

# Metadata and data management

## Data Service Centre (DSC)

Harold Kroeze

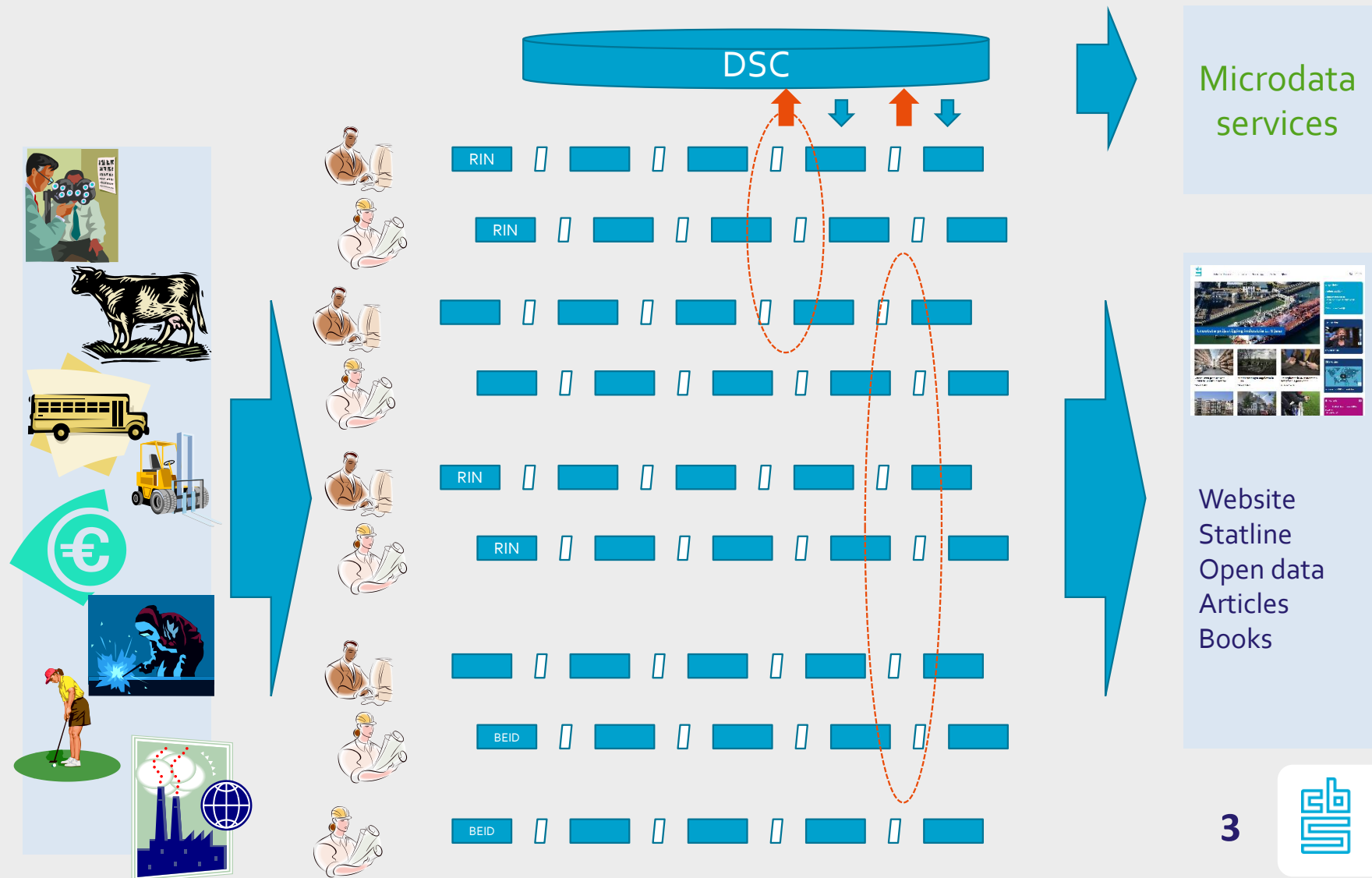


Centraal Bureau  
voor de Statistiek

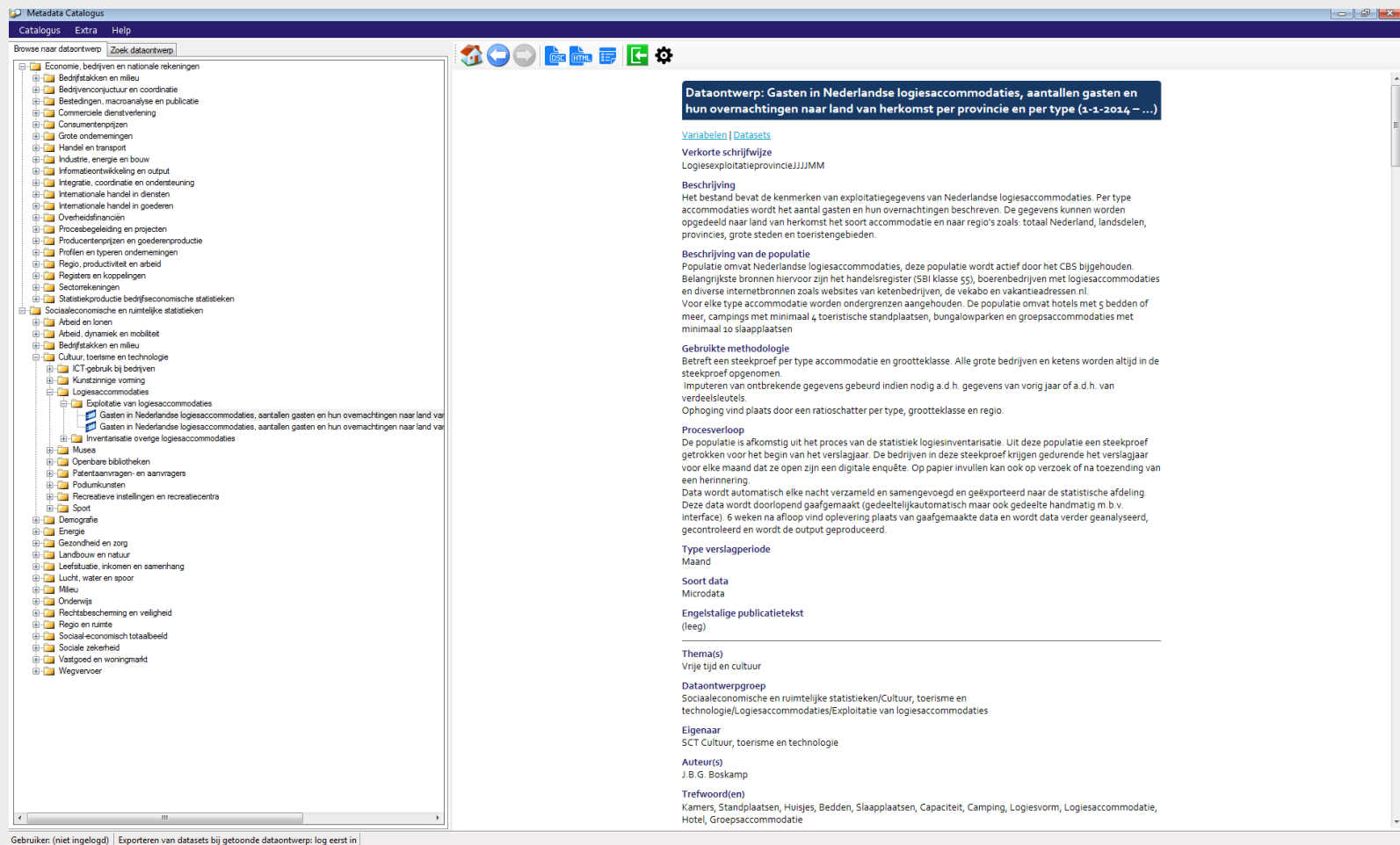
# Data Service Centre (DSC)

Corporate repository for storing and sharing statistical data and metadata

# Statistical process and DSC



# DSC Catalogue



**Metadata Catalogus**  
Catalogus Extra Help

Browse naar dataontwerp Zoek dataontwerp

**Dataset:** Gasten in Nederlandse logiesaccommodaties, aantallen gasten en hun overnachtingen naar land van herkomst per provincie en per type (1-1-2014 - ...)

**Variabelen | Datasets**

**Verkorte beschrijving**  
LogiesexploitatieprovincieJJJMM

**Beschrijving**  
Het bestand bevat de kenmerken van exploitatiegegevens van Nederlandse logiesaccommodaties. Per type accommodaties wordt het aantal gasten en hun overnachtingen beschreven. De gegevens kunnen worden opgedeeld naar land van herkomst het soort accommodatie en naar regio's zoals: totaal Nederland, landsdelen, provincies, grote steden en toeristengebieden.

**Beschrijving van de populatie**  
Populatie omvat Nederlandse logiesaccommodaties, deze populatie wordt actief door het CBS bijgehouden. Belangrijkste bronnen hiervoor zijn het handelsregister (SBI klasse 55), boerenbedrijven met logiesaccommodaties en diverse internetbronnen zoals websites van ketenbedrijven, de vekabo en vakantieadressen.nl. Voor elke type accommodatie worden ondergrenzen aangehouden. De populatie omvat hotels met 5 bedden of meer, campings met minimaal 4 toeristische standplaatsen, bungalowparken en groepsaccommodaties met minimaal 10 slaapplekken.

**Gebruikte methodologie**  
Betreft een steekproef per type accommodatie en grootteklasse. Alle grote bedrijven en ketens worden altijd in de steekproef opgenomen. Imputeren van ontbrekende gegevens gebeurt indien nodig a.d.h. gegevens van vorig jaar of a.d.h. van verdeelsleutels. Ophoging vindt plaats door een ratioschatter per type, grootteklasse en regio.

**Procesverloop**  
De populatie is afkomstig uit het proces van de statistiek logiesinventarisatie. Uit deze populatie een steekproef getrokken voor het begin van het verslagjaar. De bedrijven in deze steekproef krijgen gedurende het verslagjaar voor elke maand dat ze open zijn een digitale enquête. Op papier invullen kan ook op verzoek of na toezending van een herinnering. Deze data wordt automatisch elke nacht verzameld en samengevoegd en geëxporteerd naar de statistische afdeling. Deze data wordt doorlopend gaafgemaakt (gedeeltelijk automatisch maar ook gedeeltelijk handmatig m.b.v. interface). 6 weken na afloop vindt oplevering plaats van gaafgemaakte data en wordt data verder geanalyseerd, gecontroleerd en wordt de output geproduceerd.

**Type verslagperiode**  
Maand

**Soort data**  
Microdata

**Engelstalige publicatietekst**  
(leeg)

**Thema(s)**  
Vrije tijd en cultuur

**Dataontwerpgroep**  
Sociaaleconomische en ruimtelijke statistieken/Cultuur, toerisme en technologie/Logiesaccommodaties/Exploitatie van logiesaccommodaties

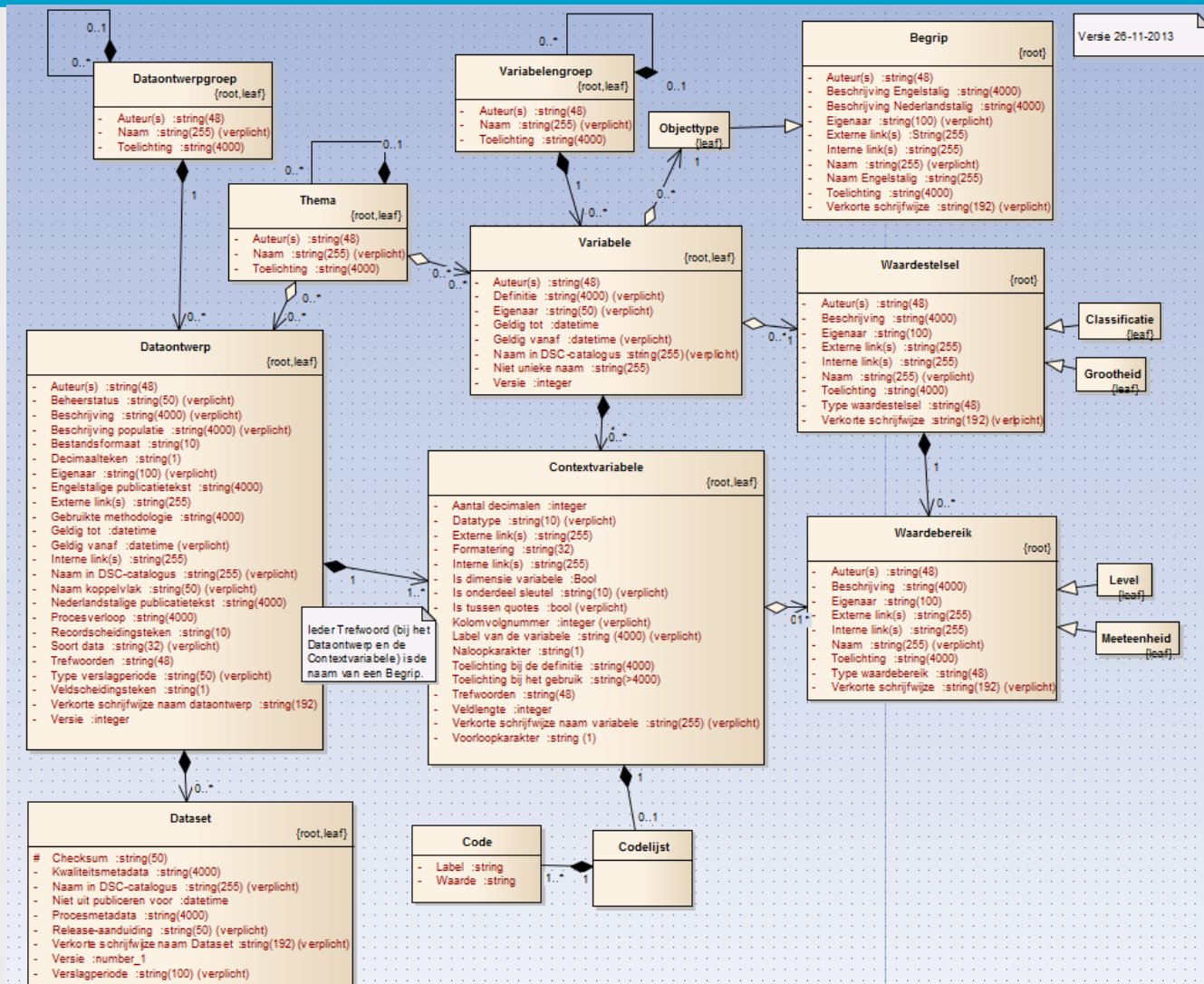
**Eigenaar**  
SCT Cultuur, toerisme en technologie

**Auteur(s)**  
J.B.G. Boskamp

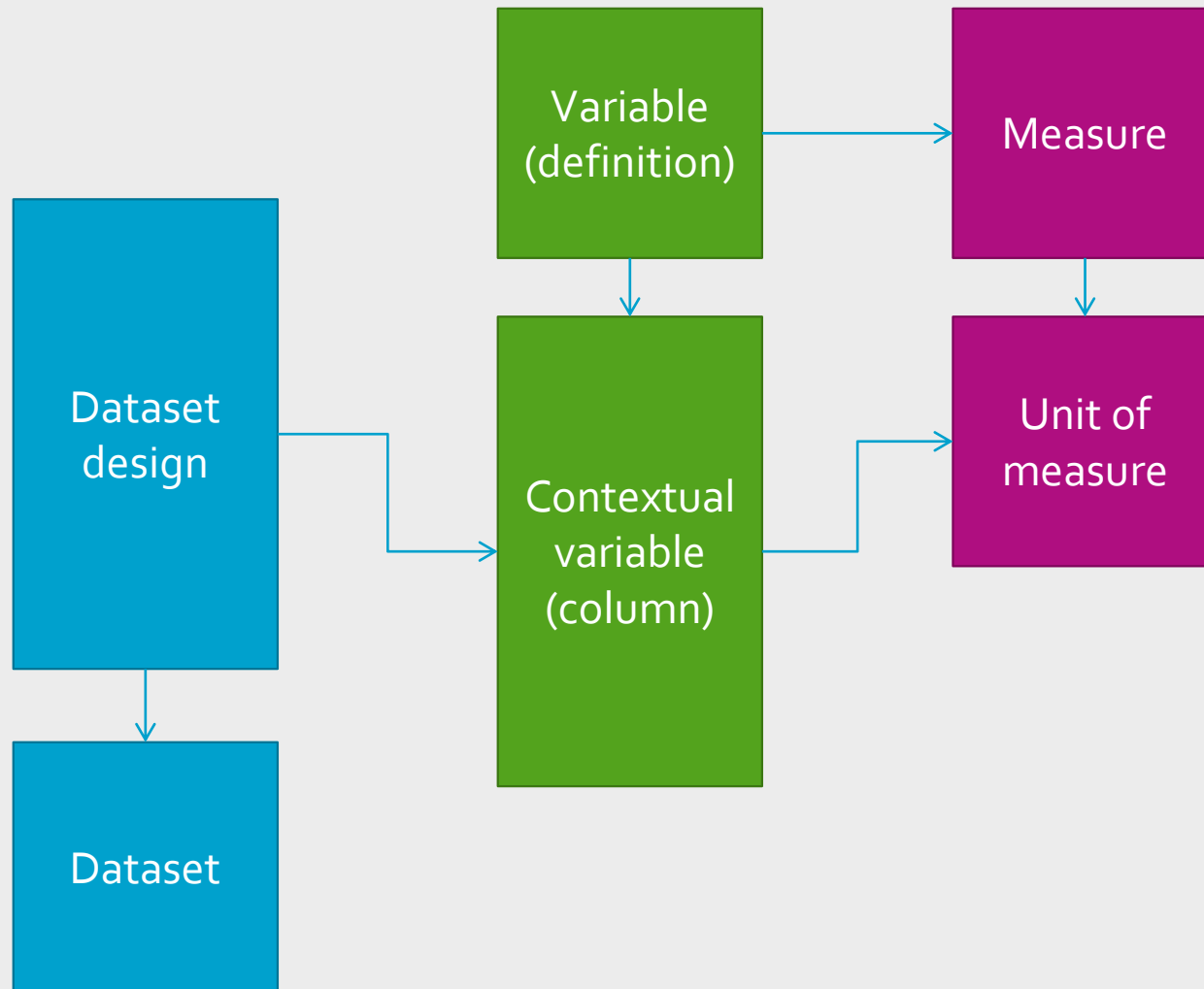
**Trefwoord(en)**  
Kamers, Standplaatsen, Huisjes, Bedden, Slaapplekken, Capaciteit, Camping, Logiesvorm, Logiesaccommodatie, Hotel, Groepsaccommodatie

Gebbruiker: (niet ingelogd) | Exporteren van datasets bij getoonde dataontwerp: log eerst in

# Metadata model



# Metadata model (overview)



# Metadata guidelines

- Document with guidelines, 2-sheet checklist, built-in checks in DSC Metadata Editor
- Support from DSC
- Examples:
  - The name of a data design must contain the most important variables and a description of the population (*Personal characteristics of all persons entered in the Municipal Personal Records Database*)
  - The name of a variable must contain the measure, if applicable (*Number of persons in household, Volume of freight*)
  - When a contextual variable uses a codelist, this codelist should be present in DSC

# DSC workflow

- Teams identify **steady state files**
- Create **metadata** (using DSC Metadata Editor)
- DSC **checks** metadata quality and uploads to DSC system
- Data owners (teams) **authorise** users (DSC Authorisation tool)
- Teams create csv / fixed-width files and **import** them into DSC system (DSC Import tool)
- Users **search** and browse DSC Catalogue
- When authorised, users can **download** datasets using DSC Catalogue or command-line tools





# DSC in numbers (2016)

**2 386** new datasets were added to DSC in 2016

By the end of 2016 DSC contained **9 115** datasets

Covering **10 461** variables

More than **600** CBS employees had access to the DSC  
Catalogue

Together they downloaded more than **40 000** datasets

# Current developments

- Implement **REST-services**, easier access to data and metadata for tools like R and Python
- Improve **search** and **retrieval**
- DSC Catalogue as intranet **web app**