10th Expert Group Meeting on Statistical Data and Metadata Exchange (SDMX)
January 25-28

“Hub of Public Statistics”

Francesco Rizzo – ISTAT, Italy
rizzo@istat.it
Content

- Italian National Statistical System (Sistan)
- Istat role within Sistan
- Hub of the public statistics (a.k.a. Sistan Hub) project:
  - main objectives
  - Implementation strategy
- Free and open source toolkit
- Hub architecture
- Statistical yearbooks: Real, Linked and Virtual DFs
- Single access point for the SISTAN HUB project
- Conclusion
- References
National Statistical System: SISTAN

Structure

- National Institute of Statistics – Istat (Central and regional offices)
- **Statistical offices** of central (*) and local government administrations
- Other public bodies and organisations dealing with statistical information

Main tasks of Istat within SISTAN

- Drafts the annual National Statistical Programme
- Coordinates, promotes and provides technical assistance and training to other bodies
- Sets nomenclatures, standards and methodologies
- Publishes and disseminates data also in collaboration with other SISTAN bodies

(*) included ONAs: Other National Agencies
National Statistical System: SISTAN

Structure

- National Institute of Statistics – Istat (Central and regional offices)
- **Statistical offices** of central (*) and local government administrations
- Other public bodies and organisations dealing with statistical information

Main tasks of Istat within SISTAN

- Coordinates, promotes and provides technical assistance and training to other bodies
- Sets nomenclatures, standards and methodologies
- Publishes and disseminates data also in collaboration with other SISTAN bodies

(*) included ONAs: Other National Agencies

Improving the quality in collecting, producing and disseminating statistics
SISTAN Hub project – main objectives

- **Improve data dissemination quality within the NSS**
  - Applying suitable dimensions of the European Code of practice / Quality Assurance Framework
  - Facilitating open (statistical) data at national level
  - Supporting the “semantic interoperability” as detailed by the Digital Agenda at European and National level

Facilitate the modernization of the dissemination information systems within the SISTAN organisations

- Standardization: International and “open” standards (SDMX, DCAT, RDF, etc.)
- Industrialisation: Metadata-drive processes and data managed in suitable databases
1) Identify concrete benefits for all the involved actors

- **ISTAT** → improve data dissemination quality
- **SISTAN Organizations** → reuse harmonized datasets for their own needs; reuse free and open source software
- **USERS** → single access point for browsing datasets produced by different organizations

2) Provide suitable solutions in order to involve as many organizations as possible

- Access to datasets that can be easily reused (e.g. for building statistical yearbooks)
- Statistical and IT consultancy (including capacity building actions)
- Free and open source Toolkit / Software as a Service
Free and open source toolkit (1/2)
Few and Clear designing principles:

- Easy to install and configure
- Business-driven
- (when possible) Reusing open software developed by others (SDMX-RI)
- Finding a good balance between componentization and simplicity
- Certainty about deadlines
- All the design decisions must follow a pragmatic approach

FREE AND OPEN SOURCE TOOLKIT (2/2)

Data Source transformation

Meta Manager

Data Management

Data and Metadata Dissemination

Reporting

Excel2CSV

Meta Manager

Data Manager

Data Browser
SDMX API
CKAN API
Hub architecture

Hub node

Satellite node

SDMX API

DDB

Satellite node

SDMX API

DDB

Satellite node

SDMX API

DDB

Satellite node

SDMX API

DDB

Industrial production index - base 2010

Frequency: Monthly, Indicator: Industrial production index

Adjustment: calendar adjusted data

<table>
<thead>
<tr>
<th>Economic Activity</th>
<th>Consumer goods</th>
<th>Capital goods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time period</td>
<td>2019-08</td>
<td>2019-09</td>
</tr>
<tr>
<td></td>
<td>112,6</td>
<td>115,4</td>
</tr>
<tr>
<td></td>
<td>115,3</td>
<td>120,6</td>
</tr>
<tr>
<td></td>
<td>111,5</td>
<td>117,3</td>
</tr>
<tr>
<td></td>
<td>75,9</td>
<td>62,7</td>
</tr>
</tbody>
</table>

RDF

XLS

CSV

PDF

Istat
Regional statistical yearbooks: Real, Linked and Virtual DFs

- **Real (Physical) DF**
- **Linked DF**
- **Virtual DF**

- Regional local data 30%
- Other Central Administrative...
- Regional slices on Istat data 50%

Format options: RDF, CSV, XLS, PDF
Single access point for the SISTAN HUB project
Conclusions

- SDMX can be used successfully beyond the data reporting/data collection use cases
- SDMX is a good candidate for supporting interoperability and interconnection among distributed statistical dissemination systems
- SDMX should evolve to support better the dissemination use case (SDMX 3.0 will facilitate these evolution but other dissemination issues must be taken in consideration)
- Hub architectures using the “pull” mode will be used more and more in the future in order to avoid data duplication, reduce the data transmission burden and facilitate automatic content reusing
- SDMX implementations can be speeded if business-driven tools will be increasingly available

Annotations can help in overcoming some SDMX deficiencies in dissemination architectures.

Some Organisations, that are developing SDMX tools, have already been shared the syntax and the semantic of a set of annotations

Why not an SDMX Annotations guideline?
Conclusions

- SDMX can be used successfully beyond the data reporting/data collection use cases
- SDMX is a good candidate for supporting interoperability and interconnection among distributed statistical dissemination systems
- SDMX should evolve to support better the dissemination use case (SDMX 3.0 will facilitate these evolution but other dissemination issues must be taken in consideration)
- Hub architectures using the “pull” mode will be used more and more in the future in order to avoid data duplication, reduce the data transmission burden and facilitate automatic content reusing
- SDMX implementations can be speeded if business-driven tools will be increasingly available
References

- Sistan Hub:
  http://sistanhub.istat.it/hub/

- Data Browser for the “Permanent census of population and housing”
  https://esploradati.censimentopopolazione.istat.it/databrowser/#/en

- SDMX Istat toolkit download:
  https://sdmxistattoolkit.github.io/index.html
  https://github.com/sdmxistattoolkit

- Data Browser demonstration website
  http://demo.databrowser.sister.it/

- More information:
  rizzo@istat.it (Francesco Rizzo)
  alcardac@istat.it (Alessio Cardacino)
thank you