New IT solutions for item list management and data validation

4th Inter-Agency Coordinating Group Meeting
October 23-25, 2017
Washington, DC
Eurostat’s PPP system supports the full life-cycle of the PPP process:

- Item list management
- Price collection
- Validation of prices and calculation survey-level PPPs
- Aggregation of PPPs and estimation of PPPs for non-survey years (extrapolations, interpolations, apply VAT and tips)
- Management of non-price data (GDP expenditure weights, CPIs, VAT, tips, exchange rates, population)
- Management and calculation of reference PPPs
- Calculation of overall results at different levels of aggregation (BHs, aggregate headings and analytical categories) and dissemination

- Actors involved: Countries, external contractors acting as survey coordinators and Eurostat
PPP Tools have been developed and are used in production since 2006

It has evolved gradually over the years

- Item List Management Tool was the first tool developed
- Validation and Data Entry tools developed initially for the consumer goods and services surveys (2006)
- Validation and Data Entry tools were gradually adapted to cover all the surveys of the PPP process (2007-2016), incl. non price data i.e. Expenditure Weights, VAT, tips, CPIs
- Aggregation Tool developed later (2010)
- National Accounts Tool developed in 2012
- Auxiliary Data Tool in 2016
- Metadata Tool: in progress
Benefits

- **Cost Effectiveness:**
  - Cost effective operations through the reduction of burden of the exchange of excel files

- **Time Efficiency:**
  - Improved the timeliness of the surveys due to the efficient collaboration of experts

- **Data Quality:**
  - Improved quality through the elimination of error prone operations.
PPP Architectural Overview

- Central Database
  - Input Data
  - Auxiliary Data
  - Metadata
  - Output Data
- Item List Management Tool
  - Item List (XML) & SDMX Specs
- Data Entry Tool
  - Price Data (SDMX-ML)
  - Survey Booklets, Questionnaires, Reports
- Data Loading Web Service
- Item List Management Tool
- Data Loader WS Client
- Dissemination Database (Eurobase)
- eDAMIS
- Validation Tool
- Auxiliary Data Tool
- National Accounts Tool
- Aggregation / Visualization Tool
- Publishing
- Reporting

- Reporting
- Data Loader WS Client
- Dissemination Database (Eurobase)
Collaborative Survey Design

The Item List Management Tool allows countries to collaborate, communicate and exchange proposals to design each survey in an iterative way.

Coordinated Data Collection

The Data Entry Tool is synchronized with the current item list and item specs via a direct XML communication to the ILMT. Price data can be entered and validated for outliers in off-line, multi-user mode.

Collaborative Data Validation

Using the Validation Tool survey coordinators perform inter-country validation and cooperate on-line with country experts to complete data editing. While validation proceeds everybody has access to indicators of other countries and to calculated PPPs.
A web application that manages the entire process of creating item lists, including:

- the creation of pre-survey lists;
- the entering of the pre-survey results by countries;
- the creation of draft and final lists;
- the translation of final item lists by countries into their own language(s);
- Management of surveys and survey instances;
- Management of BHs and SPDs (by Eurostat PPP administrators);
The Item List Management Tool has been customized to support item list management for the ICP process. In the current setup:

- **Actors involved:**
  - the Global ICP Unit (GIU)
  - and Regional Implementing Agencies (RIAs)
- The ILMT includes only items from the global core list (as a first step)
- RIAs and the GIU collaborate to compile the global core list and the final regional lists
- RIAs are able to
  - manage their own regional items
  - carry out a pre-survey (at regional level)
  - translate the item specifications, through the tool
An (off-line) Java application that is used to enter and validate the prices observed for the products on the final list.

- The tool communicates with the Item List Management Tool to retrieve item lists (in XML format).
- It operates in two modes:
  - price collector mode, used simply for data entry
  - national coordinator mode, which in addition allows the merging of price collectors’ files and performing intra-country validation of the final dataset;
- SDMX-ML format is used as internal representation format
A web application which is used to visualize the countries' price observations, to validate them and to communicate with the countries

- intra-country validation
- inter-country validation (via Quaranta Tables)
- calculation of basic heading PPPs and aggregation at survey level (Official and Ad hoc Quaranta tables)
- data versioning at observation level, with possibility to restore to a previous state
- Multiple validation rounds
- Gap-filling and fixity procedures
Involved actors work together in an iterative way, through the creation of different list types for each step, in order to compile the final list of items that will be included in the price survey.

- Unique set of reusable items, shared among the item lists
- An item list has an owner (region, country, global unit) and a specific type (pre-survey list, draft list, final list etc.)
- An item list may contain different versions of item specifications. e.g. a region may propose a different value for a parameter, thus creating a new item version of the same item.

- During a user’s session with the tool, the user works with a single ‘working list’ while at the same time can view other lists
- Work on-line. No need to exchange files
ILMT Process

Initial List
- Starting point
- copy of previous list

Pre-survey List
- add item specification
- update item specification
- remove item specification
- Add questions

Regional Pre-survey lists
- one list for each region
- each region can update / amend its own item list
- each region provides information about availability and importance of items
- each region provides a reply to the questions added by GIU

Draft List
- accommodates feedback received from regions during the pre-survey
- The list is discussed at meetings with RIAs

Final List
- Based on meeting discussions a final list is compiled
- Each item is associated with one or more regions

Final Regional Lists
- created as copy of the final list
- one list for each region
- each list includes only items available in the region
- subset of the final list

Translated regional lists
- regions translate the lists using the tool
- regions can export survey booklet (english or translated)

Validation List
- Copy of the final list
- Used by the validation tool
- Items may be split during validation if the item specification is too generic

GIU → GIU → RIAs → GIU → GIU → RIAs → RIAs → GIU
While working with an item list you can:

- Propose a modified item specification
- Introduce a new item specification
- Remove an item specification from your own list
- View what other regions are proposing for the same item
- Add to your own list an item specification proposed by another region(s)

- Compare specifications of the same item belonging to different lists.
- Add pre-survey questions/answers (case of pre-survey specific item lists)
- View pre-survey answers and notes from other regions
- Export numerous reports, including comparison reports and reports on availability/importance of items, survey catalogue etc.

Working with ILMT
Update an item specification

Insert a new item specification
Based on SPD parameters
Add a question to an item of the pre-survey list

Reply to pre-survey questions
Item List Management Tool

[Image of a software interface for managing item lists, showing detailed information about rice and other food items, with options for viewing availability, importance, pre-survey questions, and characteristics.]
### Item List Management Tool

<table>
<thead>
<tr>
<th>Code</th>
<th>11.03.12.1.01.aa</th>
<th>11.03.12.1.01.ab</th>
<th>11.03.12.1.01.bc</th>
<th>11.03.12.1.01.cd</th>
<th>11.03.12.1.01.ca</th>
<th>11.03.12.1.01.da</th>
<th>11.03.12.1.01.db</th>
<th>11.03.12.1.01.cc</th>
<th>11.03.12.1.01.de</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Men's overcoat</td>
<td>Men's overcoat</td>
<td>Men's parka</td>
<td>Men's parka</td>
<td>Men's coat</td>
<td>Men's parka</td>
<td>Men's parka</td>
<td>Men's coat</td>
<td>Men's parka</td>
</tr>
<tr>
<td>Groups</td>
<td>E N S W</td>
<td>E N S W</td>
<td>E N S W</td>
<td>E N S W</td>
<td>N W</td>
<td>N W</td>
<td>N W</td>
<td>N W</td>
<td>N W</td>
</tr>
<tr>
<td>Brand</td>
<td>Well Known</td>
<td>Well Known</td>
<td>Well Known</td>
<td>Well Known</td>
<td>Well Known</td>
<td>Well Known</td>
<td>Well Known</td>
<td>Well Known</td>
<td>Single/multiple</td>
</tr>
<tr>
<td>Brand/size/plain</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
</tr>
<tr>
<td>Style</td>
<td>casual</td>
<td>casual</td>
<td>casual</td>
<td>casual</td>
<td>casual</td>
<td>casual</td>
<td>casual</td>
<td>casual</td>
<td>casual</td>
</tr>
<tr>
<td>Material</td>
<td>70 - 100% wool,</td>
<td>50 - 80% wool,</td>
<td>cotton,</td>
<td>cotton,</td>
<td>cotton,</td>
<td>100% synthetic</td>
<td>100% synthetic</td>
<td>100% synthetic</td>
<td>100% synthetic</td>
</tr>
<tr>
<td></td>
<td>rest synthetic</td>
<td>rest synthetic</td>
<td>synthetic or mix</td>
<td>synthetic or mix</td>
<td>synthetic or mix</td>
<td>(100% synthetic)</td>
<td>(100% synthetic)</td>
<td>(100% synthetic)</td>
<td>(100% synthetic)</td>
</tr>
<tr>
<td>Lining</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>100% synthetic</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Interlining</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>100% synthetic</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Length</td>
<td>long</td>
<td>long</td>
<td>long</td>
<td>long</td>
<td>long</td>
<td>short</td>
<td>short</td>
<td>short</td>
<td>short</td>
</tr>
<tr>
<td>Colour</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- **ECP (European Comparison Programme):**
- **Personal appearance 2012:**
- **Working list:** Validation List (EU)
- **11.03.12.1.01 Men's coats and jackets:**
- **Task Report:**
- **Search:**
## Item List Management Tool

### Task Report

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.01.1.1.1.01.ab</td>
<td>Long-grain rice, not parboiled, WKB</td>
</tr>
<tr>
<td>A.01.1.1.1.01.ad</td>
<td>Long-grain rice, parboiled, in cooking bags, WKB</td>
</tr>
<tr>
<td>A.01.1.1.1.01.ca</td>
<td>Basmati rice, WKB</td>
</tr>
<tr>
<td>Q.01.1.1.3</td>
<td>Bread</td>
</tr>
<tr>
<td>A.01.1.1.3.01</td>
<td>Bread</td>
</tr>
<tr>
<td>A.01.1.3.01.bb1</td>
<td>Roll, multicorn, multipack, industrially packed, WKB</td>
</tr>
<tr>
<td>A.01.1.3.01.cc</td>
<td>Bread, white, industrially packed, WKB</td>
</tr>
<tr>
<td>A.01.1.3.01.ce1</td>
<td>Bread, white toast, large pack, BL</td>
</tr>
<tr>
<td>A.01.1.3.01.gb</td>
<td>Rye portion bread, industrially packed, WKB</td>
</tr>
<tr>
<td>A.01.1.3.01.ia</td>
<td>Pita bread, industrially packed, WKB</td>
</tr>
<tr>
<td>A.01.1.4.01.ea</td>
<td>Biscuits, salted crackers, TUC</td>
</tr>
<tr>
<td>A.01.1.4.01.eb</td>
<td>Biscuits, salted crackers, WKB</td>
</tr>
<tr>
<td>A.01.1.5.01</td>
<td>Pizza and quiche</td>
</tr>
<tr>
<td>A.01.1.5.01.ic</td>
<td>Pizza, frozen, BL</td>
</tr>
<tr>
<td>A.01.1.6.01.aa</td>
<td>Spaghetti, SB</td>
</tr>
</tbody>
</table>

### Task Report (continued)

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU</td>
<td>Long-grain rice, not parboiled, WKB</td>
</tr>
<tr>
<td>DE</td>
<td>Long-grain rice, not parboiled, WKB</td>
</tr>
<tr>
<td>EL</td>
<td>Long-grain rice, not parboiled, WKB</td>
</tr>
<tr>
<td>EU</td>
<td>Long-grain rice, parboiled, in cooking bags, WKB</td>
</tr>
</tbody>
</table>

1) In 2012 all countries could price WKB parboiled rice. Is non-parboiled WKB rice still necessary for the item list? 2) Just wondering whether parboiled is a standard feature for WKB level. Should we drop out mentioning parboiled/not parboiled in the definition, but simply have a WKB long-grain rice product?

1) No. 2) Yes we agree drop it.

1) Yes 2) We prefer to keep mentioning parboiled/not parboiled in the definition.

1) Is rice sold in cooking bags typical in your country? 2) If yes, is there a price difference compared to rice sold simipl in
Item List Management Tool
(survey catalogue preview)
Collaborative data validation in an iterative way

1. Data prepared in DET
2. Data transmitted & uploaded to the DB
3. Initial validation checks in VT
4. Initiate new Validation round

Countries recoded price data and validate their own data

Survey coordinators check for entry errors and review the thoroughness of intra-country validation carried out by countries

Detect possible errors for each country (intra-country validation)

Questions inserted to problematic items or observations

Respond to questions

Correct data, if necessary

Detect outliers using inter-country validation

Questions inserted to problematic items or observations

Respond to questions / correct data, if necessary

Quaranta Tables created

Data transmitted & uploaded to the DB

Initial validation checks in VT

Initiate new Validation round
Intra-country validation
Inter-country validation, via ad-hoc personalized Quaranta Tables (incl. ad hoc number of countries and basic headings)
Update existing observations and/or insert new observations for an item
View previous edits of the same observation (history is retained)
Explore data through a number of alternative views, with filtering capabilities and possibility to export in excel
Export reports (validation round overview, changes made since initial upload, task report etc.)
Respond to questions added by the survey coordinators (questions can be at observation, item or dataset level)
Access to average prices of other countries
Approve their data when validation is complete
Intra-country validation

Inter-country validation, via ad-hoc personalized Quaranta Tables (incl. ad hoc number of countries and basic headings)

Add questions to observation or items where outliers are detected

Explore data through a number of alternative views, with filtering capabilities and possibility to export in excel

Monitor the validation process using the task report facility
Intra-country validation

Inter-country validation, via ad-hoc or default (official) Quaranta tables

Calculation of annual national average prices using Temporal and Seasonal Adjustment factors in consumer goods and services surveys

PPP Gap filling & Fixity

Initiate new validation rounds

Finalize validation
### Validation Tool

![Validation Tool Screenshot](image-url)
The data entry tool is designed to be used mainly by national agencies.

It operates in two modes (price collector (PC) or national coordinator (NC)):
- National coordinator splits an item list (creates partial lists) and shares it to different price collectors.
- Each price collector, working in PC mode, enters price data for a set of BH items.
- The national coordinator can merge the datasets created by the different price collectors into a single national dataset, validate the dataset and send it.

It is both a data entry tool and validation tool (intra-country).

It works off-line.

In what regards the intra-country validation has the same functionalities as the Validation Tool.

For ICP could be used for regional data and/or countries data.
## Data Entry Tool

### Survey Tree

```
- 11.03.11.1 - Clothing materials
  - 11.03.12.1 - Men's clothing
    - 11.03.12.1.1 - Men's coats and jackets
    - 11.03.12.1.2 - Men's suits (two-piece)
    - 11.03.12.1.3 - Men's trousers
  - 11.03.12.2 - Women's clothing
    - 11.03.12.2.1 - Ladies' coats and jackets
    - 11.03.12.2.2 - Ladies' suits (two-piece)
    - 11.03.12.2.3 - Ladies' trousers
  - 11.03.12.4 - Ladies' skirts and dresses
  - 11.03.12.5 - Ladies' blouses, shirts, sweaters, etc.
  - 11.03.12.6 - Ladies' briefs and underwear
  - 11.03.12.7 - Brassieres
    - 11.03.12.7.1 - Brassieres, T-shirts, etc.
    - 11.03.12.7.2 - Brassieres, dress shoes
    - 11.03.12.7.3 - Brassieres, panty hose
    - 11.03.12.7.4 - Brassieres, bras
  - 11.03.12.8 - Tights and anklets
  - 11.03.12.9 - Children's and infant's clothing
  - 11.03.13.1 - Other articles of clothing and clothing accessories
  - 11.03.14.1 - Cleaning, repair and hire of clothing
  - 11.03.15.1 - Men's footwear
  - 11.03.21.2 - Women's footwear
  - 11.03.21.3 - Children's and infant's footwear
  - 11.12.11.1 - Hairdressing salons and personal care
  - 11.12.21.1 - Electric appliances for personal use
  - 11.12.31.1 - Jewellery, clocks and watches
  - 11.12.32.1 - Other personal effects
```

### Summary

<table>
<thead>
<tr>
<th>BH code</th>
<th>BH name</th>
<th>Item count</th>
<th>Priced items</th>
<th>Items with discou...</th>
<th>Items with ?</th>
<th>Seasonal items</th>
<th>No of obs. w/o discou...</th>
<th>No of obs. with discou...</th>
<th>Items with ?</th>
<th>Items with ??</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.03.11.1</td>
<td>Clothing materials</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>12</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.03.12.1</td>
<td>Men's clothing</td>
<td>76</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.03.12.2</td>
<td>Women's clothing</td>
<td>82</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.03.12.3</td>
<td>Children's and infant's clothing</td>
<td>55</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.03.13.1</td>
<td>Other articles of clothing and clothing accessories</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.03.14.1</td>
<td>Cleaning, repair and hire of clothing</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.03.21.1</td>
<td>Men's footwear</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.03.21.2</td>
<td>Women's footwear</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.03.21.3</td>
<td>Children's and infant's footwear</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.12.11.1</td>
<td>Hairdressing salons and personal care</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.12.21.1</td>
<td>Electric appliances for personal use</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.12.31.1</td>
<td>Jewellery, clocks and watches</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11.12.32.1</td>
<td>Other personal effects</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>455</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
ILMT adapted for ICP at region level

Countries could be involved and a full scenario

DET can be adapted to support the collection and intra-country validation of regional data (average prices, weights, population etc.)

Potential uses of Validation tool:

- Validation of regional data
- Validation of detailed country data (price observations)