Metadata and data management Data Service Centre (DSC)

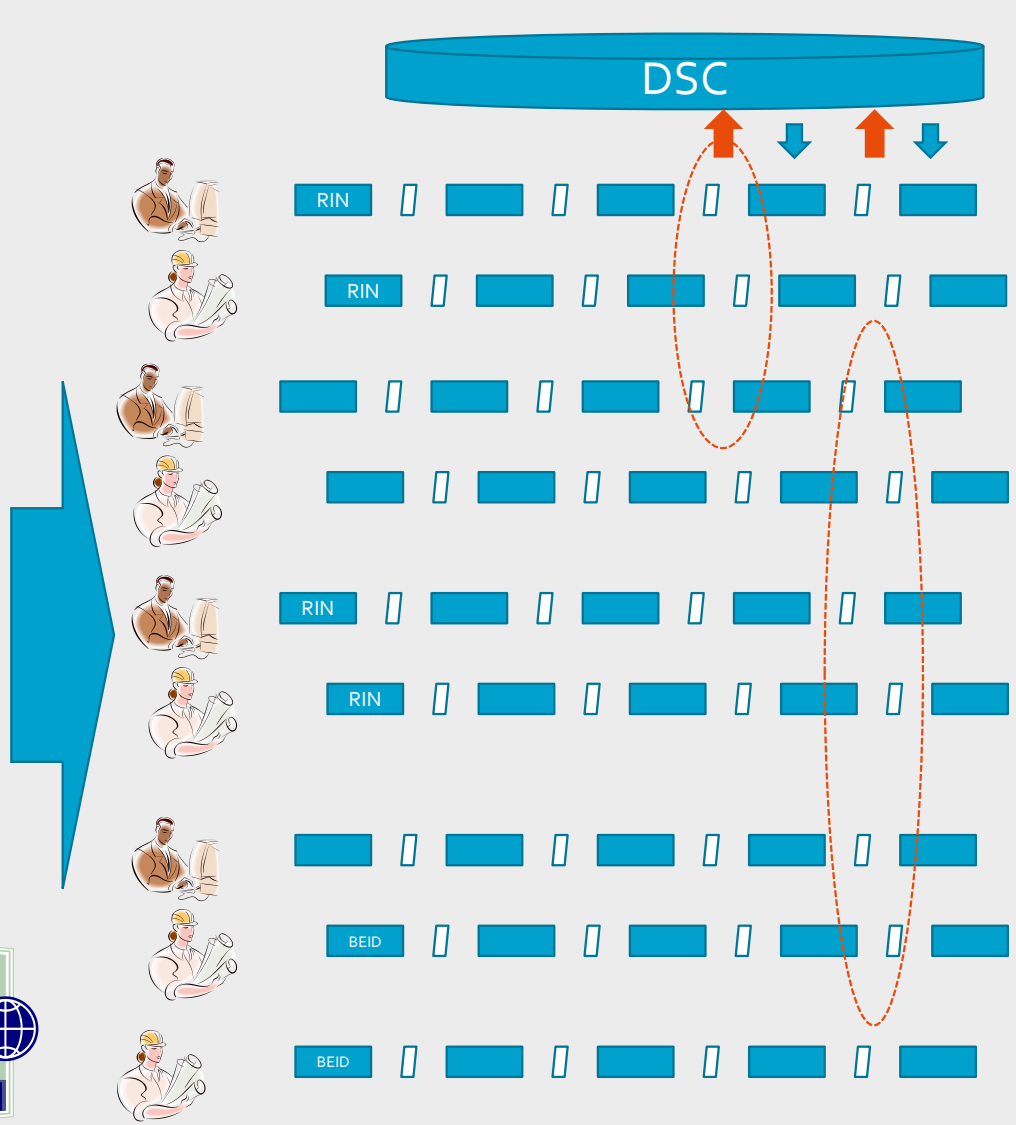
Harold Kroeze

Centraal Bureau voor de Statistiek

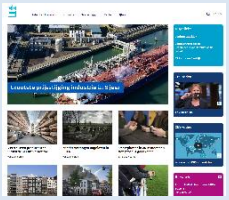
Data Service Centre (DSC)

Corporate repository for storing and sharing statistical data and metadata

Statistical process and DSC

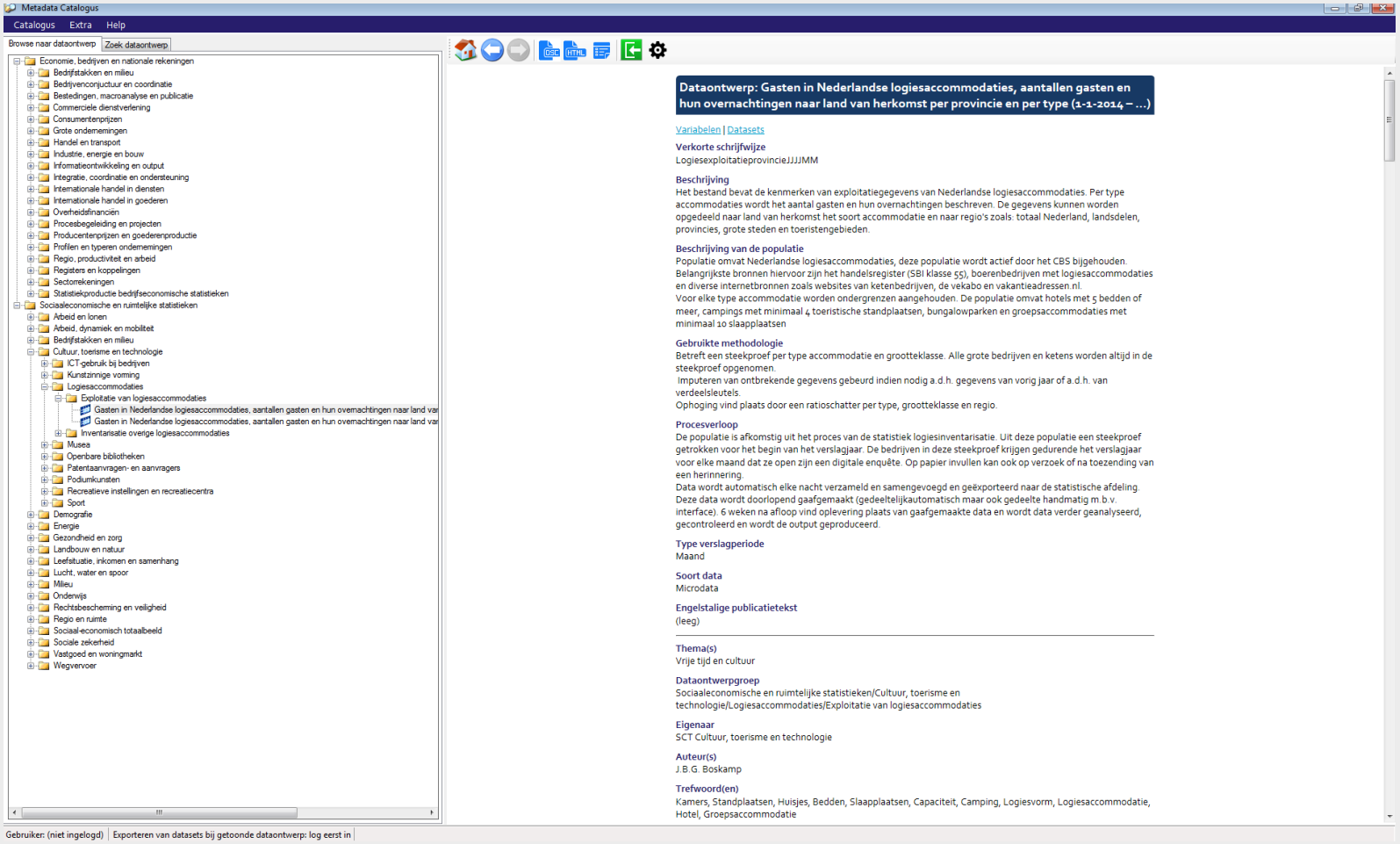


Microdata services

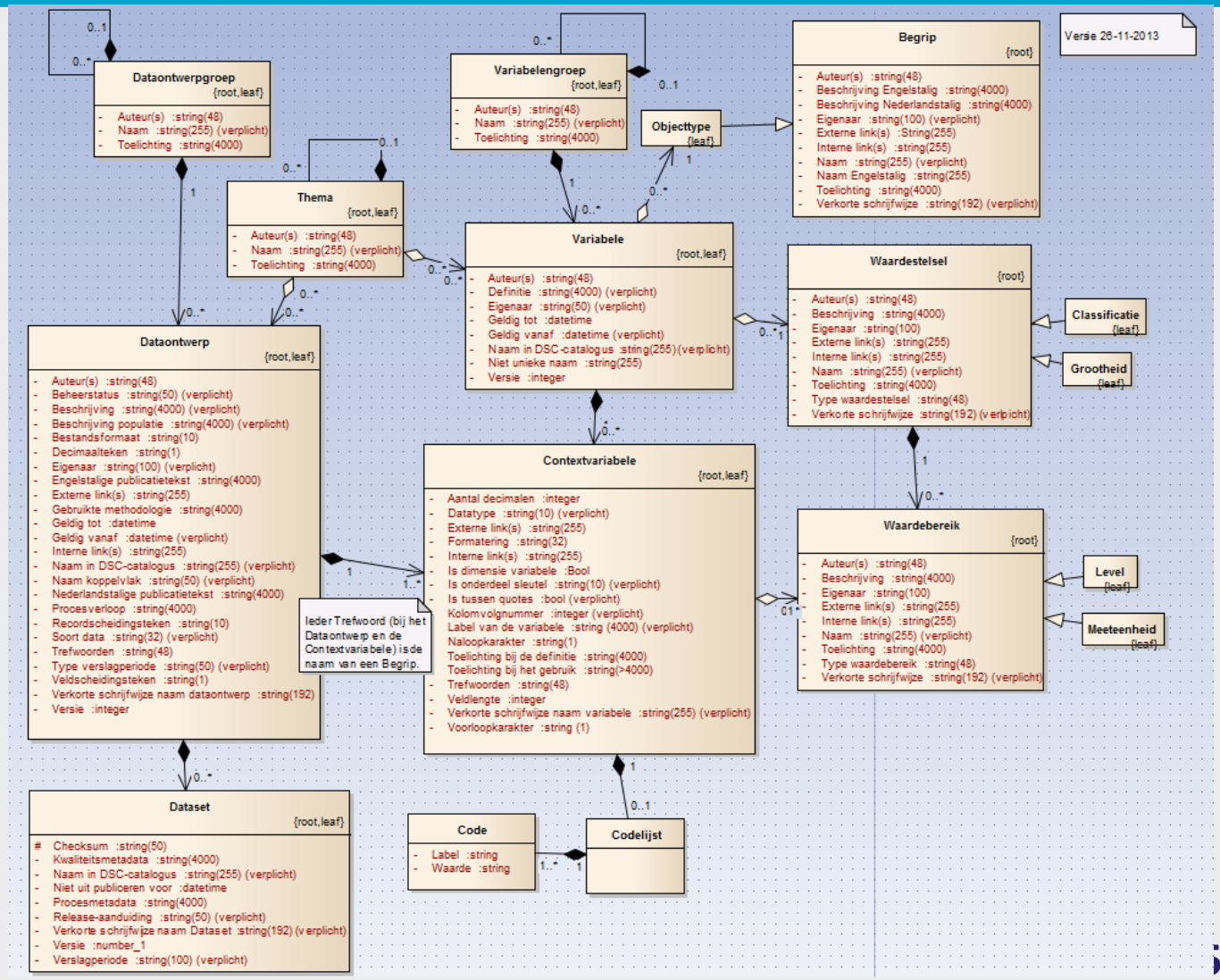


Website
Statline
Open data
Articles
Books

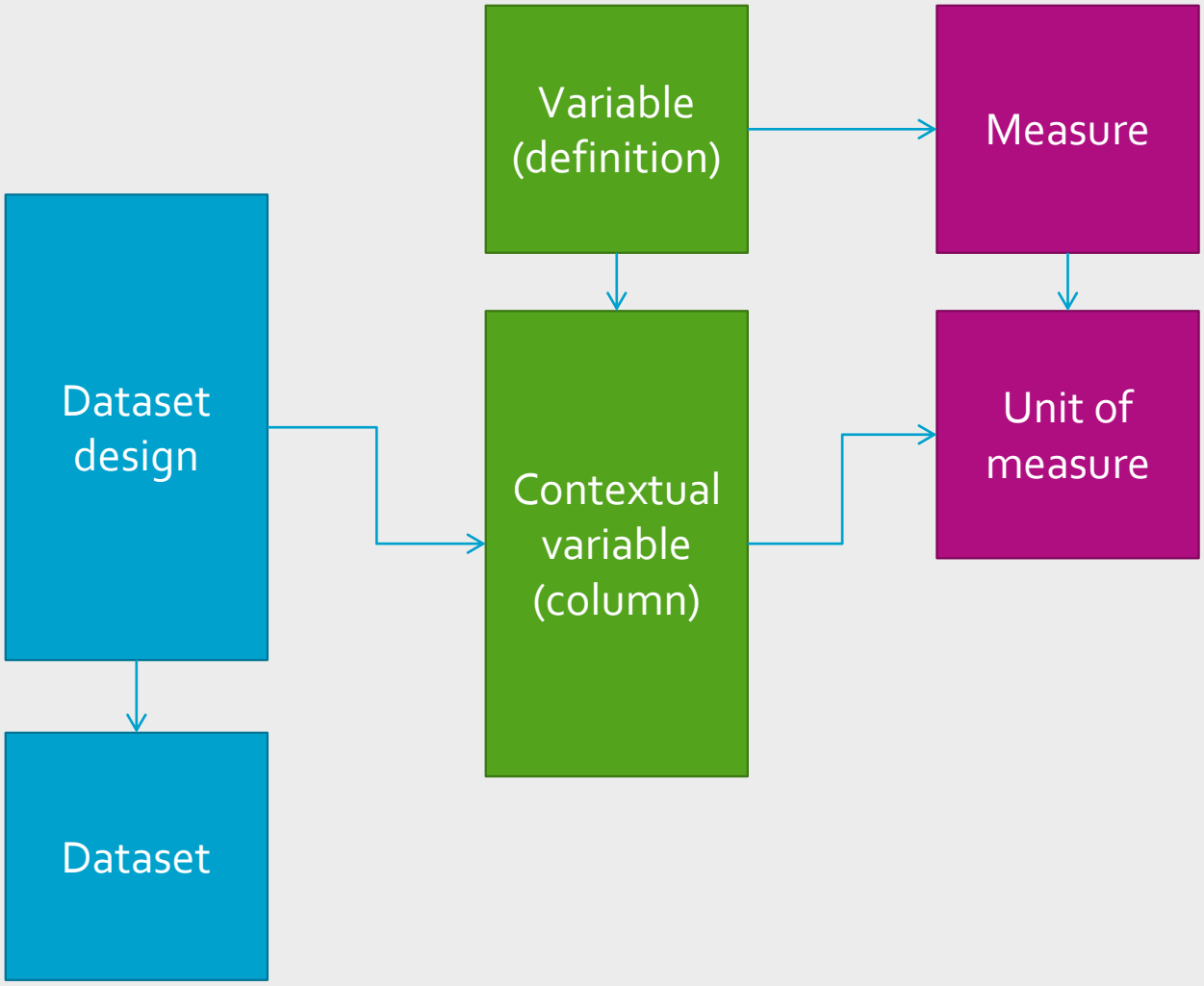
DSC Catalogue



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**