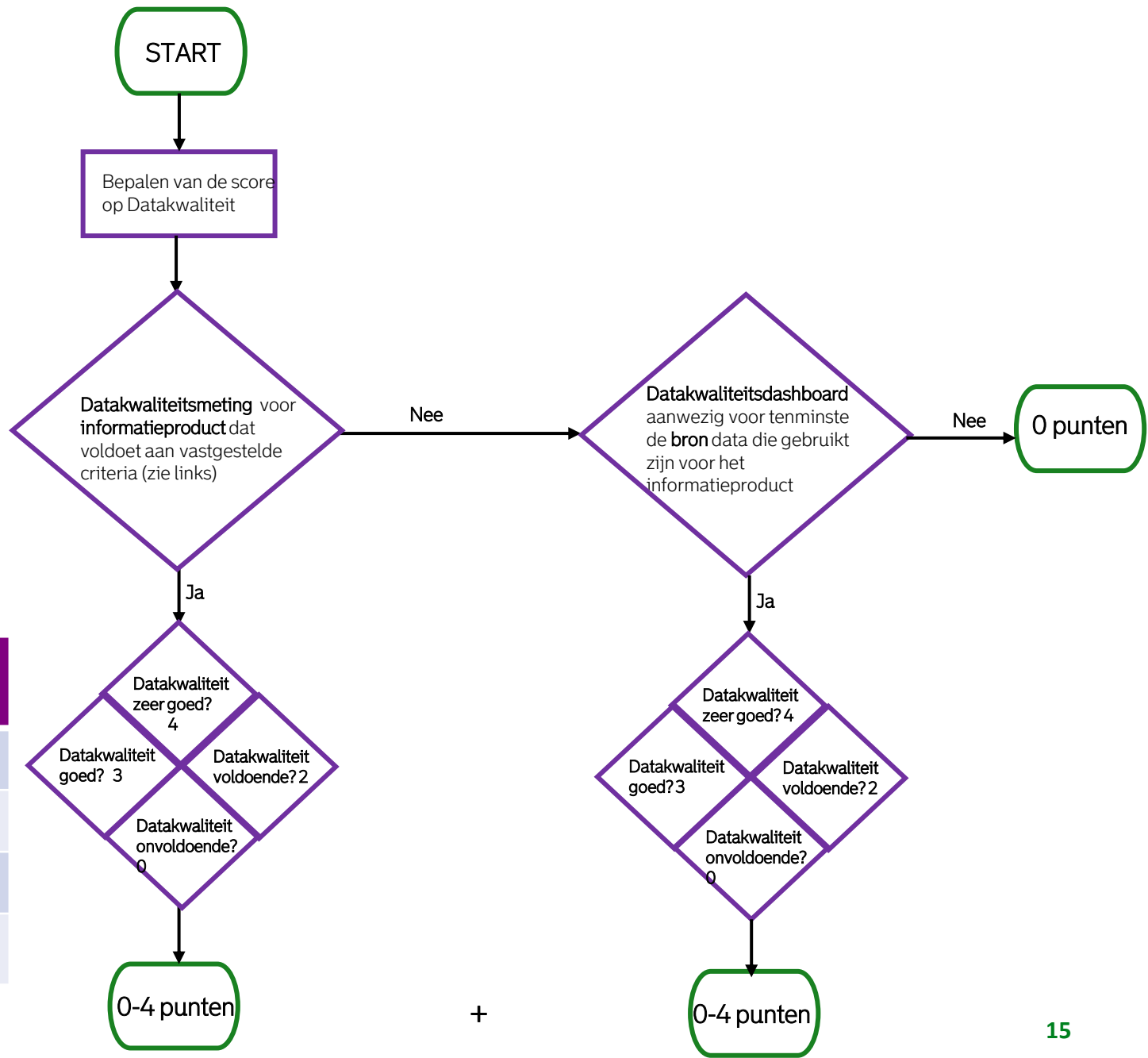
Optionele dimensies:

- Validiteit*
- Consistentie*

*Voor de kritieke data-elementen zijn kwaliteitsregels gedefinieerd.

Kwaliteitsnormen voor datakwaliteit zijn door de eigenaar van het informatieproduct vastgesteld.

DQ dashboard Bron(nen)/Informatieproduct	Zeer goed	Goed	Voldoende	Onvoldoende
Zeer goed	8	7	6	4
Goed	7	6	5	3
Voldoende	6	5	4	2
Onvoldoende of onbekend	4	3	2	0



Het totaal aantal gewogen punten en vertaling naar label

criterium	Opties(n) per criterium	Mogelijkheden punttoekenning	A	B	C	D	E
Privacy	2	0,5	5	5	5	5	variabel
Beveiliging	2	0,2	2	2	2	2	variabel
Ethiek	2	0,2	2	2	2	variabel	variabel
Algoritme	2	0,2	2	2	2	variabel	variabel
Informatie-beheer	2	0,2	2	2	2	variabel	variabel
Eigenaarschap	3	0,2,5	5	5	5	variabel	variabel
Datakwaliteit	4	0,2,3,4,5,6,7,8	7 of 8	5, 6, 7 of 8	minimaal 4	variabel	variabel
Datalineage	3	0,2,3,5	5	variabel	variabel	variabel	variabel
Specificaties	3	0,2,3,5	5	variabel	variabel	variabel	variabel
Definities	3	0,2,3,5	5	variabel	variabel	variabel	variabel
Acceptatietest	3	0,3,5	5	variabel	variabel	variabel	variabel
Puntentotaal			42-46	37-41	31-36	8-30	0-7

Gekozen rekenregels:

Score A: Score op Privacy, Beveiliging, Ethiek, Algoritme, Informatiebeheer en Eigenaarschap is altijd 18. Datakwaliteit is altijd 7 of 8. Totaal aantal punten tussen 42-26

Score B: Score op Privacy, Beveiliging, Ethiek, Algoritme, Informatiebeheer en Eigenaarschap is altijd 18. Datakwaliteit is minimaal 6. Het totaal aantal punten ligt tussen de 37-41

Score C: Score op Privacy, Beveiliging, Ethiek, Algoritme, Informatiebeheer en Eigenaarschap is altijd 18. Datakwaliteit is minimaal 4. Het totaal aantal punten ligt tussen de 31-35

Score D: Score op Privacy en Beveiliging = 7, Het totaal aantal punten ligt tussen 8-12

Score E: Totaal aantal punten is 7 of lager. Compliancy, datakwaliteit en informatieproductkenmerken onvoldoende

Beschrijving rapport per label (Gemeentebreed)

Voldoet aan
compliance eisen

A

Het informatieproduct is zeer betrouwbaar. Het risico bij het gebruik van het informatieproduct is zeer laag. Het voldoet aan de uitgangspunten van de AVG, met inachtneming van privacy, veiligheid en informatiebeheer. Daarnaast zijn de ethische kaders getoetst. Een datakwaliteitsdashboard is aanwezig voor het informatieproduct zelf. De datakwaliteit, beschikbaarheid en performance zijn zeer hoog. Het informatieproduct is inzetbaar ter ondersteuning van het primaire proces.

B

Het informatieproduct is betrouwbaar. Het risico bij het gebruik van het informatieproduct is laag. Het voldoet aan de uitgangspunten van de AVG, met inachtneming van de privacy, veiligheid en informatiebeheer. Daarnaast zijn de ethische kaders getoetst. Een datakwaliteitsdashboard is aanwezig voor het informatieproduct zelf. De datakwaliteit van het product is hoog. De gegevens worden op tijd ververst en de beschikbaarheid en performance is goed. Het product is inzetbaar voor besturing en verantwoording van de organisatie.

C

Het informatieproduct is betrouwbaar. Het risico bij het gebruik van het informatieproduct is gemiddeld. Het voldoet aan de uitgangspunten van de AVG, met inachtneming van privacy, veiligheid en informatiebeheer. Daarnaast zijn de ethische kaders getoetst. Een datakwaliteitsdashboard is aanwezig voor het informatieproduct zelf. De datakwaliteit, beschikbaarheid, specificaties, definities en performance zijn voldoende tot goed. Het product is inzetbaar voor de sturing van afdelingen of processen van de organisatie.

D

Het informatieproduct kent een hoog risico. Het voldoet aan de uitgangspunten van de AVG, met inachtneming van de privacy, veiligheid en informatiebeheer. Daarnaast zijn de ethische kaders getoetst. Echter, de datakwaliteit is echter onbekend of onvoldoende, de definities en/of specificaties zijn niet aanwezig en de data zijn niet actueel. Het product is inzetbaar voor experimentele doeleinden.

E

Het informatieproduct kent een zeer hoog risico. Het product voldoet niet aan de uitgangspunten van de AVG en de ethiek kaders zijn niet getoetst. Daarnaast zijn de datakwaliteit, specificaties, definities en acceptatie van het product onbekend of onvoldoende

Voldoet NIET aan
compliance eisen

Voor wie is het label bedoeld?

- Afnemers van informatieproducten (beleidsmedewerkers, onderzoekers, managers, etc.);
>>Misschien ook voor inwoners, ondernemers en toeristen (Digitale stad)
- De data-eigenaren van de bronnen die gebruikt zijn voor het informatieproduct;
- Data-eigenaren van het informatieproduct;

Vragen?

