INTERNATIONAL COLLABORATION USE CASE: THE OECD’S STATISTICAL INFORMATION SYSTEM COLLABORATION COMMUNITY

HLG, 25-27 November 2013
By Jonathan Challener
Community members

OECD

AUSTRALIAN BUREAU OF STATISTICS
EUROPEAN COMMISSION
INTERNATIONAL MONETARY FUND
ITALIAN NATIONAL INSTITUTE OF STATISTICS
NATIONAL BANK OF BELGIUM*

STATISTICS ESTONIA
STATISTICS NEW ZEALAND
UNESCO INSTITUTE FOR STATISTICS
UNIVERSITY OF MANCHESTER

* Pending signature of MoU
Interested organisations

INSEE    Statistics Slovenia    Central Statistics Office Ireland    Tunisian Statistics Institute

Asian Development Bank    UNESCWA

"Rich in functionalities and easy to use"
"Built for ENDUSERS"
"Very good process automation options"
"Modern user interface and functionality"
"Good perspective"
"Develop our own tool will take a long time and a lot of resources"
"Other similar available tools seem to no longer be improved or have blocking limitations"
"To be part of a collaboration community"
Community vision

Provide an international collaboration framework for a more open, more innovative, and more industrialised data dissemination to collectively develop software, leverage innovations, mutualise costs, and promote standardisation.
Community Objectives

Collectively produce and develop software, by leveraging on the OECD.Stat platform and related components, and in so doing build a robust, component-based and scalable architecture.

Share experiences, knowledge and best practices through multilateral collaboration and building of a collective capacity, concerning the Community.

Contribute to International Collaboration, by accelerating the implementation of standards and contributing to the international ‘Plug and Play’ architecture vision.
Community framework

Strategic 5 year directions

MOU
Grants
SoW
Governance framework
Product roadmap
Architecture vision
Biennial/Annual work plan
ALM and QA processes
Central to collaboration

Stat is the central repository ("warehouse") of validated statistics and related metadata.

Stat is the central hub connecting data production, sharing & dissemination processes.

It is the corporate source of data for data sharing and dissemination purposes.
Central hub

DATA DELIVERY

DATA PRODUCTION

DATA DISSEMINATION

INTERNAL DATA SHARING
Key functions

Internal Data Sharing enabling data consolidation moving away from a fragmented ‘siloh’ environment towards a coherent and strategic approach to data.

Machine to Machine Data Exchange making statistical data and metadata “open” via machine-readable formats, in particular the Statistical Data and Metadata standard eXchange format (SDMX).

Data Exploration through a feature-rich data browser providing users the ability to explore and customise data and metadata, visualise it with dynamic charts and download it for further analysis offline.

Streamlined Data Dissemination Processes enabling to structure statistical data dissemination processes and integrate a wide range of data production tools.

International Standards Setting providing a means for disseminating data in the SDMX and more recently the first working API using the common statistical JSON format as defined by the SDMX Technical Working Group.
International context
Part of the Overarching International Collaboration Framework

- CES
- HLG
- SIS-CC
- EuroStat
- SWG
- TWG
- SDMX-RI
- MSIS
- GSIM
- GSBPM
- CSPA
- SAB
- Statistical Network
- SDMX Initiative

SIS-CC
International platform for dissemination
**GSBPM v4.0 mapping**

<table>
<thead>
<tr>
<th>Quality Management / Metadata Management</th>
<th>Disseminate</th>
<th>Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Specify Needs</td>
<td>7.1 Update output systems</td>
<td>8.1 Define archive rules</td>
</tr>
<tr>
<td>1.1 Determine needs for information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 Consult &amp; confirm needs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3 Establish output objectives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4 Identify concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5 Check data availability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6 Prepare business case</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Design</td>
<td>7.2 Produce dissemination products</td>
<td>8.2 Manage archive repository</td>
</tr>
<tr>
<td>2.1 Design outputs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2 Design variable descriptions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3 Design data collection methodology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4 Design frame &amp; sample methodology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 Design statistical processing method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Build</td>
<td>7.3 Manage release of dissemination products</td>
<td>8.3 Preserve data and associated metadata</td>
</tr>
<tr>
<td>3.1 Build data collection instrument</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2 Build or enhance process components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3 Configure workflows</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.4 Test production system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5 Test statistical business process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.6 Finalize production system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Collect</td>
<td>7.4 Promote dissemination products</td>
<td>8.4 Dispose of data &amp; associated metadata</td>
</tr>
<tr>
<td>4.1 Select sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.2 Set up collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.3 Run collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4 Finalize collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.1 Integrate data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.2 Classify &amp; code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.3 Review, Validate &amp; edit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.4 Impute</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5 Derive new variables &amp; statistical units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.6 Calculate weights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.7 Calculate aggregates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.8 Finalize data files</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Analyse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.1 Prepare draft outputs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.2 Validate outputs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.3 Scrutinize &amp; explain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.4 Apply disclosure control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.5 Finalize outputs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.6 Scrutinize &amp; explain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Disseminate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.1 Update output systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.2 Produce dissemination products</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.3 Manage release of dissemination products</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.4 Promote dissemination products</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.5 Manage user support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Archive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.1 Define archive rules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.2 Manage archive repository</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.3 Preserve data and associated metadata</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.4 Dispose of data &amp; associated metadata</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Stat contributes to Planned additions*
GSBPM v5.0 mapping

Archive to be incorporated into the over-arching process of data and metadata management
GSIM v1.0 mapping
GSIM v1.0 mapping

DATASET

- Data Set
  - Fact tables + Dimensions + supporting tables (filter, etc.)

- Data Structure
  - A dataset’s dimensions, and data attributes. A dataset’s Data Structure is tied directly to a single dataset and is not shareable
  - Measure dimension is not explicitly defined by the user. It is usually called Measure/Subject/Indicator/Transaction dimension

- Data Point
  - Obs. key w/o Measure.
  - Obs. key + datum + control code + other attributes, e.g. timestamp
  - Has observation of

- Identifier Component
  - Has subtype of

- Measure Component
  - Has subtype of

- Attribute Component
  - Control codes and timestamp are the attributes. Control codes are at observation level only.

Unit

- measures

Datum

- A data value w/o attributes
- describes

Instance Variable

- Obs. key
- uses

Represented Variable

No. Stat equivalent
GSIM v1.0 mapping findings

Forces an analysis of the system which can lead to identifying improvements

Gaps/equivalences can be seen when compared to other systems/standards/methodologies/processes

“GSIM-lite” could speed-up mapping and make for greater understanding

Include an example business process and system mapping in documentation

GSIM mapping registry would be compelling to see examples of how other systems/processes have been mapped as a starting point

For more information contact: David.BARRACLOUGH@oecd.org
Standard setting and adoption

**SDMX Reference Infrastructure integration** mapping of multiple data sources to a local source and supporting 2.1 standard

**SDMX-JSON API** providing an easy to use format for developers to reuse statistical data

**Globally agreed DSDs** future support through SDMX artefacts

**SDMX consumption** providing support for both SDMX import and export
CSPA “Plug-and-play”

**Decision Principles** aligned to the SIS-CC strategic directions

**Design Principles** at the core of OECD.Stat and SIS-CC work

**Requirements** aligned to GSBPM, work underway to map to GSIM, and current architecture evolutions will strengthen alignment to CSPA
Plug-and-play “disseminate” candidate?
Thank you

Questions?

Email:

jonathan.challener@oecd.org

or

dotstatcollaborationsupport@oecd.org