Stat Suite data modelling and migration approach at OECD

02.07.2020
Context

- OECD Statistics production is decentralised (15 directorates, 90-150 teams)
- A working .Stat v7 system for dissemination - but many systems that feed into it.
- With circa 700 live/updated datasets and 500 archive datasets (close to 3000 including work in progress, private and test datasets)
- 500 new datasets every year (many of them technical)
- Few, easy to override ‘common dimensions’ (e.g. Location, Units of measure)
- Limited SDMX functionality, no artefact versioning
Motivation

The new .Stat Suite or .Stat v8 is a complete overhaul of the previous .Stat version 7. The motivation for the overhaul is fourfold:

- Improved analytical usability of the data (with more combinability through harmonisation and improved structural modelling)
- Faster and better quality data production chains, better industrialisation possibilities
- (End)user friendliness and improved discoverability
- Increase harmonization and alignment with international standards, e.g. SDMX
Broad directions and principles

- SDMX at the core of the new system
  - Larger harmonised data universes - with concept schemes defining ‘data-marts’, inspired by global DSD exercises (nonetheless translated to a dissemination context)
  - Larger shared cubes - DSDs (with DSD design principles kept at the forefront)
  - Creating connections even beyond the data-marts via globally (OECD-wide) maintained artefacts (concepts and code-lists: Reference area, Observation status, Units of measure, Industrial activities, Institutional sectors, Age, Sex, etc.)
  - A more systematic approach to identify measure + unit of measure pairs
  - Unique referential metadata model
  - Versioning introduced
Data modelling - a form of art

- Plenty of trade-offs to be made:
  - Purity of design vs simplicity
    - Considering their impact on: orthogonality, unambiguous representation, sparsity, expertise required from the user to browse the data
  - A certain economy / parsimony: balancing few DSDs, few dimensions, few codes in a code-list
Some recurring modelling choices

- Modelling exactly or modelling functionally (can the two co-exist)
  - [Age group examples: Youth, Working age vs. Y_LE15, Y15T64: consequences on scope, attributes usage, alignment of data, connectivity]
  - [Latest available period] (This could be fixed by a feature)
- Coined terms: representing parsimoniously or redundantly?

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Unit of Measure</th>
<th>Indicator</th>
<th>Unit of Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment</td>
<td>Persons</td>
<td>Unemployment</td>
<td>Persons</td>
</tr>
<tr>
<td>Unemployment</td>
<td>% of active population</td>
<td>Unemployment rate</td>
<td>% (of active population?)</td>
</tr>
</tbody>
</table>

- Generic units of measure or specific measures and change in units

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Unit of measure</th>
<th>Indicator</th>
<th>Unit of measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debt to GDP ratio</td>
<td>Percentage</td>
<td>Debt</td>
<td>Percentage of GDP</td>
</tr>
<tr>
<td>Debt</td>
<td>National currency</td>
<td>Debt</td>
<td>National currency</td>
</tr>
<tr>
<td>GDP</td>
<td>National currency</td>
<td>GDP</td>
<td>National currency</td>
</tr>
</tbody>
</table>
Organising the work

- The Practices section of the Smart Data Practices and Solutions division plays a central role.
- Within the team 4-5 people engage regularly (but not full time) in modelling/migration planning activities
- A Community of Practice for Data Modelling provides the framework for discussions with data producers
  - Many bilateral spin-off exercises
Planned OECD .Stat Suite spaces

**Longer term 2022+**

**Short term 2021**

### .Stat v7 Entry Gate

**Collect space**
- Production optimised or interim structures, public and confidential data, validation and data editing functions

**Process space**
- Production optimised structures, public and confidential data

**Disseminate - Staging**
- Dissemination optimised structures, public data, preparing artefacts and data for publication, previewing

**Disseminate - Final**
- Dissemination optimised structures, public data, externally facing

**Near identical objects, syncing between spaces**

---

**Statworks**
- Metastore

**Dir-Specific**
- ECO ADB
- SDD Browsers

**Statworks**
- SAS, SPSS...

**Sand\-box**

**Algorithms Bank (Git, JN)**

**SDMX Tools**

**EM-PROD1**

**EM-ARC1**

**EM-GIS**

**Carts & Données**

**Excel, E-Views, SAS, SPSS...**

**Dropbox**

**Ad hoc Scraping**

**Policy simulators**

**Limesurvey**

**SDMX APIs**

**Excel questionnaires**

**Email**

---

**R, Python, Stata**

**Alpha**

**Beta**

**.Stat v7**

---

**OECD. Stat Suite v7**

---

**OECD. Stat Suite v7 spaces**

**Excel importer**

**SDMX import**

**In-space calculation, validation engine**

**Transfer with mapping via StructureSets**

---

**Near identical objects, syncing between spaces**

---

**SDMX Tools**

**Algorithms Bank (Git, JN)**

**EM-PROD1**

**EM-ARC1**

**EM-GIS**

**Carts & Données**

**Excel, E-Views, SAS, SPSS...**

**Dropbox**

**Ad hoc Scraping**

**Policy simulators**

**Limesurvey**

**SDMX APIs**

**Excel questionnaires**

**Email**
Migration phases

- Stage 1: Replacing the .Stat Browser v7
  - Discussing data bilaterally, evaluating re-modeling choices [this is where we are]
  - Testing the fit of proposed models in a shared test environment
  - Preparing for beta release, connecting production systems.
- Stage 2: Gradually replacing production systems.
Functional testing aimed at

While discussing data and structures bilaterally with producer teams

- Monitor the stability of the demo environment
- Check basic Create/Read/Update/Delete functionality
- Check features related to authentication/authorisation
- Clarify attributes capabilities, referential metadata, annotations and hierarchies related behaviour

- So it is a two way exercise:
  - modeling the data for the Suite
  - feeding back information to the developer team
Moving forward

Testing the fit of proposed models in a shared test environment

- A shared, realistically sized test environment on OECD premises by end 2020
- Versioning and Agency consistent CRUD operations and basic helper functionalities
- Basic referential metadata functions
Moving even further

Connecting production systems with the new DotStat

- Improved EDD mechanism
- Mapping mechanism between spaces
- Complete referential metadata management and presentation
The OECD data migration approach for .Stat Suite
Approach in general

- Directorate by directorate: review all indicators/variables/other dimensions and try to take out all the information that could be put elsewhere.
  - Some examples of existing indicators:
    - Total tourism employment (direct) as % of total employment
    - Young migrants share (migrants 15 to 29 over total migrants)
    - Share of SME outstanding loans (% of total outstanding business loans)

- Separate:
  - Measure (e.g. Employment (direct))
  - Unit of measure (e.g. Percentage of total employment)
  - Unit multiplier (e.g. Units)
  - Other breakdowns (e.g. Industrial activity = Tourism)

- Model similar indicators in a similar way
- Use the DSD Matrix Generator to do the analysis and generate SDMX structures
## Decomposing indicators in Matrix Generator

<table>
<thead>
<tr>
<th>CONCEPTS</th>
<th>V</th>
<th>REF_AREA</th>
<th>REF_REGION</th>
<th>REPORTING_COUNTRY</th>
<th>COUNTERPART_COUNTRY</th>
<th>MEASURE</th>
<th>UNIT_MEASURE</th>
<th>UNIT_MULTI</th>
<th>TOURISM_FORM</th>
<th>VISITOR_TYPE</th>
<th>ACTIVITY</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DOMESTIC_TOURISM</strong></td>
<td>Total domestic trips</td>
<td>#</td>
<td><em>Z</em></td>
<td>Number</td>
<td>Tourism trips</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td>Tourist #</td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
</tr>
<tr>
<td><strong>DOMESTIC_TOURISM</strong></td>
<td>Overnight visitors (tourists)</td>
<td>#</td>
<td><em>Z</em></td>
<td>Number</td>
<td>Tourism trips</td>
<td>Units</td>
<td>Domestic</td>
<td><em>OVERNIGHT</em></td>
<td>Tourist #</td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
</tr>
<tr>
<td><strong>DOMESTIC_TOURISM</strong></td>
<td>Same-day visitors (excursionists)</td>
<td>#</td>
<td><em>Z</em></td>
<td>Number</td>
<td>Tourism trips</td>
<td>Units</td>
<td>Domestic</td>
<td><em>SAME_DAY</em></td>
<td>Tourist #</td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
</tr>
<tr>
<td><strong>DOMESTIC_TOURISM</strong></td>
<td>Nights in all types of accommodation</td>
<td>#</td>
<td><em>Z</em></td>
<td>Number</td>
<td>Nights spent in tourism accommodation establishments</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td>Tourist #</td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
</tr>
<tr>
<td><strong>DOMESTIC_TOURISM</strong></td>
<td>Hotels and similar establishments</td>
<td>#</td>
<td><em>Z</em></td>
<td>Number</td>
<td>Nights spent in tourism accommodation establishments</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td>HOTELS</td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
</tr>
<tr>
<td><strong>DOMESTIC_TOURISM</strong></td>
<td>Other collective establishments</td>
<td>#</td>
<td><em>Z</em></td>
<td>Number</td>
<td>Nights spent in tourism accommodation establishments</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td>OTHER</td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
</tr>
<tr>
<td><strong>DOMESTIC_TOURISM</strong></td>
<td>Private accommodation</td>
<td>#</td>
<td><em>Z</em></td>
<td>Number</td>
<td>Nights spent in tourism accommodation establishments</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td>PRIVATE</td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
<td><strong>#</strong></td>
</tr>
</tbody>
</table>
Approach in general

- Reuse latest versions of the existing artefacts whenever possible
  - Cross-domain artefacts maintained by SDMX
    - [https://registry.sdmx.org/FusionRegistry/export.html](https://registry.sdmx.org/FusionRegistry/export.html)
  - Global artefacts maintained by SDMX sponsor organisations
    - [https://registry.sdmx.org/FusionRegistry/export.html](https://registry.sdmx.org/FusionRegistry/export.html)
    - [https://webgate.ec.europa.eu/sdmxregistry/](https://webgate.ec.europa.eu/sdmxregistry/) (ESTAT)
    - [https://sdmxcentral.imf.org](https://sdmxcentral.imf.org) (IMF)
  - BIS, UN, WB
  - Internationally agreed artefacts (Global DSDs for National Accounts, Labour, Education...)
  - Nationally agreed artefacts
  - Artefacts already designed and used by the organisation

- Contact maintenance agency to make needed improvements to the artefact (e.g. National Accounts DSD).

- If this fails, copy the artefact to .Stat Suite, make the necessary changes (extend or narrow but do not re-write codes) and change the maintenance agency to your agency.

- Create a completely new artefact only when the needed artefact does not exist at all.
What types of artefacts are needed for .Stat Suite?

- Concept schemes
  - Define concepts that will be used in DSDs as Dimensions and Attributes, also MSD concepts

- Code lists
  - Define representation of the concepts

- Data Structure Definitions (DSDs)
  - Reuse concepts as Dimensions, Attributes or Primary Measure

- Dataflow and Content constraints
  - Constrained views on a DSD

- Category scheme and Categorisations
  - To categorise Dataflows

- Metadata Structure Definition (MSD) and Metadataflows (not possible to manage referential metadata in .Stat Suite yet)
  - Defines structure for metadata and views on a MSD
Interrelations of SDMX artefacts
A Concept Scheme is a collection of Concepts with the help of which data can be fully described. The example concepts are:

- Concepts used as Dimensions: Reference area, Age, Location, Frequency, Time period etc.
- Concepts used as Attributes: Observation status, Source, Unit of measure, Unit multiplier etc.
- Additional Concepts required in Data Structure Definition, e.g. Observation value.

Each Concept has its identifier and name in one or several languages.

Several concept schemes may be maintained
- Organisation-specific common concepts
- Concepts related to one statistical domain
- Concepts used for referential metadata
- Concept schemes maintained by other agencies ([SDMX Cross-domain Concept Scheme](#))

In OECD re-modelling exercise we use: one concept scheme per statistical domain and one common concept scheme for referential metadata.
Concept scheme

- Add “CS_” to the Concept Scheme Identifier
- Concepts may be used as Dimensions, Data Attributes, Metadata Attributes in DSDs and MSDs.
- It is important to re-use cross-domain Concepts:

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Role (identified by SDMX cross-domain Concept)</th>
<th>Other variations used previously should be re-engineered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>Dimension</td>
<td>FREQ</td>
<td>FREQUENCY, PERIODICITY</td>
</tr>
<tr>
<td>Reference Area</td>
<td>Dimension</td>
<td>REF_AREA</td>
<td>LOCATION, COUNTRY</td>
</tr>
<tr>
<td>Measure</td>
<td>Dimension</td>
<td>MEASURE</td>
<td>INDICATOR, VARIABLE</td>
</tr>
<tr>
<td>Time</td>
<td>Time Dimension</td>
<td>TIME_PERIOD</td>
<td>TIME, YEAR</td>
</tr>
<tr>
<td>Unit of measure</td>
<td>Dimension/Data Attribute</td>
<td>UNIT_MEASURE</td>
<td>UNIT_MEAS, UNITS, UNIT</td>
</tr>
<tr>
<td>Unit multiplier</td>
<td>Data attribute</td>
<td>UNIT_MULT</td>
<td>UNIT_MULTIPLIER, POWERCODE, DENOMINATOR</td>
</tr>
<tr>
<td>Primary measure</td>
<td>Primary measure</td>
<td>OBS_VALUE</td>
<td></td>
</tr>
</tbody>
</table>
Codelist

- Codelist is a list of Codes from which Concepts take their values.
- Use prefix “CL_” in the Codelist identifier.
- Several Codelists can be used with one Concept.
  - For example, different codelists for Countries and Regions can be used with the “Reference area” concept in different DSDs.
- Use SDMX cross-domain Codelists where possible.
  - List of cross-domain Codelists: [https://registry.sdmx.org/FusionRegistry/codelists.html](https://registry.sdmx.org/FusionRegistry/codelists.html)
- Do not create a new Codelist if the existing one can be reused fully or partially (e.g. constrained to certain existing values).
- If a codelist maintained by SDMX or another agency fits almost, the missing codes should be proposed to be added to the maintenance agency.
- If it is decided to create a new codelist, follow the [Guidelines on creating a new codelist](https://registry.sdmx.org/FusionRegistry/codelists.html).
## Codelist

- Always use reserved codes:

<table>
<thead>
<tr>
<th>Recommended Code ID</th>
<th>Recommended name</th>
<th>Not correctly used codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>_L</td>
<td>Local extension</td>
<td></td>
</tr>
<tr>
<td>_N</td>
<td>Non response</td>
<td>NSP</td>
</tr>
<tr>
<td>_O</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>_S</td>
<td>Subtotal</td>
<td></td>
</tr>
<tr>
<td>_T</td>
<td>Total</td>
<td>TOT, TOTAL</td>
</tr>
<tr>
<td>_U</td>
<td>No data/unknown</td>
<td>UNK</td>
</tr>
<tr>
<td>_X</td>
<td>Not allocated/unspecified</td>
<td></td>
</tr>
<tr>
<td>_Z</td>
<td>Not applicable</td>
<td>NAP</td>
</tr>
</tbody>
</table>
Codelist

- Codelists support simple hierarchy in the Codelist
  - Parent code can be defined for each code in the Codelist

- Order of codes in the Codelist is not explicit
  - Use annotation of type ORDER to define order that will be applied in the Data Explorer visualisation pages.

- For better understanding of hierarchies outside of the context of hierarchy, alternative names can be provided for lower-level codes (*feature not implemented in Data Explorer yet*):
  - Using Annotations
  - Annotation type: FULL_NAME
  - Text in all relevant languages
DSD defines

- Dimensions (Dimension, Time Dimension, Frequency Dimension)
- Attributes (mandatory or optional)
- Primary measure specifies what type of data is handled by the DSD
  - Use SDMX cross-domain Concept OBS_VALUE

In OECD re-modelling exercise DSD uses one Concept Scheme and re-uses Codelists where possible.

One or several DSDs per statistical domain are usually maintained.

Unnecessary dimensions can be later “switched off” in the Dataflows by constraining them to one value (_Z Not applicable or _T Total).
- Switching off works only in Data Explorer, in data messages _Z and _T will still appear

IDs of the Data Structure Definition could always follow the same convention. For example:
- <ID of the statistical domain>_<meaningful abbreviation of name>, e.g. NA_MAIN
## Dimensions in a DSD

- Dimensions uniquely identify observations
- Order of Dimensions in a DSD defines the order of filters in Data Explorer.
  - Exception for Frequency and Time Period that have their own special behaviour
- Dimensions usually are:

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>REF_AREA</td>
<td>Reference area (ex Country, Location)</td>
<td>SDMX cross-domain Concept REF_AREA</td>
</tr>
<tr>
<td>MEASURE</td>
<td>Measure (ex Indicator, Variable)</td>
<td>SDMX cross-domain Concept MEASURE</td>
</tr>
<tr>
<td></td>
<td>• Dimensions based on common Concepts/Code lists: Age, Sex, Industry classification, etc.)</td>
<td>Break-downs</td>
</tr>
<tr>
<td></td>
<td>• Dimensions not based on common Concepts.</td>
<td></td>
</tr>
<tr>
<td>UNIT_MEASURE</td>
<td>Unit in which the data values are expressed.</td>
<td>Might be also an attribute SDMX cross-domain Concept UNIT_MEASURE</td>
</tr>
<tr>
<td>FREQ</td>
<td>Frequency</td>
<td>SDMX cross-domain Concept FREQ + Codelist CL_FREQ</td>
</tr>
<tr>
<td>TIME_PERIOD</td>
<td>Time</td>
<td>Time dimension SDMX cross-domain Concept TIME_PERIOD</td>
</tr>
</tbody>
</table>
Attributes in a DSD

- Attributes provide additional information to Dimensions
- Attributes can be mandatory or optional
- Attribute allowed attachment level should be defined in the DSD
  - Dataset
  - Dimension or group of dimensions
  - Observation

- Commonly used attributes:

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNIT_MEASURE</td>
<td>Unit of measure used</td>
</tr>
<tr>
<td>UNIT_MULT</td>
<td>Unit multiplier</td>
</tr>
<tr>
<td>BASE_PER</td>
<td>Base period</td>
</tr>
<tr>
<td>DECIMALS</td>
<td>Number of decimal digits displayed in the data table</td>
</tr>
<tr>
<td>OBS_STATUS</td>
<td>Provides coded information on 1) the quality of a value or 2) unusual or missing values.</td>
</tr>
</tbody>
</table>
Attributes vs Referential metadata

- Attributes are closely related to data and need to travel together with the data
  - For example: Unit of measure, Unit multiplier, Reference period, Base period
  - Supported format:
    - Textual information (in one language)
    - Coded information

- Metadata - information that can be transmitted without data
  - For example: Data source, Last updated date, Contact person
  - Supported format:
    - Textual information (multilingual)
    - Coded information
Referential metadata implementation proposal

- A new reference metadata classification has been proposed in OECD to replace the existing 42 categories:
  - Contains 17 categories
  - Based on “Global SDMX metadata concept scheme” that standardizes a set of metadata concepts for exchange between agencies.

- Metadata concepts will be:
  - Data source
  - Contact
  - Compiling agency
  - Dissemination agency
  - Date updated - last updated
  - Statistical concepts and definitions
  - Coverage
    - Geographical coverage
    - Sector coverage
    - Population coverage
    - Time coverage
  - Data compilation
  - Seasonal adjustment
  - Data validation
  - Comment
  - Quality management - quality assessment
  - Recommended uses and limitations
The Coverage concepts allow for these mutually exclusive uses:

1. Put all coverage metadata into the single Coverage concept, or;
2. Split coverage metadata into the specific ...coverage concepts.

Using both options for coverage metadata in a dataset should not be done to avoid ambiguity for end-users when reading metadata concepts. In the mapping lines, a dotted line depicts when a choice must be made to put the migrate the metadata to Coverage or a specific ...coverage concept.
Dataflow and Content Constraints

- Dataflows are constrained views on the content of a DSD
- IDs of the Dataflows could always follow the same convention to avoid duplication of IDs.
  - For example, composed of the name of subdomain or reporting table that is followed by the “@” character and the Data Structure Definition ID: DOMESTIC@TOURISM_TRIPS
- The more limiting the Dataflow constraint is, the better user experience in .Stat Suite is achieved.
- It is recommended to limit the number of dimensions not constrained to a single value in a Dataflow relatively low.
- Constraining a Dimension to a single value:
  - When a dimension is constrained to one of these codes: Total (T), Local extension (L) or Non applicable (Z), it will be hidden from Data Explorer visualisation page
  - When a dimension is constrained to one specific value like Women, Annual etc.:
    - It will be displayed in the title on Data Explorer visualisation page
    - Filter box will not be displayed for this dimension
Dataflow and Content Constraints

- Stat Suite supports Content Constraints of two types:
  - “Allowed” Content Constraints
  - “Actual” Content Constraints
- “Allowed” Content Constraints are uploaded with other structural artefacts
  - Used to validate data during the data uploads
  - Applied in data queries to return constrained data for a specific Dataflow
  - Should be created in the following way (this complexity can be hidden with the tools like Matrix Generator):
    - Content constraint type: “Allowed”
    - Cube region include = “true”
    - Key Value: ID must match Dimension ID or Attribute ID
    - Value: Limit the codes only to the ones that are used in this dataflow
- “Actual” Content Constraints are automatically generated during the data uploads
  - Indicate available values for dimensions and attributes in a Dataflow
  - Applied in search features to index and display only codes for which data exists
  - Applied in data visualisation pages to hide filter values with no data
Dataflow default view

- Dataflow layout for .Stat Suite tables can be specified using different types of annotations attached to each Dataflow:
  - LAYOUT_ROW: defines what dimensions will be shown in rows.
  - LAYOUT_COLUMN: defines what dimensions will be shown in columns.
  - LAYOUT_ROW_SECTION: defines what dimensions will be shown in row sections.
  - NOT_DISPLAYED: defines what dimensions or attributes or their single values will be hidden from the table output.
  - DEFAULT: defines dimension values that will be selected by default for example, REF_AREA=AU.
  - UNIT_MEASURE_CONCEPTS: defines what dimensions/attributes are to be shown together as part of Unit of measure. If not provided, then the default Unit of measure concepts will be: Unit of measure, Unit multiplier and Base period (configurable).
Category scheme and Categorisations

- One „main“ Category Scheme defined for each data space.
- Category Scheme needs to have a short and clear name in each .Stat Suite language that will be visible on the Data Explorer homepage (e.g. Topic).
- Categories in a Category Scheme may be organised in hierarchies.
  - Usually not more than 10-15 categories on the top-level
  - Only the first level of hierarchy is selectable on the Data Explorer home page
  - Lower level categories are shown as information on the home page and can be used for filtering when higher-level category is previously selected by the user.
- All Dataflows in a space should be categorised using the “main” Category Scheme to be indexed and become discoverable for the end-users.
- Each Dataflow can be attached to one or more Categories in the Category Scheme
Microdata

- Micro-datasets may be very large
  - Microdata typically has many “fields”. Creditor Reporting System (CRS) dataset has approximately 50 fields
  - Microdata fields have different data types (date, string, number, boolean etc.)
- A “drill-down” feature from the aggregated data view may be required to display the microdata in the microdata browser.
Microdata implementation proposal

- Microdata structure is defined in a DSD:
  - Aggregated data dimensions are usual dimensions in a DSD
  - Microdata fields (displayed in the microdata browser) are dimensions or attributes.
    - There may be need for a dimension that uniquely identifies microdata records (e.g. MD_ID) - “Total”/”0” for aggregated data, another integer for microdata
    - Additional dimensions can be hidden with the NOT_DISPLAYED annotations from the aggregated view.
  - A special dimension that distinguishes aggregated data from microdata (e.g. MD_DIM) needs to be added. This dimension allows two values:
    - Total (_T) for aggregated data
    - Microdata (MD) for microdata
    - ID of this dimension is then provided in the DF annotation with type MICRODATA.
  - Microdata fields that would be displayed in the microdata browser can be defined in an optional annotation with type MICRODATA_CONCEPTS.
General recommendations

- Dataflows, Category schemes, Codelists and Concepts should have names in all languages used in the .Stat Suite instance (English, French, Arabic, Spanish, Khmer, ...).
- Names of artefacts used for dissemination should be kept brief and understandable.
- Use the same IDs and labels for the same concepts and codes.
- Additional information or complete definition of the artefact can be provided as a description.
- Make sure files have the right encoding and special characters are entered correctly.
- Avoid grammatical errors and typos.
General recommendations

- Follow guidelines for versioning
  - If the artefact is marked as final, only these changes can be done without increasing the version:
    - Name (translations), Description
    - Annotation
    - Valid from and valid to
    - StructureURL, ServiceURL, URI
    - isExternalReference
  - Increase the MAJOR version (1.0->2.0) when changes are not backward compatible.
    - Data message produced using and validated with the previous version of an artefact (e.g. a DSD) cannot be successfully validated using the newest version of the same artefact.
      - Added dimensions/attributes to a DSD
      - Removed dimensions/attributes from a DSD
      - Removed codes from the Codelist
  - Increase the MINOR version (1.0->1.1) when changes are backward compatible.
    - Data message produced using and validated with the previous version of an artefact (e.g. a DSD) can still be successfully validated using the newest version of the same artefact
    - Data message that is produced and validated with the new version of an artefact (e.g. a DSD) cannot be validated using the previous version of the same artefact.
      - Added new codes to the Code list
Preparing data for .Stat Suite uploads

Several options are available. All of these could be considered depending on the use case:

- Use any of the available tools to generate data in CSV or XML format (SDMX Converter, ILO SMART, ISTAT Excel2CSV etc.)
- Prepare SDMX-CSV files manually using the csv output from .Stat v7
- If there is SDMX-RI implementation available, in case of global DSDs data can be mapped using the Mapping Assistant
Mapping Assistant as a .Stat Suite V8 Migration Tool

- Mapping Assistant already configured with OECD.Stat V7 for creating SDMX-ML data messages to send to Eurostat and the ECB
- Allows early and easy migration from V7 to V8 production versions
- Temporary solution and will be obsolete when V7 is no longer updated
- Future purpose: Could mappings be exported and used in EDD mapping files?
Mapping Assistant: National Accounts Migration
NAMAIN T0101 example

Added .Stat Suite V8 SDMX Structure file
- AGENCY: OECD.SDD
- ID: NAMAIN
- VERSION: 1.0
Mapping Assistant: National Accounts
Migration
NAMAIN T0101 example

- Created new category scheme for National Accounts
- Created category NAMAIN
- Added data flow NAMAIN_0101_A
Mapping Assistant: National Accounts Migration
NAMAIN T0101 example

- SQL query on OECD.Stat V7 dataset
Mapping Assistant: National Accounts
Migration
NAMAIN T0101 example

- Added mapping set NAMAIN_0101_A
- Used CSV files to map between V7 and V8 datasets
- Dataflow set in Production
Test Client: Discover query syntax to create an SDMX data message for uploading to the DLM

- SDMX-ML 2.1 Generic
- REST compatible query
- Dimension at Observation: TIME_PERIOD
- Data via web service updated when dataset in OECD.Stat V7 is updated: https://stats.oecd.org/soisws/rest/data/OECD.SDD,NAMAIN_0101_A,1.0/all/ALL/?detail=full&dimensionAtObservation=TIME_PERIOD
Tourism Datasets Re-modelling Process Using SDMX Matrix Generator
SDMX Matrix Generator

- Purpose: allow statisticians to easily create and maintain dataset-oriented structural metadata without knowing a lot about SDMX
- Inspired by design of Global DSD design matrix. Made it functional by adding the SDMX generation and prefilling part
- Open source tool hosted on GitHub with a user guide
  - https://github.com/OECDSTD/sdmx-matrix-generator
  - or Google for “Github SDMX matrix generator”
- Template already contains commonly used cross-domain concepts and codelists
- Excel workbook, step-by-step wizard approach
- Can import and export from/to files and SDMX web services such as .Stat, Global Registry
- For different domains and structures although there are some structural prerequisites
Re-modelling Tourism datasets

- Analysed the following Stat v7 datasets:
  - Inbound tourism
  - Outbound tourism
  - Domestic tourism
  - Key tourism indicators
  - Enterprises and employment in tourism
Decomposing indicators in Matrix Generator

1. Copy all indicators from the datasets under analysis.
2. For each original indicator identify what concepts describe it and what the possible values for each concept might be.
   1. Add new columns for the identified concepts
   2. Fill the cells for each used concept with a single value, % (several values) or # (all possible values) or blank where the concept is unused.
## Decomposing indicators in Matrix Generator

<table>
<thead>
<tr>
<th>Source</th>
<th>Existing indicators</th>
<th>UNIT_MEASURE</th>
<th>MEASURE</th>
<th>UNIT_MULTI</th>
<th>TOURISM_FORM</th>
<th>VISITOR_TYPE</th>
<th>ACTIVITY</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOMESTIC TOURISM</td>
<td>Total domestic trips</td>
<td>Number</td>
<td>Tourism trips</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td><em>T</em></td>
<td>Tourism</td>
</tr>
<tr>
<td>DOMESTIC TOURISM</td>
<td>Overnight visitors (tourists)</td>
<td>Number</td>
<td>Tourism trips</td>
<td>Units</td>
<td>Domestic</td>
<td>OVERNIGHT</td>
<td><em>T</em></td>
<td>Tourism</td>
</tr>
<tr>
<td>DOMESTIC TOURISM</td>
<td>Same-day visitors (excursions)</td>
<td>Number</td>
<td>Tourism trips</td>
<td>Units</td>
<td>Domestic</td>
<td>SAME_DAY</td>
<td><em>T</em></td>
<td>Tourism</td>
</tr>
<tr>
<td>DOMESTIC TOURISM</td>
<td>Nights in all types of accommodation</td>
<td>Number</td>
<td>Nights spent in tourism accommodation establishments</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td><em>T</em></td>
<td>Tourism</td>
</tr>
<tr>
<td>DOMESTIC TOURISM</td>
<td>Hotels and similar establishments</td>
<td>Number</td>
<td>Nights spent in tourism accommodation establishments</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td>HOTELS</td>
<td>Tourism</td>
</tr>
<tr>
<td>DOMESTIC TOURISM</td>
<td>Other collective establishments</td>
<td>Number</td>
<td>Nights spent in tourism accommodation establishments</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td>OTHER</td>
<td>Tourism</td>
</tr>
<tr>
<td>DOMESTIC TOURISM</td>
<td>Private accommodation</td>
<td>Number</td>
<td>Nights spent in tourism accommodation establishments</td>
<td>Units</td>
<td>Domestic</td>
<td><em>T</em></td>
<td>PRIVATE</td>
<td>Tourism</td>
</tr>
</tbody>
</table>
Defining concept scheme based on the decomposed indicators

1. Move identified concepts from “Decompose indicators” sheet to “Concept scheme” sheet.
2. Assign a role in DSD(s) to each concept: Dimension, Time Dimension, Attribute (with attachment level and mandatory/optional selection), Measure.
3. Provide clear names and descriptions for each concept in necessary languages.
4. Decide if the concept is coded (has a pre-defined list of values) or un-coded.
   1. For each coded concept a separate codelist sheet will be generated when moving to the next step. All codelist sheets should be updated to have all possible codes there.
Defining concept scheme based on the decomposed indicators

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>Code List and Registry Link</th>
<th>Type</th>
<th>Min length</th>
<th>Max length</th>
<th>Concept Role</th>
<th>Concept Name:can</th>
<th>Concept Definition:en</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>REPORTING_COUNTRY</td>
<td>CL_AREA</td>
<td>Reporting country</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>COUNTERPART_COUNTRY</td>
<td>CL_AREA</td>
<td>Top counterpart country</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>RBI_AREA</td>
<td>CL_AREA</td>
<td>Reference area</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>MEASURE</td>
<td>CL_TOURIST_MEASURE</td>
<td>Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>TOURISM_FORM</td>
<td>CL_TOURIST_FORM</td>
<td>Tourism form</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>VISITOR_TYPE</td>
<td>CL_VISITOR_TYPE</td>
<td>Visitor type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>ACCOMMODATION_TYPE</td>
<td>CL_ACCOMMODATION_TYPE</td>
<td>Accommodation type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>ACTIVITY</td>
<td>CL_TOURIST_ACTIVITY</td>
<td>Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>SOURCE</td>
<td>CL_TOURIST</td>
<td>Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>UNIT_MEASURE</td>
<td>CL_UNIT_MEASURE</td>
<td>Unit of measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>FREQ</td>
<td>CL_FREQ</td>
<td>Frequency of observation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D(Time)</td>
<td>TIME_PERIOD</td>
<td>wooded</td>
<td>ObservationalTimePeriod</td>
<td>Time period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>OBS_VALUE</td>
<td>wooded</td>
<td>Observation value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A(optional)</td>
<td>UNIT_MULT</td>
<td>CL_UNIT_MULTI</td>
<td>Unit multiplier</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A(optional)</td>
<td>OBS_STATUS</td>
<td>CL_OBS_STATUS</td>
<td>Observation status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A(optional)</td>
<td>OBS_STATUS2</td>
<td>CL_OBS_STATUS</td>
<td>Observation status nr 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A(optional)</td>
<td>CONF_STATUS</td>
<td>CL_CONF_STATUS</td>
<td>Confidentiality status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A(optional)</td>
<td>DECIMALS</td>
<td>CL_DECIMALS</td>
<td>Decimals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Filling the DSD-Concept Matrix

1. Move previously identified concepts to the DSD-Concept Matrix sheet.
2. Fill the matrix in the following way:
   1. In rows identify Dataflows (visualisation tables) and what concepts they are using. Mark each concept as follows:
      1. # - concept is used, no constraint
      2. % - concept is used with a subset constraint
      3. <code> - Concepts is used, but is fixed Single code constraint
      4. Blank - concept is not used
   2. For each Dataflow define the DSD structure that it is built on.
# Filling the DSD-Concept Matrix

**Matrix Legend**
- Any code
- Several codes allowed
- Single code
- Unused

<table>
<thead>
<tr>
<th>DSD Id</th>
<th>Data flow id</th>
<th>CONCEPTS</th>
<th>REPORTING COUNTRY</th>
<th>SOMETHING ELSE</th>
<th>REF. AREA</th>
<th>MEASURE</th>
<th>TOURISM FORM</th>
<th>VISITOR</th>
<th>ACTIVITY</th>
<th>SOCIAL ECONOMIC</th>
<th>UNIT MEASURE</th>
<th>UNITS</th>
<th>TIME PERIOD</th>
<th>OBS. VALUE</th>
<th>OBS. UNIT</th>
<th>OBS. VALUE</th>
<th>OBS. UNIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOURISM_TRIPS</td>
<td>DOMESTIC@TOURISM_TRIPS</td>
<td>#_Z</td>
<td>%</td>
<td>D # %</td>
<td>AGG TOUR</td>
<td># NUMBER</td>
<td># # # #</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOURISM_TRIPS</td>
<td>INBOUND@TOURISM_TRIPS</td>
<td>#_#</td>
<td>%</td>
<td>I # #</td>
<td>AGG TOUR</td>
<td># NUMBER</td>
<td># # # #</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOURISM_TRIPS</td>
<td>OUTBOUND@TOURISM_TRIPS</td>
<td># #</td>
<td>TOTAL VISITORS</td>
<td>O #</td>
<td>AGG TOUR</td>
<td># NUMBER</td>
<td># # # #</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOURISM_KEY</td>
<td>EMP@TOURISM_KEY</td>
<td>#_EMP</td>
<td>#</td>
<td>#</td>
<td>PERSONS</td>
<td># # # # # #</td>
<td># # # # # #</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOURISM_KEY</td>
<td>ENT@TOURISM_KEY</td>
<td>#_ENT</td>
<td>%</td>
<td>#</td>
<td>NUMBER</td>
<td># # # # # #</td>
<td># # # # # #</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOURISM_KEY</td>
<td>ENT.EMP@TOURISM_KEY</td>
<td>#_%</td>
<td>%</td>
<td>%</td>
<td># # # # # #</td>
<td># # # # # #</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOURISM_KEY</td>
<td>KEY.IND.ROC@TOURISM_KEY</td>
<td>#_%</td>
<td>AGG TOUR</td>
<td>%</td>
<td># # # # # #</td>
<td># # # # # #</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Updating the DSD and Dataflow details

- For each DSD and Dataflow defined on DSD-Concept sheet it is important to provide further details on the sheets DSDs and Dataflows:
  - Identifier (ID, Agency, Version)
  - Names and descriptions
  - Is Final (true or false)

- For Dataflows it is possible to additionally specify:
  - Categorisation (relation to a Category Scheme and Category in it)
  - Default layout
    - What dimensions go to rows, columns and rows sections
    - Pre-selected dimension values
    - Hidden dimension and attribute values
## Updating the DSD and Dataflow details

<table>
<thead>
<tr>
<th>ID*</th>
<th>Agency*</th>
<th>Version*</th>
<th>IsFinal*</th>
<th>Name:en</th>
<th>Description:en</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOURISM_KEY</td>
<td>OECD.CFE</td>
<td>1.0</td>
<td>TRUE</td>
<td>DSD for tourism</td>
<td></td>
</tr>
<tr>
<td>TOURISM_TRIPS</td>
<td>OECD.CFE</td>
<td>1.0</td>
<td>TRUE</td>
<td>DSD for tourism trips</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID*</th>
<th>Name:en</th>
<th>Description:en</th>
<th>LAYOUT_ROW*</th>
<th>LAYOUT_COLUMN*</th>
<th>LAYOUT_ROW_SECTION*</th>
<th>NOTDISPLAYED*</th>
<th>DEFAULT*</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOMESTICTOURISM_TRIPS</td>
<td>Domestic tourism</td>
<td>Domestic tourism compt.</td>
<td>TIME_PERIOD</td>
<td>SOURCE</td>
<td>REPORTING_COUNTRY</td>
<td>UNIT, MULT=0, ACTIVITY, TOURISM, FORM_FR</td>
<td>SOURCE=TDSS</td>
</tr>
<tr>
<td>INBOUNDTOURISM_TRIPS</td>
<td>Inbound tourism</td>
<td>Inbound tourism compt.</td>
<td>TIME_PERIOD</td>
<td>SOURCE</td>
<td>REPORTING_COUNTRY</td>
<td>UNIT, MULT=0, ACTIVITY, TOURISM, FORM_FR</td>
<td>SOURCE=TDSS</td>
</tr>
<tr>
<td>OUTBOUNDTOURISM_TRIPS</td>
<td>Outbound tourism</td>
<td>Outbound tourism compt.</td>
<td>TIME_PERIOD</td>
<td>SOURCE</td>
<td>REPORTING_COUNTRY</td>
<td>UNIT, MULT=0, ACTIVITY, TOURISM, FORM_FR</td>
<td>REPORTING_COUNTRY=AU</td>
</tr>
<tr>
<td>ENGTOURISM_KEY</td>
<td>Employment in tourism</td>
<td>Employment in tourism</td>
<td>TIME_PERIOD</td>
<td>REF_AREA,UNIT, MEASURE</td>
<td>UNIT, MULT=0, CONF_STATUS=F, FREQ, MEASURE, REF_AREA=AU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENTERPRISETOURISM_KEY</td>
<td>Enterprises in tourism</td>
<td>An establishment is at ACTIVITY</td>
<td>TIME_PERIOD</td>
<td>REF_AREA</td>
<td>UNIT, MULT=0, CONF_STATUS=F, FREQ, MEASURE, REF_AREA=AU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENT ENTERPRISETOURISM KEY</td>
<td>Enterprises and employment</td>
<td>Data on enterprises at ACTIVITY</td>
<td>TIME_PERIOD</td>
<td>REF_AREA,MEASURE,UNIT, MEASURE</td>
<td>UNIT, MULT=0, CONF_STATUS=F, FREQ, MEASURE, REF_AREA=AU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEY. RECORDS</td>
<td>Key tourism indicators</td>
<td>Tourism GDP corresponds REF_AREA</td>
<td>TIME_PERIOD</td>
<td>REF_AREA,MEASURE,UNIT, MEASURE</td>
<td>UNIT, MULT=0, ACTIVITY, CONF_STATUS=F, FREQ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Defining content constraints

- For those dimensions and attributes that have % or <code> in DSD Matrix next to them, content constraint details need to be additionally specified on the codelist sheet - “1” has to be put to next to the used codes.
Generating XML and uploading structures and data to .Stat Suite
### Domestic tourism

**Reporting country:** Australia  
**Source:** Tourism demand surveys  
**Unit of measure:** Number

<table>
<thead>
<tr>
<th>Measure</th>
<th>Visitor type</th>
<th>Accommodation type</th>
<th>Time period</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total visitors</td>
<td>Total visitors</td>
<td>Total</td>
<td>(≈) 214,272,000</td>
<td>(≈) 213,264,000</td>
<td>(≈) 223,218,000</td>
<td>(≈) 227,264,000</td>
<td></td>
</tr>
<tr>
<td>Total visitors</td>
<td>Overnight visitors (tourists)</td>
<td>Total</td>
<td>(≈) 73,527,000</td>
<td>(≈) 68,539,000</td>
<td>(≈) 68,143,000</td>
<td>(≈) 70,977,000</td>
<td></td>
</tr>
<tr>
<td>Total visitors</td>
<td>Same-day visitors (excursionists)</td>
<td>Not applicable</td>
<td>(≈) 140,745,000</td>
<td>(≈) 144,725,000</td>
<td>(≈) 155,075,000</td>
<td>(≈) 158,387,000</td>
<td></td>
</tr>
<tr>
<td>Nights spent in tourism accommodation establishments</td>
<td>Overnight visitors (tourists)</td>
<td>Total</td>
<td>(≈) 265,390,000</td>
<td>(≈) 263,405,000</td>
<td>(≈) 264,290,000</td>
<td>(≈) 269,235,000</td>
<td></td>
</tr>
<tr>
<td>Nights spent in tourism accommodation establishments</td>
<td>Overnight visitors (tourists)</td>
<td>Hotels and similar establishments</td>
<td>(≈) 79,938,000</td>
<td>(≈) 70,178,000</td>
<td>(≈) 72,430,000</td>
<td>(≈) 72,043,000</td>
<td></td>
</tr>
<tr>
<td>Nights spent in tourism accommodation establishments</td>
<td>Overnight visitors (tourists)</td>
<td>Private accommodation</td>
<td>(≈) 137,138,000</td>
<td>(≈) 125,022,000</td>
<td>(≈) 122,003,000</td>
<td>(≈) 125,093,000</td>
<td></td>
</tr>
<tr>
<td>Nights spent in tourism accommodation establishments</td>
<td>Other collective accommodation</td>
<td>(≈) 69,314,000</td>
<td>(≈) 67,306,000</td>
<td>(≈) 69,028,000</td>
<td>(≈) 69,100,000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Visualising Dataflows in .Stat Suite

**Key tourism indicators**

Measure: Share of GVA (direct) • Unit of measure: Percentage of total GVA

<table>
<thead>
<tr>
<th>Reference area</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>(w) 2.6</td>
<td>(w) 2.7</td>
<td>(w) 2.8</td>
<td>(w) 2.8</td>
</tr>
<tr>
<td>Canada</td>
<td>(w) 2.1</td>
<td>(w) 2.2</td>
<td>(w) 2.2</td>
<td>(w) 1.9</td>
</tr>
<tr>
<td>Switzerland</td>
<td>(w) 2.9</td>
<td>(w) 2.8</td>
<td>(w) 2.9</td>
<td>(w) 2.9</td>
</tr>
<tr>
<td>Chile</td>
<td>(w) 3.2</td>
<td>(w) 3.1</td>
<td>(w) 3.2</td>
<td>(w) 3.2</td>
</tr>
<tr>
<td>Colombia</td>
<td>(w) 2.4</td>
<td>(w) 2.2</td>
<td>(w) 2.1</td>
<td>(w) 2.1</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>(w) 4.2</td>
<td>(w) 4.2</td>
<td>(w) 4.2</td>
<td>(w) 4.5</td>
</tr>
<tr>
<td>Croatia</td>
<td>(w) 2.4</td>
<td>(w) 2.4</td>
<td>(w) 2.6</td>
<td>(w) 2.8</td>
</tr>
<tr>
<td>Denmark</td>
<td>(w) 4.3</td>
<td>(w) 5.5</td>
<td>(w) 5.5</td>
<td>(w) 5.5</td>
</tr>
<tr>
<td>Denmark</td>
<td>(w) 2.2</td>
<td>(w) 2.2</td>
<td>(w) 2.2</td>
<td>(w) 2.2</td>
</tr>
<tr>
<td>Estonia</td>
<td>(w) 3.7</td>
<td>(w) 3.6</td>
<td>(w) 3.7</td>
<td>(w) 3.7</td>
</tr>
<tr>
<td>Finland</td>
<td>(w) 4.6</td>
<td>(w) 4.7</td>
<td>(w) 4.9</td>
<td>(w) 3.9</td>
</tr>
<tr>
<td>Greece</td>
<td>(w) 4.7</td>
<td>(w) 4.7</td>
<td>(w) 4.9</td>
<td>(w) 3.9</td>
</tr>
<tr>
<td>Croatia</td>
<td>(w) 5.3</td>
<td>(w) 5.5</td>
<td>(w) 5.0</td>
<td>(w) 5.1</td>
</tr>
<tr>
<td>Iceland</td>
<td>(w) 2.8</td>
<td>(w) 2.7</td>
<td>(w) 2.7</td>
<td>(w) 2.6</td>
</tr>
<tr>
<td>Italy</td>
<td>(w) 5.7</td>
<td>(w) 5.8</td>
<td>(w) 5.3</td>
<td>(w) 5.1</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>(w) 1.3</td>
<td>(w) 1.4</td>
<td>(w) 1.5</td>
<td>(w) 1.1</td>
</tr>
<tr>
<td>Lithuania</td>
<td>(w) 3.7</td>
<td>(w) 3.8</td>
<td>(w) 3.3</td>
<td>(w) 3.1</td>
</tr>
<tr>
<td>Latvia</td>
<td>(w) 7.1</td>
<td>(w) 7.4</td>
<td>(w) 4.4</td>
<td>(w) 4.2</td>
</tr>
<tr>
<td>Mexico</td>
<td>(w) 8.4</td>
<td>(w) 8.2</td>
<td>(w) 8.2</td>
<td>(w) 8.5</td>
</tr>
<tr>
<td>Netherlands</td>
<td>(w) 5.7</td>
<td>(w) 5.2</td>
<td>(w) 3.3</td>
<td>(w) 3.8</td>
</tr>
</tbody>
</table>
Visualising Dataflows in .Stat Suite

Enterprises and employment in tourism
Reference area: Australia

<table>
<thead>
<tr>
<th>Activity</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tourism: Tourism services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accommodation services for visitors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food and beverage services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road passenger transport</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rail passenger transport</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hotel, camping, and other accommodation services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other tourism services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cultural activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sports and recreation services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail trade of souvenirs, specific tourist characteristics goods</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure: Employment</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tourist</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accommodation services for visitors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food and beverage services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road passenger transport</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rail passenger transport</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hotel, camping, and other accommodation services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other tourism services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cultural activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sports and recreation services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail trade of souvenirs, specific tourist characteristics goods</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>