Going from Disaggregated to Aggregated Data

28.01.2021

Presented by Anastassia Samsonova

10th Expert Group Meeting on Statistical Data and Metadata Exchange

Disaggregated data

- Disaggregated data is broken down by more granular categories
  - Local level vs Regional level vs Country level
  - Information at the level of individual respondents, e.g. microdata
- Disaggregated data amounts may be huge
  - OECD's Creditor Reporting System dataset has approximately 50 fields
  - Data fields have different data types (date, string, number, Boolean etc.)
- When available, disaggregated data gives more opportunities for analysis and research

Aggregated data

- Aggregated data is a summary of the disaggregated data that is grouped into more general categories
  - Individual records are aggregated by geographic areas, age groups, sex, education level, level of income etc.
  - Total counts, sums, averages, percentages etc are calculated
- A “drill-down” feature from the aggregated data view may be needed to display the disaggregated data in a separate data browser
  - By default, an aggregated data view is displayed, and disaggregated data is hidden
  - User can click on an observation value to view disaggregated data related to the selected value

Modelling disaggregated and aggregated data in SDMX 2.1

- DSDs may be designed to store both aggregated and disaggregated data
- This requires adding dimensions and attributes to accommodate unique records of:
  - Individual microdata records or disaggregated data that is has more granular breakdowns
  - Aggregated data where some dimensions and attributes are fixed to “Total”, “Not applicable” or have more generalized breakdowns
- A special dimension that distinguishes aggregated data from disaggregated data may be added with the following enumeration:
    - Total (_T): Aggregated data
    - Disaggregated data (DD): Disaggregated data

Disaggregated data display

- Display of aggregated and disaggregated data may be driven by two types of standard SDMX annotations:
  - **DRILLDOWN**

    Denotes the concept (dimension) that specifies whether observations are components of an aggregate value or not (e.g., DD_DIM).

    Is attached to Dataflow, Data Structure Definition or Concept in a Concept scheme.

```
▼<common:Annotation>
   <common:AnnotationTitle>DD_DIM</common:AnnotationTitle>
   <common:AnnotationType>DRILLDOWN</common:AnnotationType>
 </common:Annotation>
```

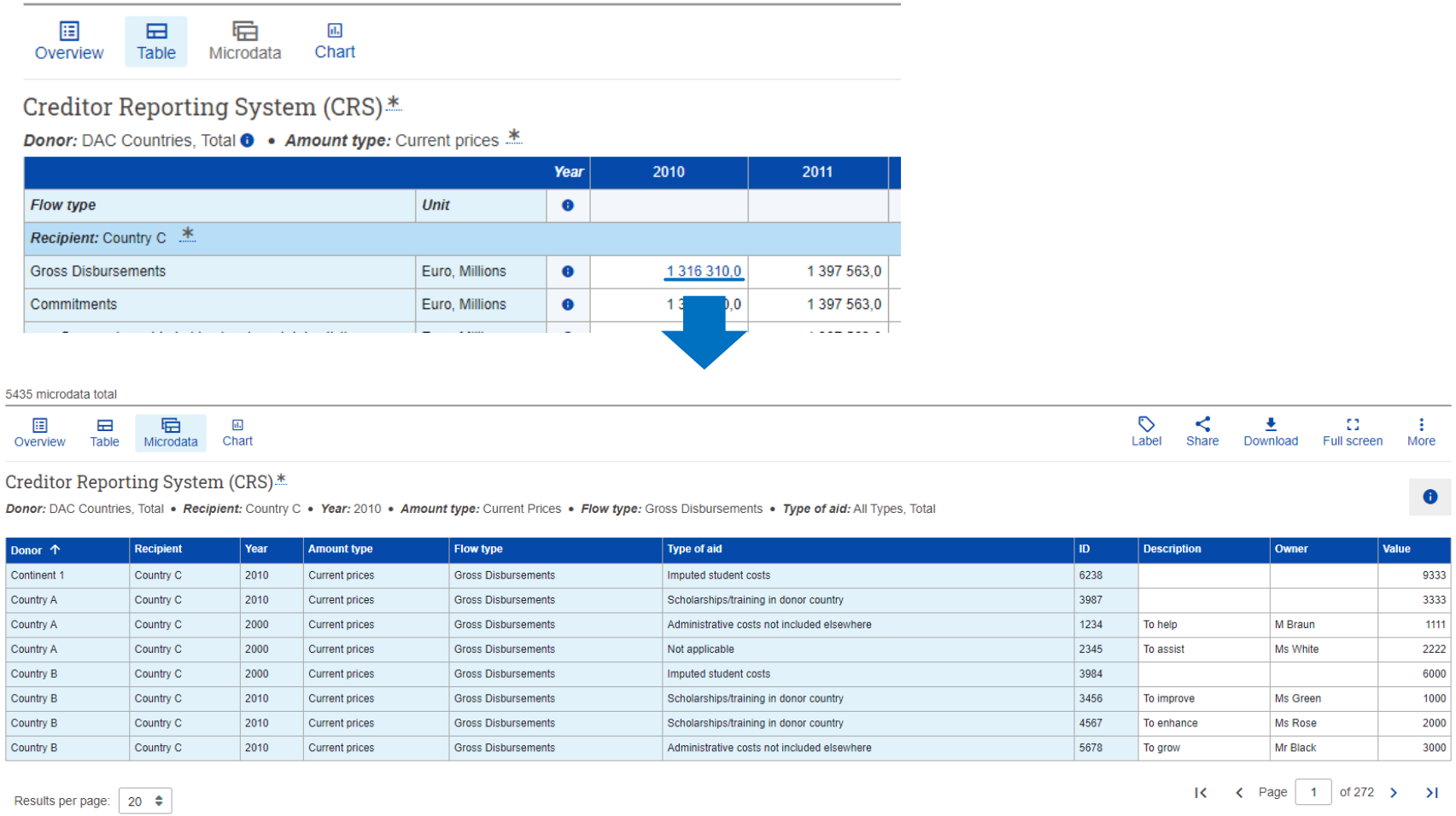
  - **DRILLDOWN_CONCEPTS**

    Defines what dimensions and attributes will be displayed in the drilldown view.

    Is attached to Dataflow or Data Structure Definition.

```
▼<common:Annotation>
   <common:AnnotationTitle>DD_ID,MEASURE,UNIT_MEASURE,REF_AREA,SEX,AGE,DATE,RESULT,RESULT_TIME,INSERT_TIME,OBS_VALUE</common:AnnotationTitle>
   <common:AnnotationType>DRILLDOWN_CONCEPTS</common:AnnotationType>
 </common:Annotation>
```

Aggregated and disaggregated data display in a microdata viewer



* This is an example mock-up of the microdata (drilldown) feature in .Stat Suite. Actual implementation may slightly differ from this mock-up.

Thank you!

OECD
BETTER POLICIES FOR BETTER LIVES