

# Data Foundations

## Hoofdstuk 4

## Data Management

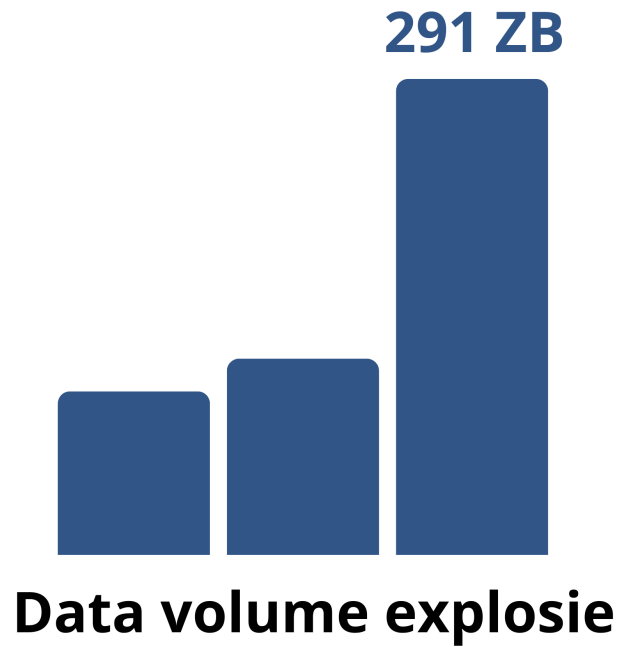
Hassan Haddouchi

In samenwerking met Collibra



# Agenda

1. Wat is data management?
2. Metadata management
3. Data governance
4. Data management in de praktijk



**\$12.9M**

**Slechte datakwaliteit**

**\$31M**

**Privacy uitgaven**

- Gemiddeld kost slechte datakwaliteit organisaties \$ 12,9 miljoen per jaar.
- Het kan organisaties tot \$ 31 miljoen kosten om te voldoen aan de privacyregelgeving.
- International Data Corporation (IDC) voorspelt dat de datasfeer zal groeien van 101 ZB in 2022 tot 291 ZB in 2027.

# Waarom hebben we data management nodig?

1. Data zonder bruikbaarheid is nutteloos.
2. Datagedreven besluitvorming elimineert vooroordelen en subjectieve gevoelens.
3. Data = macht.

**Om data waardevol te maken, moet data bruikbaar zijn.**

# Hoe kunnen we data bruikbaar maken?

1. **Zoek:** vind de benodigde data.
2. **Begrijp:** weet wat de data betekenen en wat het doel ervan is.
3. **Vertrouw:** zorg voor de kwaliteit en nauwkeurigheid van de data.

4. **Werk samen:** deel en publiceer data effectief.
5. **Toegang:** krijg snel toegang tot data en gebruik deze op een manier die aan de wet voldoet.

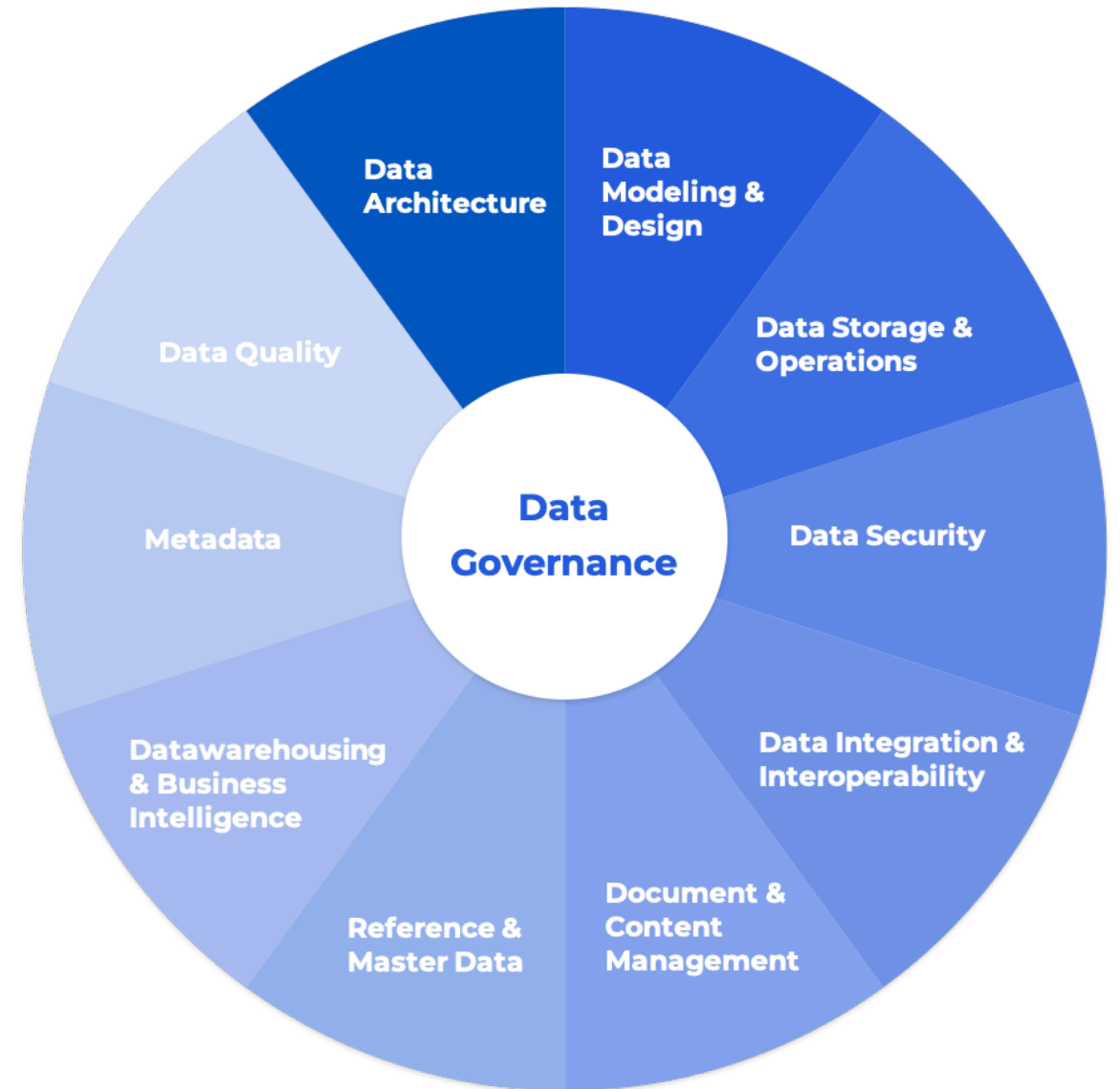


# Waar te beginnen?

Data Management Body of Knowledge (DMBoK) Wheel

- Verschillende focusgebieden voor datamanagement.
- Data governance in het midden.

©ITSV



# Componenten in Data Management Body of Knowledge Wheel

**Data-architectuur:** definieert de structuur van data-activa en beheerprocessen.

**Datamodellering en -ontwerp:** ontdek, analyseer en representeer datavereisten.

**Dataopslag en -bewerking:** ontwerp en ondersteun dataopslag.

# Componenten in Data Management Body of Knowledge Wheel

**Data-integratie en -interoperabiliteit:** beheer dataverplaatsing en -consolidatie.

**Referentie- en masterdata:** onderhoud cruciale gedeelde data.

**Datakwaliteit:** zorg ervoor dat data geschikt is voor gebruik.

# Metadata Management

# Melk = onze data



Hoe kies je de juiste dataset voor uw use case?



Metadata: een contextuele laag die het de gebruiker eenvoudiger maakt om de data beter te begrijpen.

Nutrition Facts		
SERVING = A STANDARD NUTRITION LABEL		
Serving 8		
Serving Size		
Serving Size	Nutrition Facts Label	
Calories	Calories	Serving Sizes
Serving Size   Serving Fat	Calories	12.3% 1.5%
Calories	Total Fat	1.5% 5.5%
Total Fat, Cholesterol	Total Fat	5.5% 1.5%
Total Fat	Sodiums	1.5% 1.2%
Total Cholesterol,	Total Carbohydrates	39 1.5%
Total Carbohydrates	Sugars	2.3% 1.2%
Total Carbohydrates	Total Carbohydrates	12.9% 1.5%
Total Protein	Vitamins	5.5% 1.5%
No Vitamins	Pratamins	1.2% 1.5%
Vitamins	Vitamins	5.5% 1.5%

## Types metadata:

- Bedrijfsmetadata
  - Voorbeeld: de definitie van een 'klant' in een CRM-systeem.
- Technische metadata
  - Voorbeeld: database schema's (VARCHAR, INT, ...)
- Structurele metadata
  - Voorbeeld: een XML-schema dat de structuur van het bestand beschrijft.



- Operationele metadata
  - Voorbeeld: Logging-gegevens van een data-extractie proces.
- Governance-metadata
  - Voorbeeld: Een classificatieschema dat aangeeft dat een verzekeringsnummer worden geclassificeerd volgens de GDPR-wetgeving.

# Data Governance

Het kader van processen, regels en verantwoordelijkheden dat ervoor zorgt dat data binnen een organisatie accuraat, veilig, toegankelijk en consistent wordt beheerd en gebruikt.

## **Data governance creëert een gemeenschappelijke taal.**

Iedereen binnen de organisatie gebruikt dezelfde termen en definities voor data.

Teams, ongeacht hun afdeling, begrijpen dezelfde gegevens op dezelfde manier. Dit draagt bij aan consistente besluitvorming.

**Data governance brengt mensen samen om samen te werken.**

Duidelijk vastgelegde rollen en verantwoordelijkheden voor data-eigenaren die toezicht houden op de datakwaliteit en naleving.

## **Data governance vermindert beveligingsrisico's.**

Strengere toegangscontroles en toezicht op datagebruik helpen bij het minimaliseren van beveligingsrisico's.

**Data governance maakt data betekenisvol.**

Data wordt bijgevolg verrijkt met de juiste context, zodat het effectief gebruikt kan worden binnen de bedrijfsprocessen.

# Een voorbeeld van data governance

ING Bank implementeert data governance voor GDPR naleving.

## Wat is de uitdaging?

ING werd geconfronteerd met de uitdaging om te voldoen aan de strenge vereisten van de General Data Protection Regulation (GDPR). De bank had grote hoeveelheden gegevens verspreid over meerdere afdelingen en systemen, en er was een gebrek aan centraal beheer en controle over de datastromen. Dit verhoogde het risico op inbreuken op dataprivacy en maakte het moeilijk om te voldoen aan de nieuwe regelgeving.



# Wat is de oplossing?

ING implementeerde een uitgebreide strategie met een focus op het verbeteren van de naleving van de privacyregels.

- Data Catalogus: een centrale data catalogus die alle gegevensbronnen binnen de organisatie in kaart brengt.
- Data Stewardship: de bank wees data stewards aan voor elk datadomein. Deze stewards zijn verantwoordelijk voor de datakwaliteit, governance en naleving van de regelgeving.
- Geautomatiseerde monitoring: geautomatiseerde tools voor datakwaliteitscontrole en het monitoren van datagebruik.

# Data management in de praktijk

- 66% van de bedrijven verspilt tijd met het stellen van dezelfde vragen over data.
- 61% van de dataconsumenten heeft geen inzicht in wie welke data gebruikt.
- Slechts 25% van de bedrijven neemt datagedreven beslissingen.

# Data Management Strategy

- **Defensief gebruik van data:** richt zich op data kwaliteit, beveiliging, privacy en naleving van regelgeving.
- **Offensief gebruik van data:** richt zich op monetisatie, kostenreductie en operationele efficiëntie.

# Master Data Management (MDM)

Master data zijn de kerngegevens die essentieel zijn voor bedrijfsvoering.

Deze gegevens worden door meerdere afdelingen binnen een organisatie gebruikt, waardoor consistentie en nauwkeurigheid cruciaal zijn.

Voorbeelden: productinformatie, leveranciersgegevens.

## Het doel van MDM

Ervoor zorgen dat er één betrouwbare, consistente versie is van de waarheid voor alle kerngegevens in een organisatie.

Dit wordt ook **single source of truth** genoemd.

# Quiz

# Wat is het primaire doel van Master Data Management (MDM) binnen een organisatie?

- A. Het dupliceren van gegevens voor redundantie
- B. Het onderhouden van één enkele bron van waarheid voor kritieke gegevens
- C. Het verhogen van de datavolumes binnen een organisatie
- D. Het ontwikkelen van nieuwe data-opslagmethoden

**Welke van de volgende soorten metadata wordt gebruikt om de structuur van een bestand te beschrijven?**

- A. Bedrijfsmetadata
- B. Operationele metadata
- C. Structurele metadata
- D. Governance-metadata



# Wat is een van de belangrijkste voordelen van Data Governance?

- A. Het verhogen van de hoeveelheid verzamelde data
- B. Het verminderen van beveiligingsrisico's door strikte toegangscontroles
- C. Het versnellen van de datainvoer in systemen
- D. Het verminderen van de kosten voor datakwaliteitscontrole

# Wat was de belangrijkste uitdaging voor ING bij de implementatie van Data Governance volgens de case study?

- A. Het naleven van de vereisten van de General Data Protection Regulation (GDPR)
- B. Het implementeren van nieuwe databasesystemen
- C. Het verhogen van de winst door middel van data-monetisatie
- D. Het automatiseren van klantinteracties met AI