

# 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*

- ❑ 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



Main  
actors

## 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



Main actors

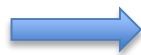


## Structure

- National Institute of Statistics – Istat (Central and regional offices)
- **Statistical offices** of central (\*) and local

**Improving the quality in collecting, producing and disseminating statistics**

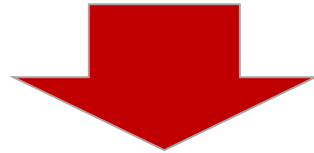
Main tasks of Istat within SISTAN



- 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

- ❑ **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)

The diagram illustrates the data pipeline. On the left, a spreadsheet shows a table with columns for 'Year' and 'Country' and rows for 'DSDCOP Basic Index', 'Real and non-financial', 'Manufacturing', 'Construction', 'Wholesale', 'Retail trade', 'Transport and storage', 'Information and communication', 'Energy supply', and 'Household consumption'. A blue arrow points from the spreadsheet to a central window titled 'M&D Manager' showing 'Data Structure Definitions'. Another blue arrow points from this window to a table on the right containing 16 rows of metadata, including frequency, indicator, and territory information.

The screenshot shows the M&D Manager v1.5.1 interface. At the top, there are navigation icons for MM, DM, RM, and a settings gear. The main area is titled 'Data Structure Definitions' and includes a search bar, 'Import' and 'New' buttons, and a table of definitions. The table has columns for 'ID', 'Ag.', 'Name', 'Vers.', and 'Final'. The definitions listed are:

ID	Ag.	Name	Vers.	Final
DCSP_PRICES	IT1	Consumer Prices	1.0	✓
DSD_INDUSTRY	IT1	Industrial new orders and turnover monthly data	1.0	✓
DSD_SDG DISS	IT1	SDG for dissemination	1.0	✓
DSD_SDG DISS	IT1	SDG for dissemination	1.1	✓
DSD_SDG DISS	IT1	SDG for dissemination	1.2	✓
		Public water supply use	1.0	✓
		Mining and quarrying	1.0	✓
		Households energy consumption	1.0	✓

The screenshot shows the MULTIDOMAIN statkit interface. The main title is 'Industrial production index - base 2015=100 (\*)'. Below the title, it specifies 'Frequency: Monthly, Indicator: Industrial production index - base 2015=100, Territory: Italy'. The interface includes a search bar, a language selector (UK flag), and a 'LOGIN' button.

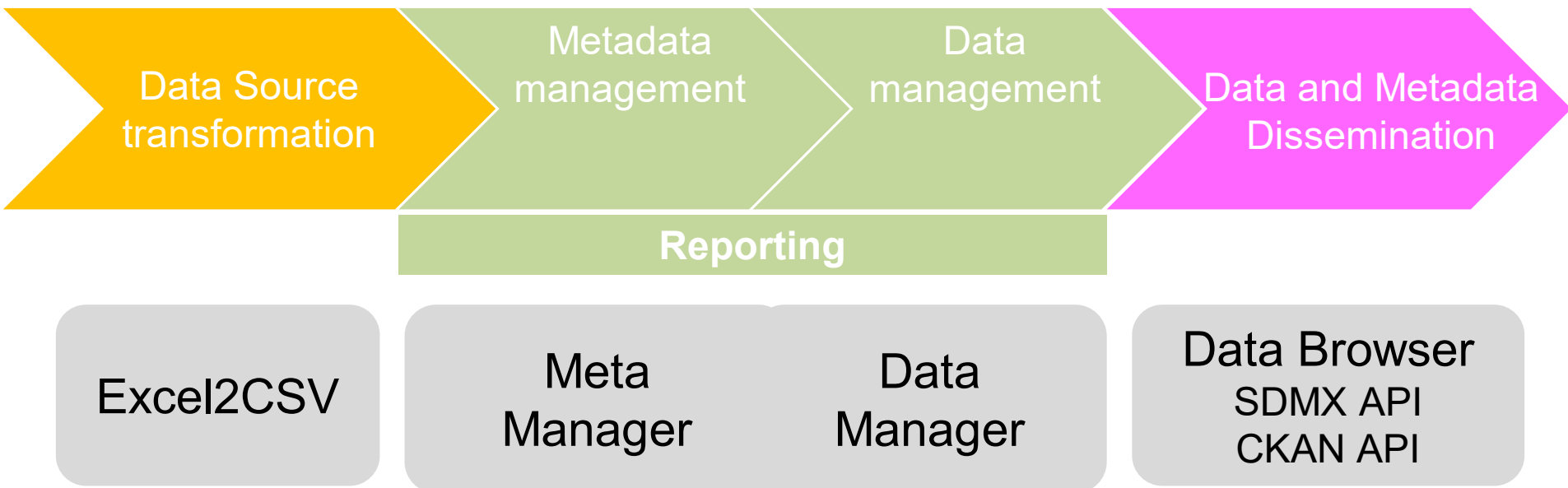
The screenshot shows the data table in the MULTIDOMAIN interface. The table is titled 'calendar adjusted data' and has columns for 'Economic Activity', 'Consumer goods', 'Capital goods', 'Intermediate goods', and 'TOTAL INDUSTRY EXCLUDING CONSTRUCTION (b to e)'. The rows represent time periods from 2019-08 to 2020-05.

Economic Activity	Consumer goods	Capital goods	Intermediate goods	TOTAL INDUSTRY EXCLUDING CONSTRUCTION (b to e)
2019-08	75,9	62,7	59,8	70,1
2019-09	111,5	117,3	108,8	110,7
2019-10	115,3	120,6	112,2	113,4
2019-11	112,6	115,4	105,9	109,3
2019-12	89,9	94	80,7	89,4
2020-01	100,7	102,2	97	100,7
2020-02	102,8	115,8	104,1	105,3
2020-03	83,4	76,9	82,7	81,9
2020-04	58,7	51,5	54,3	57,8
2020-05	91	96,1	88,1	90,5

# Free and open source toolkit (2/2)

Few and Clear designing principles:

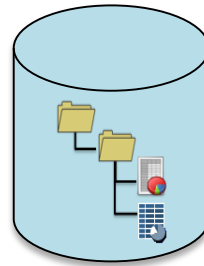
- ❑ Easy to install and configure
- ❑ Business-driven
- ❑ (when possible) Reusing open software developed by others (SDMX-RI)
- ❑ Finding a good balance between componentization and simplicity
- ❑ Certainty about deadlines
- ❑ All the design decisions must follow a pragmatic approach





# Hub architecture

## Hub node



SDMX  
API

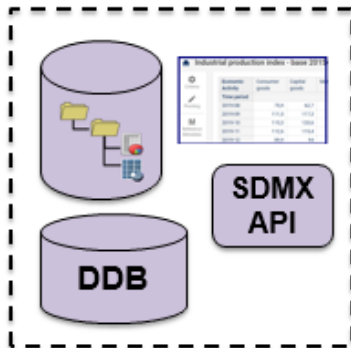
Industrial production index - base 201...

Frequency: Monthly, Indicator: Industrial production in...

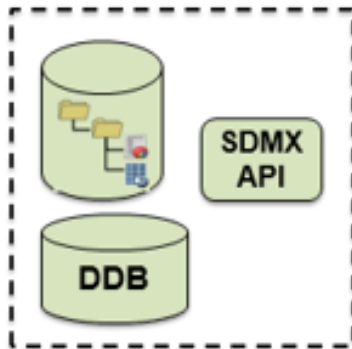
Adjustment  
calendar adjusted data

Economic Activity	Consumer goods	Capital goods
Time period		
2019-08	75,9	62,7
2019-09	111,5	117,3
2019-10	115,3	120,6
2019-11	112,6	115,4
2019-12	89,9	94
2020-01	100,7	102,2

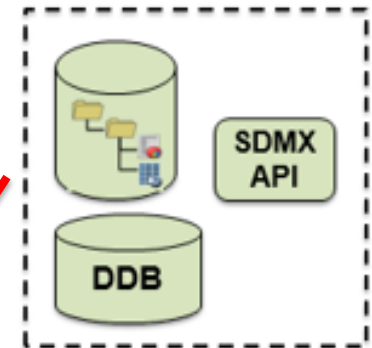
## Satellite node



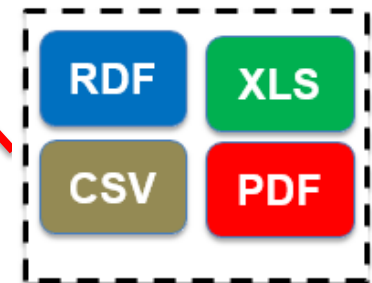
## Satellite node



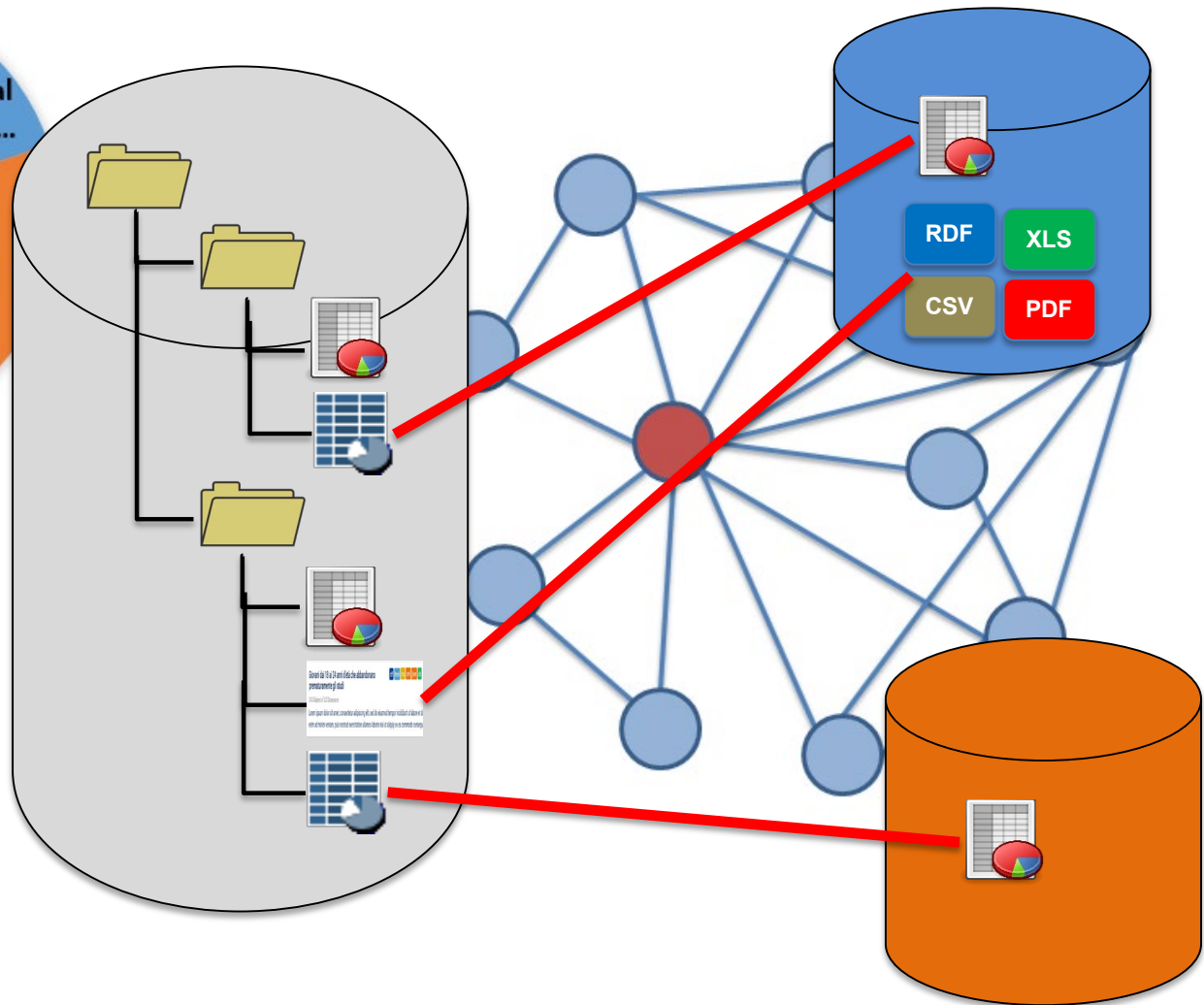
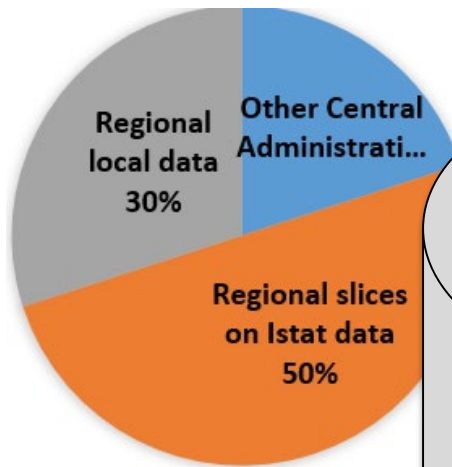
## Satellite node



## Satellite node



# Regional statistical yearbooks: Real, Linked and Virtual DFs



Real (Physical) DF



Linked DF



Virtual DF

# Single access point for the SISTAN HUB project

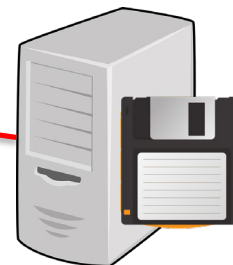
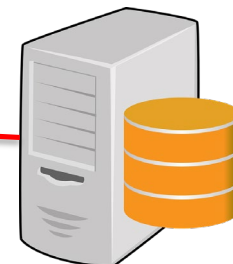
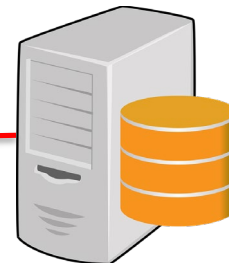



**Themes**

- Population and households
  - + Population
  - + Foreigners and immigrants
- + Environment and energy
- + Health statistics
- + Enterprises
- + Services
- Prices
  - + Consumer Prices
- Labour
  - + Employment
  - + Unemployment
- + Social security and welfare
- + Education and training
- + Justice and security

**Territory**

- + Lombardia
- + Veneto
- + Sicilia



# 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

**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?**

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

# 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](mailto:rizzo@istat.it) (Francesco Rizzo)  
[alcardac@istat.it](mailto:alcardac@istat.it) (Alessio Cardacino)

