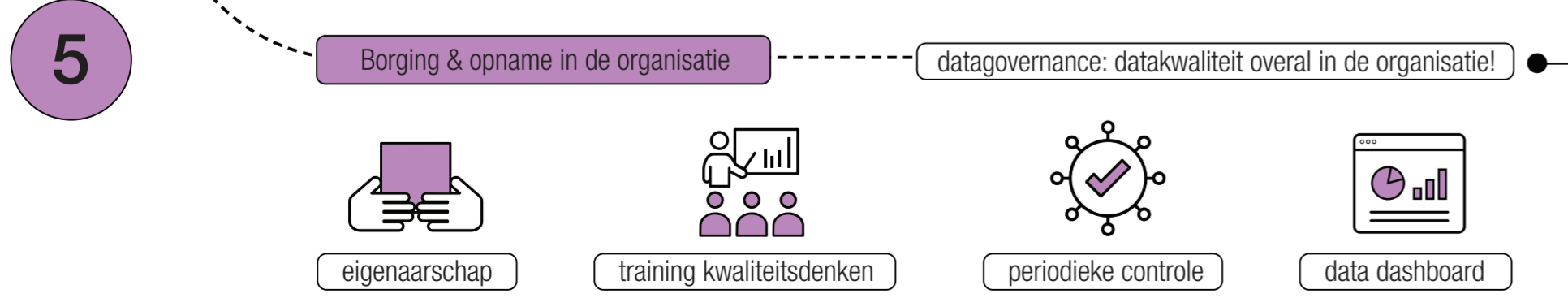
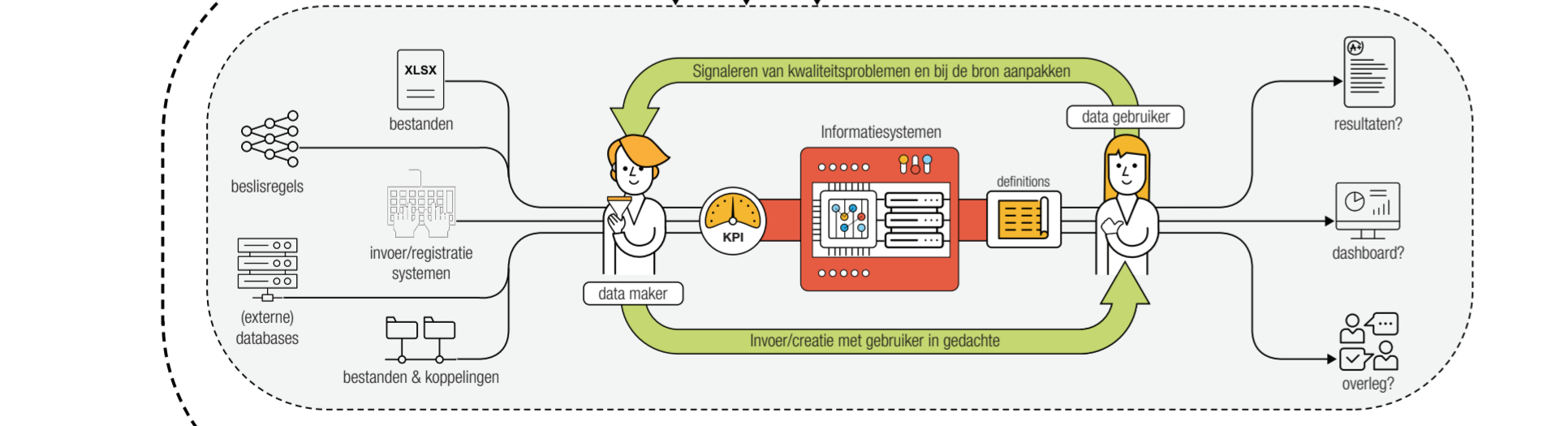
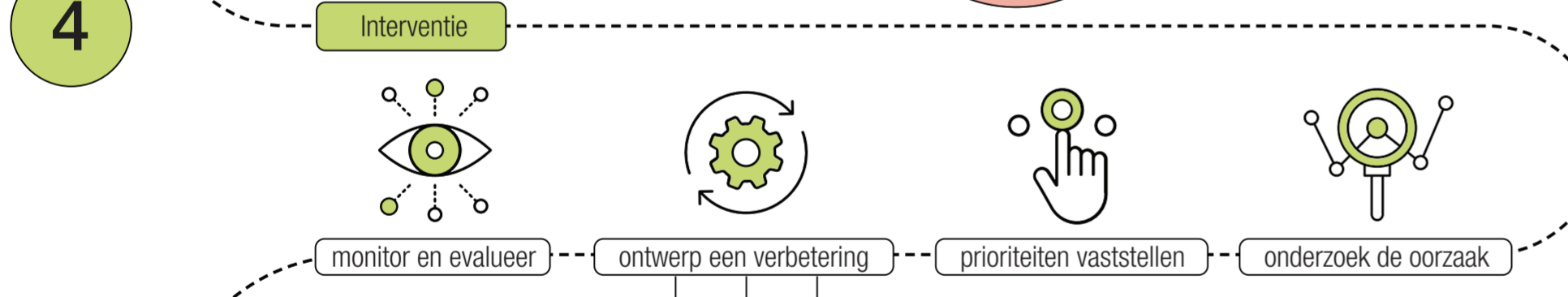
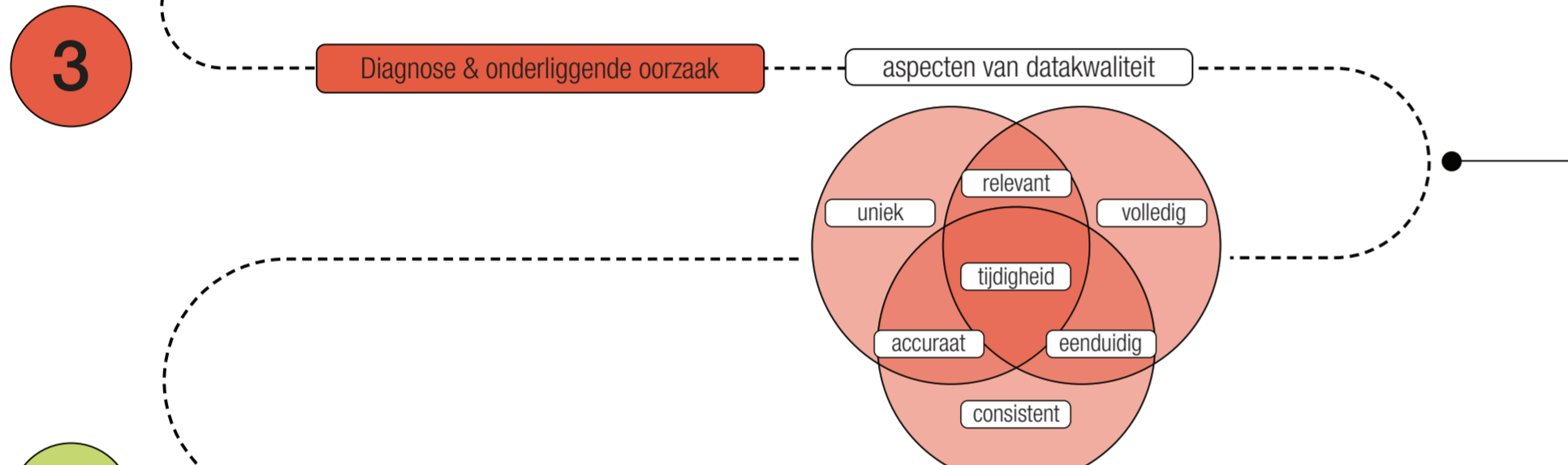
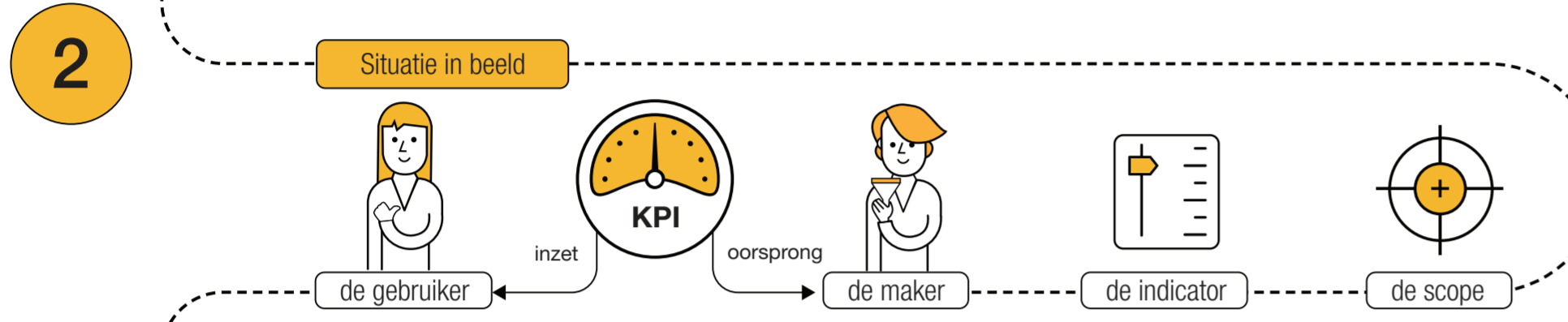
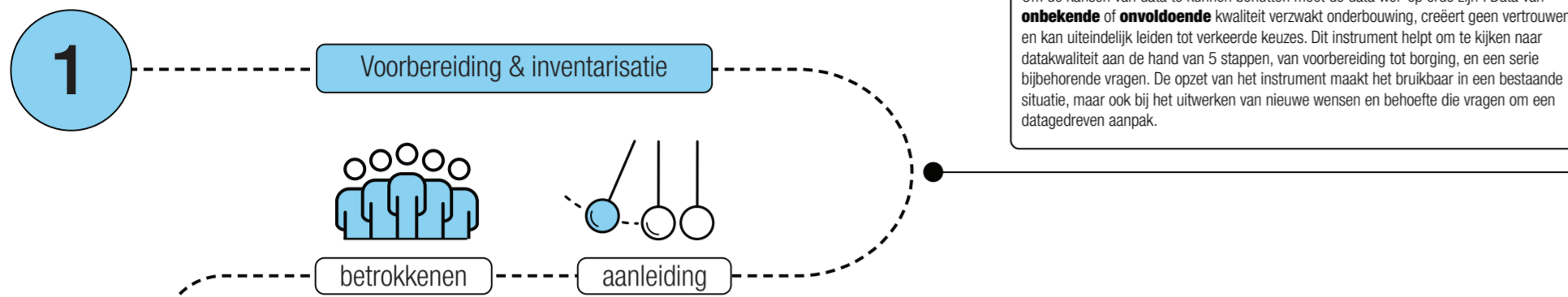


# Cheatsheet: onderzoeken en verbeteren van datakwaliteit

## Over deze cheatsheet

Om de kansen van data te kunnen benutten moet de data wel 'op orde zijn'. Data van **onbekende of onvoldoende** kwaliteit verzwakt onderbouwing, creëert geen vertrouwen en kan uiteindelijk leiden tot verkeerde keuzes. Dit instrument helpt om te kijken naar datakwaliteit aan de hand van 5 stappen, van voorbereiding tot borging, en een serie bijbehorende vragen. De opzet van het instrument maakt het bruikbaar in een bestaande situatie, maar ook bij het uitwerken van nieuwe wensen en behoeftes die vragen om een datagedreven aanpak.



## Verdiepende vragen

Deze verdiepende vragen helpen om te onderzoeken waar de belangrijkste aandachtspunten zitten. Gebruik ze als startpunt of ter begeleiding en vraag verder waar het nodig is.

- 1 aanleiding**
- 1.1 Wat is de aanleiding/rede om dit stappenplan te doorlopen?
  - 1.2 Is er een bestaande visie en beleid met betrekking tot datakwaliteit?
  - 1.3 Hoe wordt datakwaliteit nu in kaart gebracht? Zijn daarvoor metingen of raamwerken?
  - 1.4 Hoe signaleren we dataproblemen, welke processen hebben we daarvoor?
- 2 betrokkenen**
- 2.1 Hoe hebben we datakwaliteit in de organisatie beled? Welke medewerkers zijn er betrokken en welke rollen/verantwoordelijkheden zie ik?
  - 2.2 Is er voldoende mandaat voor de betrokken medewerkers om datakwaliteit te monitoren en verbeteren?

- 1 de scope**
- 1.1 Wat is de focus van dit onderzoek? Is dat een onderzoek van de kwaliteit van een definitie, een bronstelsel, een indicator of een weergave via een rapport/dashboard?
  - 1.2 Welk rapport/dashboard/indicator kijken we naar?
  - 1.3 Welk doel heeft dit rapport/dashboard/indicator?
  - 1.4 Wie is de gebruiker?
  - 1.5 Hoe is dit rapport/dashboard/indicator ontstaan?
- 2 de indicator**
- 2.1 Welke definitie wordt gebruikt?
  - 2.2 Uit welke onderdelen bestaat het rapport/dashboard/indicator?
  - 2.3 Wat is de bron?
- 3 de maker**
- 3.1 Wie legt / leggen de gegevens vast?
  - 3.2 Is er een eigenaar/ verantwoordelijke?
  - 3.3 Legt een persoon de gegevens vast of komt het uit een ander (extern) systeem?
  - 3.4 Is de maker tevens gebruiker? Of is de maker op de hoogte van het gebruik?
- 4 de gebruiker**
- 4.1 Binnen welke processen wordt het gebruikt?
  - 4.2 Op welke momenten (datum & tijdstip) wordt de data gebruikt?
  - 4.3 Wat is de benodigde kwaliteit?
  - 4.4 Is er een kwaliteitsnorm vastgesteld? Bijvoorbeeld in de vorm van % gevuld, niet ouder dan x dagen etc.

- A uniek**
- A.1 Is er sprake van dubbelleningen?
  - A.2 Hoe is dit vastgesteld?
  - A.3 Hoe kan dit zijn ontstaan?
  - A.4 Wat is de impact?
- B relevant**
- B.1 Is de gepresenteerde indicator nodig voor het beoogde doel?
  - B.2 Is de gebruikte data en het detailniveau van de data passend bij het doel?
- C volledig**
- C.1 Is er sprake van ontbrekende data?
  - C.2 Hoe is dit vastgesteld?
  - C.3 Is er een normering voor hoeveel data er mag ontbreken?
  - C.4 Wordt er gebruik gemaakt van historische data? Zo ja, hoe ver willen we daarin terug kunnen kijken?
- D eenduidig**
- D.1 Is er een heldere, werkbare definitie opgesteld?
  - D.2 Wordt die definitie overal gehanteerd (zowel door systemen als actoren)?
- E consistent**
- E.1 Zijn er tegenstrijdige bronnen? Bijvoorbeeld: verschillende aantallen studenten in verschillende bronnen?
  - E.2 Zijn er gegevens die verschillende informatie geven? Bijvoorbeeld: studenten die zijn uitgeschreven maar nog wel ingelogd zijn in de ELO?
  - E.3 Hoe wordt omgegaan met veranderende dimensies? Bijvoorbeeld veranderende opleidingsnamen of organisatieindelingen?
- F accuraat**
- F.1 Geeft de data de werkelijke situatie weer?
  - F.2 Is er een beoogde kwaliteit?
  - F.3 Is het detailniveau passend bij het doel?
- G tijdigheid**
- G.1 Is er sprake van tijdsensitieve data?
  - G.2 Wanneer zou deze data beschikbaar moeten zijn?
  - G.3 Hoe vaak wordt de data ververs/ingeladen? Komt dat overeen met de behoefte?
- H validiteit**
- H.1 Wordt de data vastgelegd in een juist formaat?

- 1 onderzoek de oorzaak**
- 1.1 Wat zijn de belangrijkste problemen samengevat?
  - 1.2 Hoe is het ontstaan?
  - 1.3 Is er sprake van incidentele of structurele oorzaak?
  - 1.4 Wat voor soort probleem is het (bv vastlegging, technisch, werkspraken)?
- 2 prioriteiten vaststellen**
- 2.1 Waar hebben we het meeste last van?
  - 2.2 Wat zijn de grootste risico's?
- 3 ontwerp een verbetering**
- 3.1 Welke verbeteringen zien we in het kader van I.) Fouterstel en II.) Procesverbeteringen
  - 3.2 Hoe kunnen we de verbeteringen doorvoeren?
    - 3.2.1 Wie is er verantwoordelijk?
    - 3.2.2 Wanneer kunnen we verbeteringen het beste doorvoeren?
    - 3.2.3 Hoe meten we het effect?
- 4 monitor en evalueer**
- 4.1 Doorvoeren verbeteringen
  - 4.2 Testen oplossing

- A eigenaarschap**
- Zorg voor eigenaarschap bij de proceseigenaren (en niet alleen bij IT of informatievoorziening)
- B training kwaliteitsdenken**
- Training in kwaliteitsdenken op datagebied
- C periodieke controle**
- Maak het makkelijk om issues te melden en zorg periodieke controle van je belangrijkste data
- D data dashboard**
- Data dashboard voor de controle van de gegevens