



Presentatie en Discussie KNVI

10 Beginnersfouten in het Data-landschap

Waar Testen en Toetsen het verschil kan maken

Armando Dörsek

22 juni 2021

Introductie en opbouw

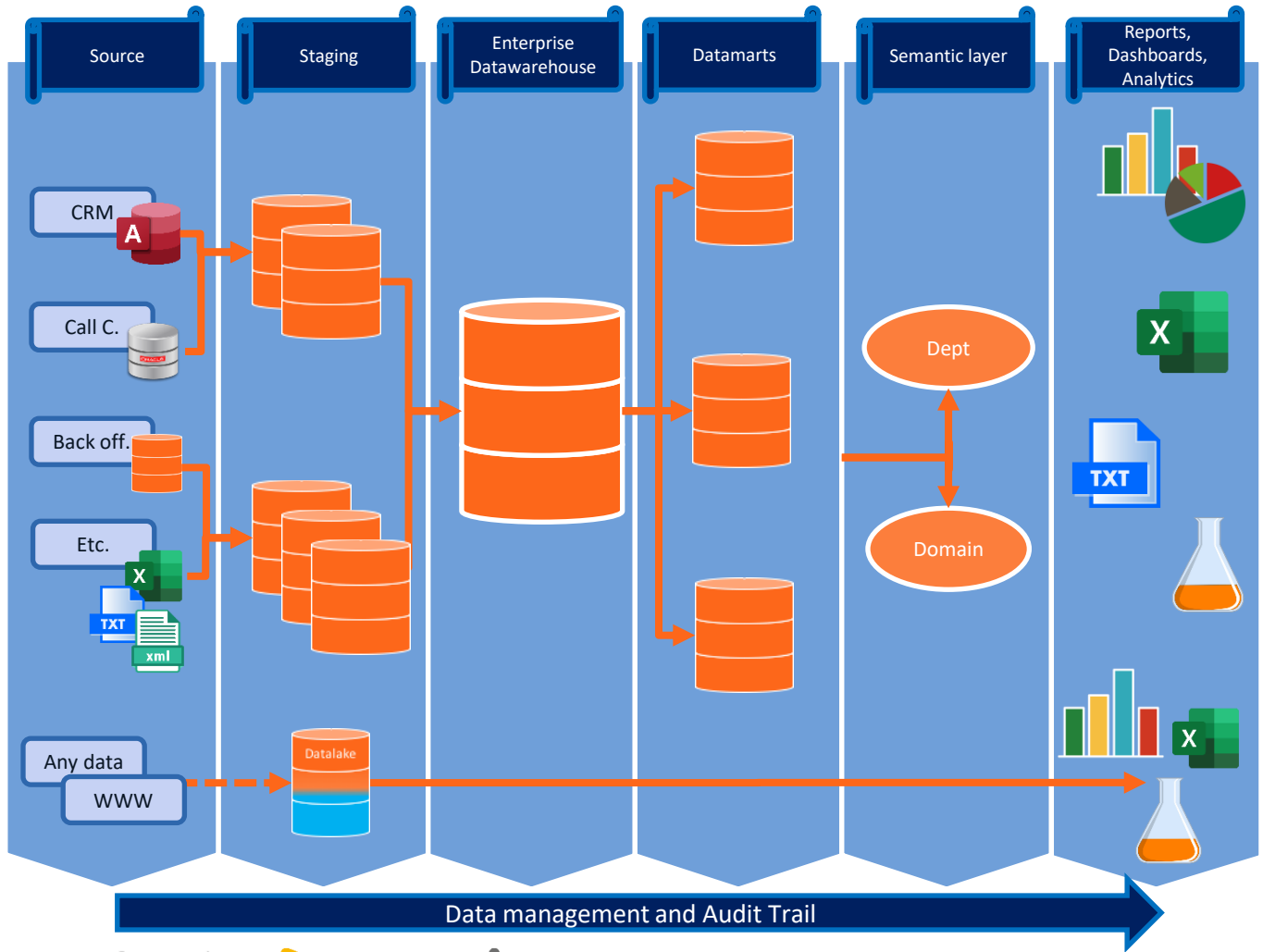
- Intro
 - Over mij
 - Waarom luisteren / kijken?
- Inhoud
 - Mentimeter
 - Deel 1
 - Pauze*
 - Deel 2
 - Samenvatting/Conclusie
- Discussie

Armando Dörsek

- Verified.nl
- NL + AT
- HEAO Arnhem, Radboud Universiteit
- Testen & Toetsen
- BI/DWH & Analytics
- (Test) consultancy, -management en –opleidingen
- Trusted advisor
- Verbindend en hulpvaardig



Een gemiddeld data-landschap...



U mag, gezien de tijd, kiezen uit 5 “beginnersfouten”.
Welke kiest u?



Go to
www.menti.com
and use the code
7880 4645

<https://www.menti.com/wwy6c9acqa>

Beginnersfouten bij Testen in het Data Landschap

“Voor datakwaliteit zijn wij toch zeker niet verantwoordelijk?”

“Garbage in = Garbage out”

Het team geeft aan geen invloed te hebben op de kwaliteit van de data in bronsystemen, daar niet verantwoordelijk voor te zijn.

Mogelijke oorzaken: geen datamanagement, eigenaarschap, rolonduidelijkheid, onwil (extra werk) bij de bron, geen notie van het effect van fouten in data.

“We kunnen Datakwaliteit niet eens meten”

Toch wel: inmiddels zijn diverse DQ methodieken voorhanden, incl. DAMA, ISO 25012, IDQ (IPS) en bedrijfs(tak)eigen kwaliteitsstandaarden. Meten is weten. NB: Als geen data steward of business analist DQ bepaalt, kan een tester (voorlopig) deze werkzaamheden oppakken.

Noot: DQ-issues leveren problemen op voor beheerders, bouwers en testers: uitzonderingen kunnen leiden tot uitval en/of aanpassingen in de code. Dit leidt tot extra bouw- en testwerk. Hebben we tijd en geld voor extra bouw- en testwerk?



Beginnersfouten bij Testen in het Data Landschap

Slepen met Data

Meerdere Waarheden

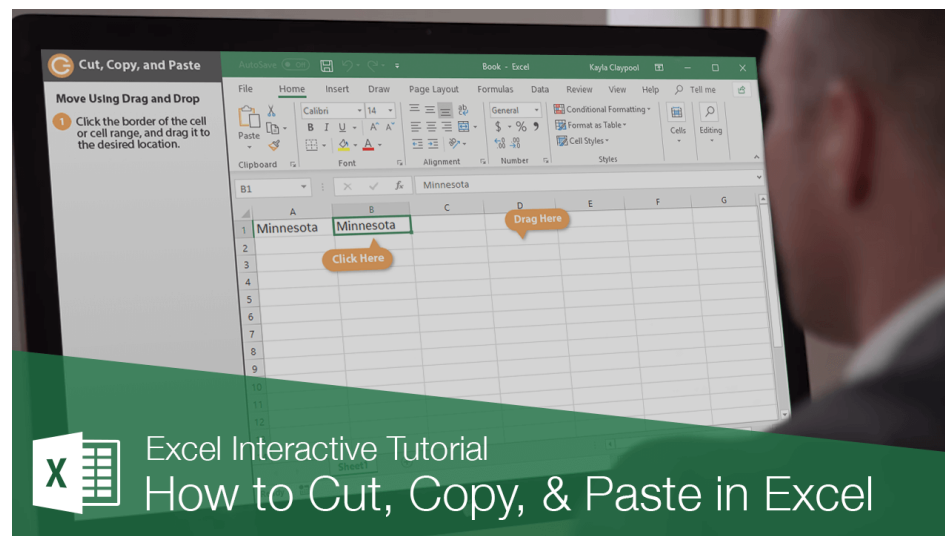
Ongebreideld kopiëren van data door business (en techniek) leidt tot het ontstaan van verschillen en meerdere “waarheden”.

Dit leidt tot méér (te testen) producten, met hun eigen versies van een dataset, meer uitzonderingen en meer zoekwerk. Bugfixing leidt tot onnodige belasting van beheerders en analisten.

Potentiële datalekken

Hoe meer data je verspreidt, hoe meer kans op datalekken.
NB: Wie heeft autorisatie tot welke data?

Noot: ontwerpers, testers en ontwikkelaars kunnen toetsen op het onnodig inzetten van “kopieën van data”. Middels Testdata Management kan iedereen voorzien worden van geverifieerde testsets.



Pauze



Beginnersfouten bij Testen in het Data Landschap

“Ja maar... we hebben toch **Productiedata** om mee te testen?”

Meer is niet altijd beter...

- Volume is niet zaligmakend: bevat vooral véél dezelfde cases, niet de juiste (test) cases – en vergt wel heel veel doorlooptijd
- Wat er vandaag nog niet in zit, kan morgen wel aangeboden worden...
- Indicatief voor Performance, maar hoe zit het met zaken zoals toekomstige groei en laadprocessen?
- Een afslag van productie veroudert, een live connectie met productie is mogelijk te vluchtig voor testers.
- Is de huidige productie-bron de goede versie om tegen te testen? Of hebben we een nieuwere versie nodig?

Tip: Maak duidelijk onderscheid in functioneel testen en non-functionals incl. doorlooptijden. Met testontwerp-technieken kun je gericht testen (zoals Klassen met EP, Grenzen met BVA, als-dan-anders met bv DTT). Neem Testdata Management op als rol.

Onderschatting van Kaders o.g.v. Privacy en Security

- AVG vraagt om het anonimiseren van Data; tijdig starten, in een datawarehouse vaak complex.
- In een datawarehouse kunnen verbanden worden gelegd die in losse bronsystemen niet aanwezig zijn, waardoor latente PII in PII verandert.
- “Testdata” naar buitenlandse leveranciers brengen kan een issue vormen (contractueel af te hechten).
- Productiedata in een Testomgeving? Dan ook securitymaatregelen zoals in Productie.

Noot: Op enig moment zullen we met Productie(like) data willen testen bv. voor een ketentest/end-to-end test, zorg op dat moment ook voor de juiste beveiligingsmaatregelen incl. autorisatiestructuren, bv. op de acceptatie-omgeving.

Beginnersfouten bij Testen in het Data Landschap

Eén klant is geen klant

Wil de echte klant opstaan?

Heeft de klant meerdere organisatieonderdelen of regio's die bediend worden? Zorg dan voor een representatieve vertegenwoordiging bij de Acceptatietesten.

Noot:

Lokale, afwijkende werkprocessen kunnen leiden tot andere (onverwachte) data. Als dit te laat onderkend wordt, zal dit leiden tot discussies, "niet-accepteren" en uitloop van het project.



Discussie



Dank u wel

Vragen? Neem gerust contact op met:

- Armando Dörsek
- (06) 23 054 054
- armando.dorsek@verified.nl



(c) 2021 Verified.nl en KNVI - Hergebruik toegestaan met toestemming van de auteurs