



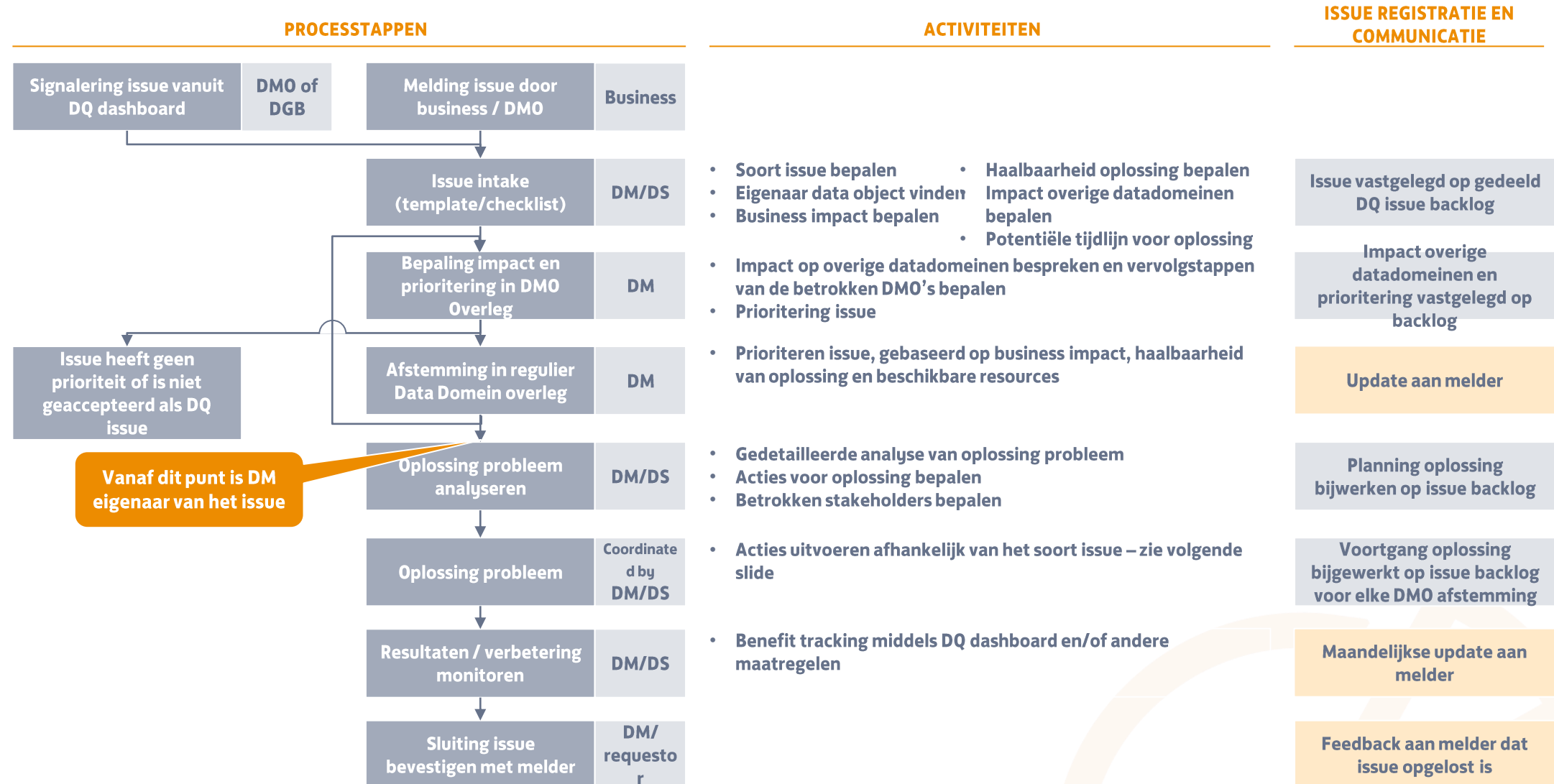
Data Quality Issue Management

December 2021

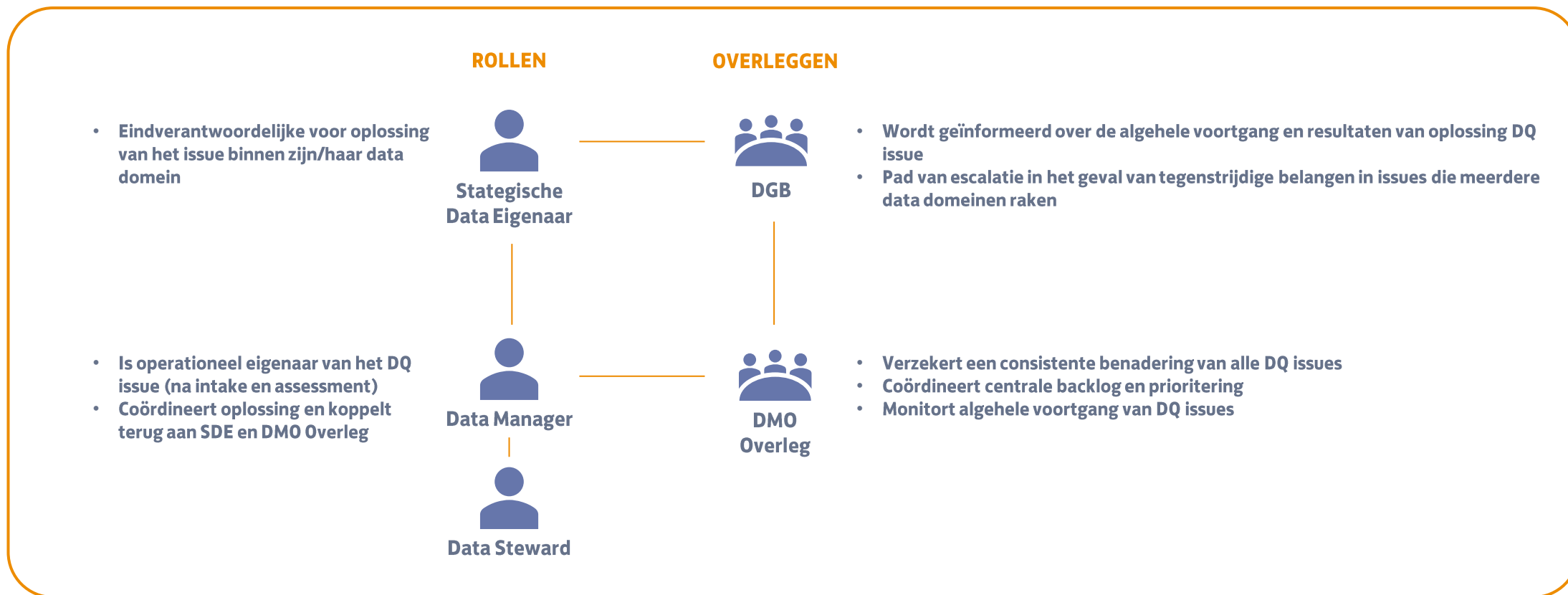
Enterprise Data Management
(EDM)



DQ issue management process



Rollen en governance van DQ issue management



DQ issues en oplossingen

SOORTEN DATA QUALITY ISSUES

- Onvolledige data
- Onnauwkeurige data
- Inconsistente data
- Invalide data
- Gedupliceerde data (uniciteit)
- Onvoldoende tijdigheid

Andere data issues kunnen ook gesignaleerd worden in dit proces, maar dat zijn niet altijd DQ issues, bijvoorbeeld:

- Missende velden in het ontwerp van het data object
- Compliance issues

TO DO: proces voor non-DQ data issues vaststellen

CATEGORIEËN OPLOSSINGEN

Correcte data

- Handmatig opvoeren of corrigeren data
- Ontwikkelen en draaien update scripts om data te corrigeren
- Data pensioneren (of inactief maken)

Verbetering van data opvoer

- Verandering van data opvoer processen/procedures
- Verandering geautomatiseerde data inname, van sensoren of geautomatiseerde input vanuit andere bronnen (inclusief scheduling van batch processen)

Verbetering life cycle management

- Toepassen of veranderen vervaldata for data validiteit
- Update data retentie (beleid)

Compliance

- Toegankelijkheid van data
- Retentie (beleid) van data

Veranderingen in architectuur

- Veranderingen data model
- Veranderingen architectuur van applicaties

Vereisten en vervolgstappen (to do's)

PREREQUISITES FOR ADEQUATE DQ ISSUE MANAGEMENT

- De eigenaar van het data object is bekend
 - Voor bronsystemen, data eigenaarschap voor elk data object wordt nu vastgesteld als follow-up van DataFit
 - Voor het data warehouse wordt eigenaarschap van data objecten/tabellen niet consequent vastgesteld en dat moet ook gebeuren
- Data Managers en Data Stewards moeten opgeleid worden in het DQ issue management proces, in het bijzonder in:
 - Intake issues
 - Voorbereiden voor prioritering/besluiten nemen in DMO afstemming en data domein overleg
 - Het coördineren van de oplossing van issues en het monitoren van de vooruitgang en resultaten
 - Consequent vastleggen en communiceren gedurende het hele proces

TO DO

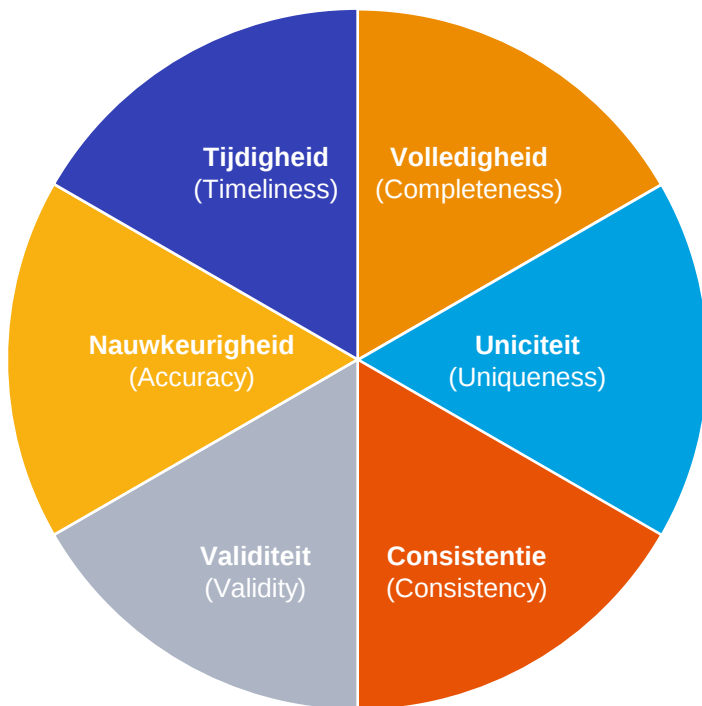
- Maken van een (simpele) template voor intake issue
- Gedeelde (centrale) backlog voor DQ issues opzetten, beschikbaar voor alle DMOs
- Afstemmen met Data Scientists en BI developers wat hun verwachtingen van DQ issue management zijn
- Door het DQ issue management proces heen lopen met (nieuwe) Data Managers en Data Stewards
- Afstemmen met digital consultants/rollen in het agile proces

Appendix

Definitie, scope en positionering data kwaliteit

Voor héél PostNL én voor álle data

Datakwaliteit is de mate waarin de (hergebruikte) data(set) geschikt is voor het doel waarvoor het gebruikt wordt. **Data Quality Management** betreft de processen en technologieën om dat voor de gehele PostNL organisatie én voor álle data op een structurele manier in te richten en te borgen. Het is een programma, géén project.



- **Volledigheid**: Mate waarin verwachte data aanwezig is
- **Uniciteit**: Mate waarin de data uniek is, afhankelijk van het gedefinieerde niveau
- **Tijdigheid**: Mate waarin data beschikbaar is op het moment dat het nodig is
- **Validiteit**: Mate waarin de data valt binnen het gedefinieerde formaat, type en reeks
- **Nauwkeurigheid**: Mate waarin de data overeenkomt met de werkelijkheid
- **Consistentie**: Mate waarin de data tussen verschillende datasets overeenkomt