Istat | Istituto Nazionale di Statistica

SDMX Global Conference 2023

Kingdom of Bahrain

29 October 2023 – 02 November 2023

Data engineering for moving out from legacy systems to SDMX compliant systems: the ISTAT experience

Francesco Rizzo (SDMX dissemination platform – project manager), Carlo Boselli (Dissemination database initiative - coordinator), Alessio Cardacino (SDMX toolkit development initiative – coordinator)

Summary

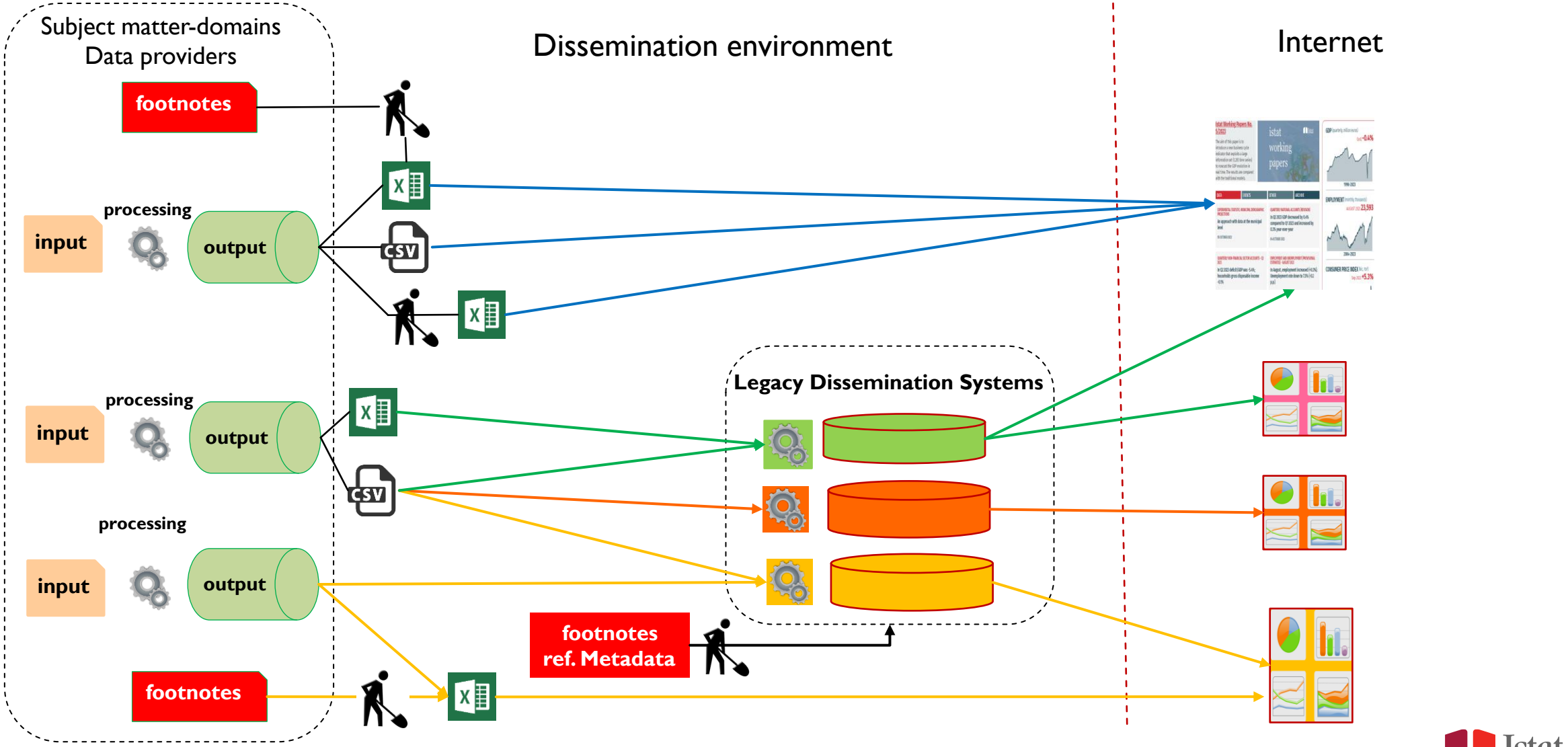
The presentation summarizes the lesson learnt on migration processes, from legacy systems to SDMX information architectures

- Migration processes of dissemination/reporting systems – some experiences
  - Legacy scenarios - synthesis of a nightmare
- Figures of the Istat migration process (2022 – 2023)
- Migration Maturity Model (The Drivers of the systems migrations, complexities and main barriers)
- Main inescapable constraints to consider and solve
- Enhancements of the IT platform for reducing complexities and facilitating the overcome of the barriers
- Lesson learnt

Migration processes of dissemination/reporting systems – some experiences

- ISTAT Italy (Dissemination and reporting)
- NSI Tunisia (MDM as back-office of the new dissemination system)
- FAO (MDM as back-office of the new dissemination system)
- SSO Macedonia (Data reporting to Eurostat and IMF SDDS plus)
- ASK Kosovo (Data reporting to Eurostat)
- BHAS Bosnia (Data reporting to Eurostat)
- CARICOM (data warehouse for dissemination)
- African Development Bank (Evolution of African Information Highway)
- National Bank of Chile (Pilot exercise for evaluation)
- World Bank (Pilot exercise for evaluation)

Legacy scenarios – synthesis of a nightmare



Figures of the Istat migration process (2022 – 2023)

- Corporate data warehouse (IstatData):
  - 2 billions of data records spread on 450 data cubes, disseminated through 2900 dataflows
- External trade data warehouse (CoeWeb):
  - 10 billions of data records spread on 45 data cubes, disseminated through 78 Data flows
- Permanent Population and household Census
  - 700 millions of data records spread on 25 data cubes, disseminated through 564 dataflows

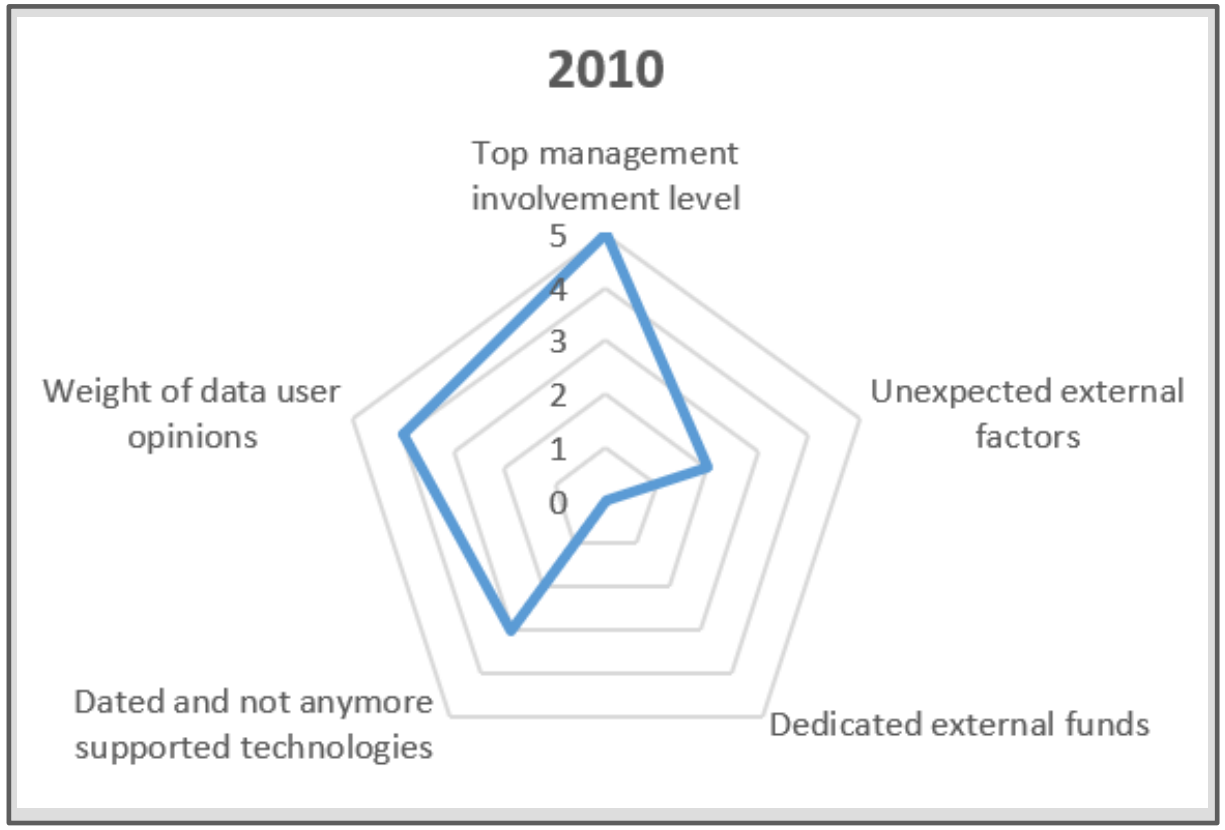
Migration Maturity Model

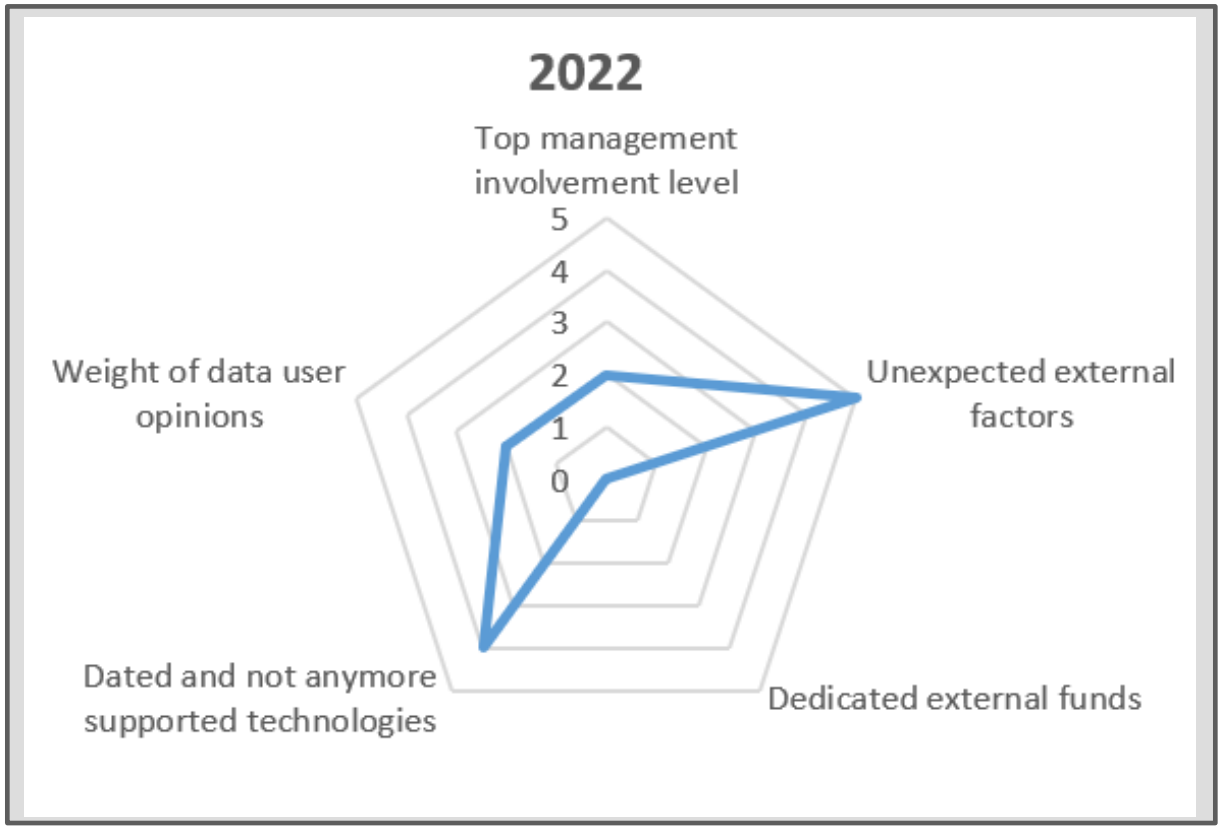
- The Drivers of the migration process
- Complexities and main barriers

The Drivers of the migration process

- Top management modernization strategy
- Unexpected external factors (including new National or International regulations)
- Dedicated external funds (grants, twinning, etc.)
- Dated and not anymore supported technologies (including security aspects)
- Weight of data user opinions

Drivers impact level in the migration of the dissemination legacy systems in Istat





Top management involvement level	5
Unexpected external factors	2
Dedicated external funds	0
Dated and not anymore supported technologies	3
Weight of data user opinions	4

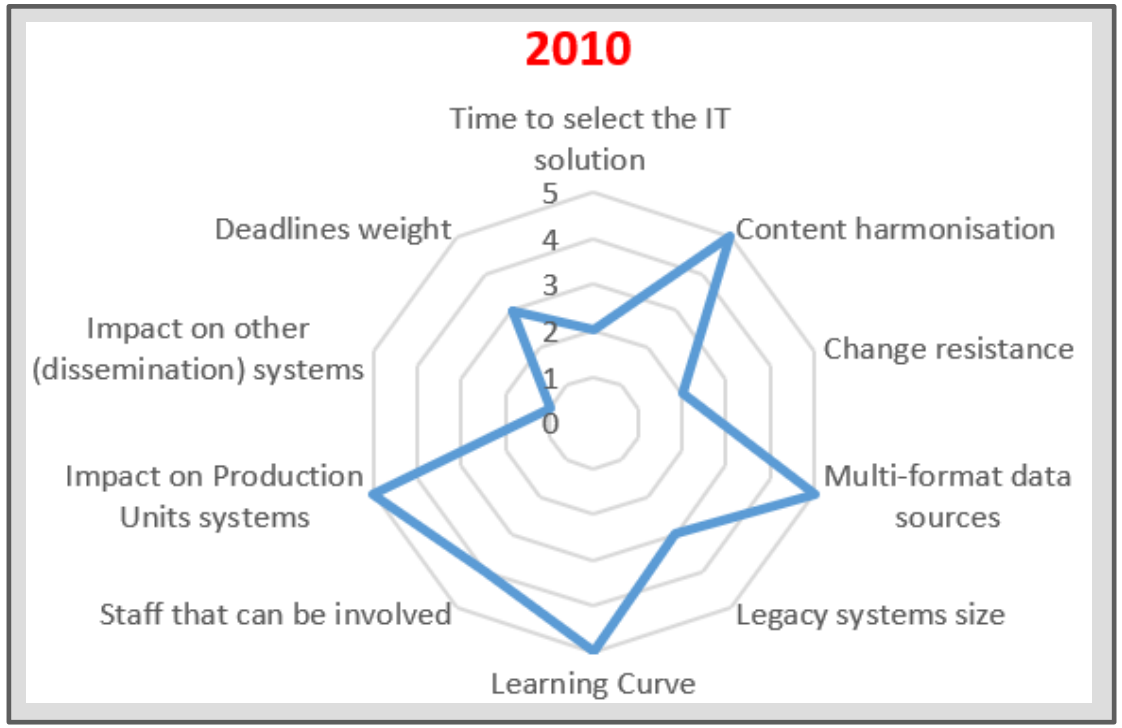
Top management involvement level	2
Unexpected external factors	5
Dedicated external funds	0
Dated and not anymore supported technologies	4
Weight of data user opinions	2

Complexities and main barriers

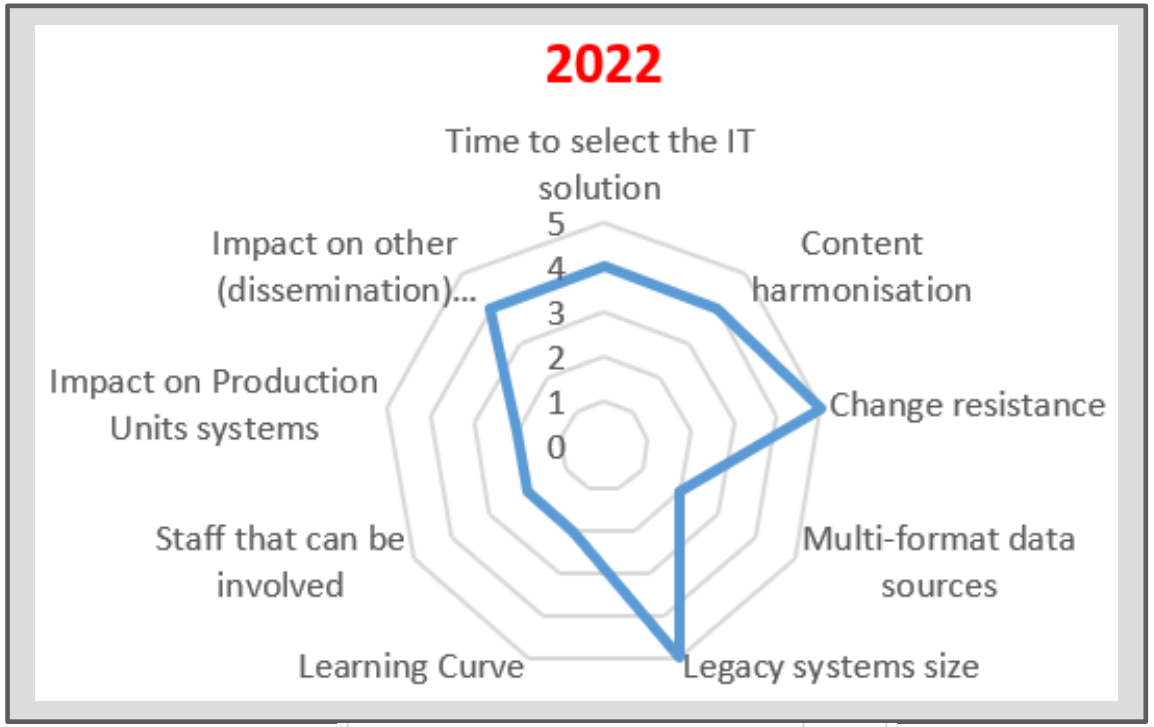
- Time to select the IT solution (in-house Vs market solution)
- Content harmonisation (including DSDs definition)
- Change resistance
- Multi-format data dissemination (Excel files, Pdf statistical yearly books, multi-technologies databases, SAS, etc.)
- Legacy systems size (number of datasets, indicators, data points)
- Learning Curve
- Staff (IT, dissemination experts, etc.) that can be involved
- Impact on Production Units systems
- Impact on other (dissemination) systems
- Deadlines weight

Long transition period

Complexities and main barriers impact level in Istat



Time to select the IT solution	2
Content harmonisation	5
Change resistance	2
Multi-format data sources	5
Legacy systems size	3
Learning Curve	5
Staff that can be involved	4
Impact on Production Units systems	5
Impact on other (dissemination) systems	1
Deadlines weight	3



Time to select the IT solution	4
Content harmonisation	4
Change resistance	5
Multi-format data sources	2
Legacy systems size	5
Learning Curve	2
Staff that can be involved	2
Impact on Production Units systems	2
Impact on other (dissemination) systems	4
Deadlines weight	5

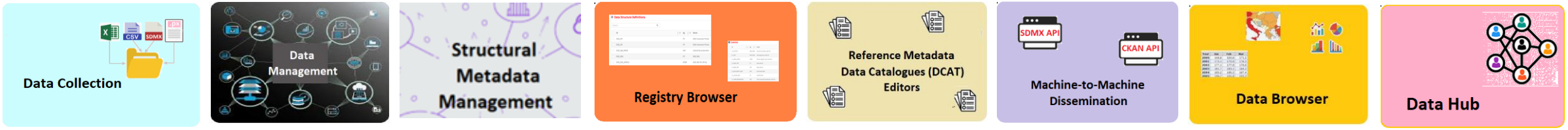
Main inescapable constraints to consider and solve

- Impact on Production Units systems
  - Data input format
  - Attribute (footnotes) management
- Performance
  - Data cube greater than 1K millions of data points
  - Sparse data cubes
- Impact on other (dissemination) systems
  - SDDS Plus
  - Data reporting to Eurostat
- SDMX complexity

Enhancements of the IT platform for reducing complexities and facilitating the overcome of the barriers

- Absorb the changes that Production Units would have been called upon to fulfill
- Implement the legacy systems features more appreciated by data users
- Ensure that performance will be at least the same or better of the legacy systems
- Provide a unique solution valid for both dissemination and reporting

Supporting the sustainability beyond pilot projects (SDMX Istat Toolkit)



NEW **Transformation**

NEW **Mapping handler**

NEW **Plugin Manager**

NEW **Data Browser caching**

Processing NEW

Easy loader NEW

Istat

Lesson learnt

- Before starting a migration, a Maturity Model should be considered
  - identify clearly the inescapable constraints
- A migration process is a long running activity
  - it is not sufficient planning the migration as such, but
  - the short and middle period must be (strongly) considered and planned to succeed shut downing the legacy systems
- The implementation of a new SDMX architecture adds more complexity and barriers
  - New capabilities (e.g., data modelling)
  - Biases (SDMX is too difficult)
- Collaborations, Capacity building actions and suitable software tools can help a lot

Thanks

- SDMX Istat toolkit download: https://sdmxistattoolkit.github.io/index.html
- Permanent census of population and housing: https://esploradati.censimentopopolazione.istat.it/databrowser/#/en
- Corporate dissemination data warehouse (IstatData): https://esploradati.istat.it/databrowser/#/en
- External trade dissemination system (CoeWeb): https://esploradati.istat.it/coeweb/databrowser/#/en

Francesco Rizzo | rizzo@istat.it
Carlo Boselli | carlo.boselli@istat.it
Alessio Cardacino | alessio.cardacino@istat.it

Istat | Istituto Nazionale di Statistica