



Releasenieuws

Datadash 1.8

Elke dag speelt software een rol in de ondersteuning van uw bedrijfskritische processen. Belangrijke applicaties in uw organisatie staan vol met privacygevoelige persoonsgegevens van uw medewerkers, klanten of patiënten. En elke dag komen daar weer nieuwe gegevens bij.

Als u deze applicaties wil testen of de data uit de applicaties wil gebruiken om medewerkers te opleiden of onderzoek te doen, dan doet u dit het liefst met waarheidsgetrouwe gegevens. Maar dit mag niet volgens de Nederlandse en Europese privacywet.

Door de privacygevoelige gegevens te anonimiseren, zijn de gegevens niet meer te herleiden naar een individueel persoon. Zo voldoet u aan de wet- en regelgeving en heeft u een kwalitatief goede da-

taset om te gebruiken voor testen, onderzoek en ontwikkeling. Data volledig anonimiseren kan een flinke klus zijn en daar kan de anonimiseeroplossing Datadash u bij helpen.

Datadash 1.8

In Datadash 1.8 zijn een aantal nieuwe functionaliteiten beschikbaar voor het anonimiseren van persoonsgegevens.

- Seed-functionaliteit
- Remove outliers

Daarnaast controleren wij vanaf Datadash 1.8 dagelijks de kwaliteit en de snelheid van de algoritmes die gebruikt worden in de shufflemethode.



1 Seed-functionaliteit

Bij het testen van software kun je als gebruiker meerdere keren dezelfde test set willen gebruiken. Met de seed-functionaliteit is het mogelijk om de randomisatie van data te reproduceren. Dat betekent dat wanneer u dezelfde dataset randomiseert, u elke keer hetzelfde resultaat krijgt. Op de manier kunt u uw testcases hergebruiken. Ook kan de seed-functionaliteit van toegevoegde waarde zijn in een ketentest. Zodat u, indien nodig, maar een gedeelte uit de ketentest hoeft te anonimiseren.

Voorbeeld

We geven aan dat de kolommen 'voornaam' en 'achternaam' geanonimiseerd moeten worden door de shuffle-anonimiseermethode met de seed-functionaliteit. Deze shufflemethode is gebaseerd op een zelf ontwikkeld algoritme, dat gebruikt maakt van randomisatie om de data te anonimiseren. Wanneer we de originele tabel voor een tweede keer anonimiseren met hetzelfde commando ziet u in het voorbeeld hiernaast dat het resultaat identiek is aan de eerste keer.

Omdat de randomisatie in het geheugen van de database wordt uitgevoerd en er geen sleutel is, is hier alsnog sprake van anonimisering.

Data voor anonimiseren

	Patientnummer	Voornaam	Achternaam
1	76020088408	Simonas	Burch
2	79954552053	Witold	Limbek
3	74655241640	Theijs	Delisle
4	73258162498	Oskar	Mora
5	72231460069	Myllo	Gjertsen
6	75271692533	Long	Bild
7	77035094215	Cavian	Limoges
8	71574471824	Jawhar	Lefrançois
9	74911062869	Tygo	Rank

Data na eerste anonimisering

	Patientnummer	Voornaam	Achternaam
1	76020088408	Massoud	Fensmark
2	79954552053	Iero	of Metz
3	74655241640	Iliaz	Rehbinder
4	73258162498	Sukur	Novgorodskij
5	72231460069	Thorvild	Røssing
6	75271692533	Vigo	Armour
7	77035094215	Jagger	von Eickstedt
8	71574471824	Arbi	Gill
9	74911062869	Youns	Tang

Data na tweede anonimisering

	Patientnummer	Voornaam	Achternaam
1	76020088408	Massoud	Fensmark
2	79954552053	Iero	of Metz
3	74655241640	Iliaz	Rehbinder
4	73258162498	Sukur	Novgorodskij
5	72231460069	Thorvild	Røssing
6	75271692533	Vigo	Armour
7	77035094215	Jagger	von Eickstedt
8	71574471824	Arbi	Gill
9	74911062869	Youns	Tang



2 Remove outliers

Drie belangrijke elementen bij het anonimiseren van data zijn de herleidbaarheid, deduceerbaarheid en de koppelbaarheid van data. Voor volledige anonimisering moeten aan die drie elementen voldaan worden. De optie 'remove outliers' helpt om beter te kunnen voldoen aan het principe van deduceerbaarheid. Met deze optie kunt u als gebruiker aangeven hoe vaak een waarde moet voorkomen om meegenomen te worden in de shufflemethode.

Voorbeeld

Sommige persoonlijke kenmerken zijn zo specifiek,

dat u niet wilt dat deze terugkomen in het anonimiseerresultaat omdat deze dan nog steeds herkenbaar zijn. Bijvoorbeeld een achternaam die erg weinig voorkomt, zoals 'Van Oranje-Nassau'. Als u een tabel waarin de achternaam 'Van Oranje-Nassau' één keer voorkomt en de optie 'remove outliers' niet gebruikt, komt de achternaam mogelijk terug in het resultaat na anonimisering. Door de optie aan te zetten en de 'outliers threshold' op 1 te zetten, komt de achternaam dan niet meer terug in het resultaat na anonimiseren.

Datadash Configuration Tool

File Edit Help

Database Connection
 Connection:

Target Database
 Target Database:

Anonymize Configuration

Table Selection
 Table: ☐ Hide non-configured tables

Column Configuration

	Column Name	Column Type	Anonymize Method	Replace Value	Group Name
▶	PatientNr	varchar			
	Voornaam	varchar	SHUFFLE		
	Achternaam	varchar	SHUFFLE		
	GebDatum	datetime	SMARTREPLACE		
	Geslacht	varchar			
	Land	varchar	SHUFFLE		
	Plaats	varchar	GROUPEDSHUF...		PC_Plaats
	Adres	varchar	SHUFFLE		
	Postcode	varchar	GROUPEDSHUF...		PC_Plaats
	HuisNr	int	SHUFFLE		
	Tel1	varchar	REMOVE		
	Tel2	varchar			

Anonymize Options

REPLACE SMARTREPLACE HASH SHUFFLE GROUP

Keep NULLs ☐

Force Unique ☐

Remove Outliers ☐

Outliers Threshold



3 Kwaliteit en snelheid shuffle-algoritme

Om ervoor te zorgen dat u profiteert van optimaal presterende software en u gebruikt kunt maken van onherleidbare data testen wij de algoritmes, die onder andere gebruikt worden in de shufflemethode, vanaf Datadash 1.8 op de volgende onderdelen:

- Codekwaliteit van Datadash
- Kwaliteit van de installatie scripts
- Kwaliteit van de randomisatie in de shufflemethode
- Snelheid van Datadash
- Inhoudelijke uitvoering van de anonimiseermethodes

Whitepaper Datadash

Meer weten over de shuffle-anonimiseermethode en de onderliggende algoritmes? In het whitepaper: '*Datadash deep dive - de shuffle anonimiseermethode*' geven we uitgebreid inzicht in hoe de anonimiseermethode tot stand is gekomen.

[Download whitepaper](#)

Meer weten over Datadash?

In dit releasenieuws las u meer over de belangrijkste vernieuwingen in Datadash 1.8.

Datadash 1.9 verwachten wij in januari 2018 op te leveren. In deze nieuwe versie komen we met nieuwe functionaliteiten waarmee u als gebruiker de relaties tussen verschillende databasetabellen in stand kunt houden.

Heeft u vragen naar aanleiding van dit document? Of wilt u meer informatie over een van deze onderwerpen of de implementatie van Datadash? Neem contact op met Mark Jansen.



Mark Jansen

Productmanager Datadash
mark.jansen@pinkroccade.nl
06 – 4339 2983