ISWC 2014
Riva del Garda - Trentino, Italy
The 13th International Semantic Web Conference
19-23 October 2014

Smart Statistics: Data Analytics Services Based on Open Data Platforms

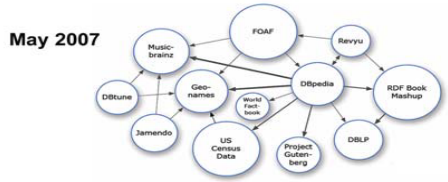
Emanuele Baldacci

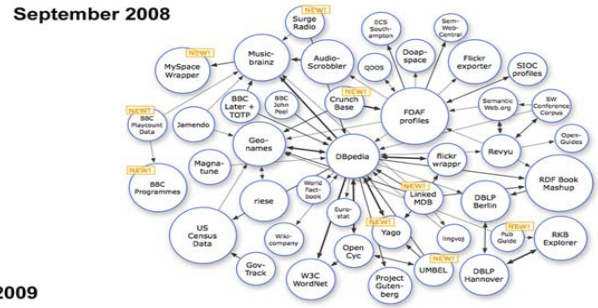
Italian National Institute of Statistics (Istat)

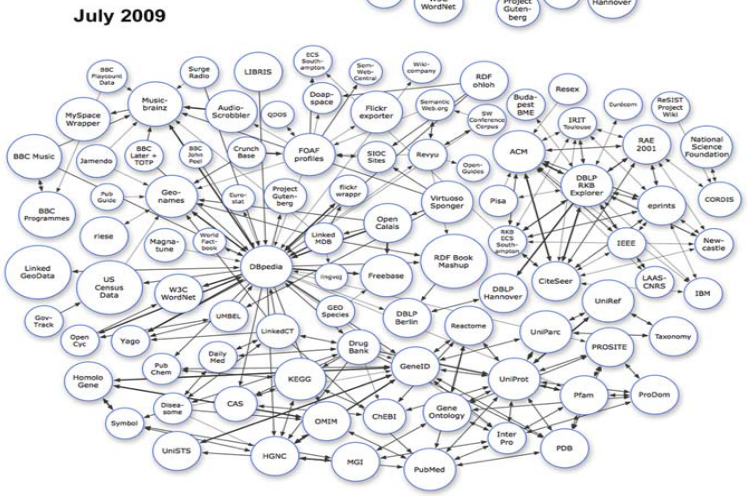
Head of Department for Integration, Quality, Research and Production Networks Development (DIQR)

Outline

- Times are changing
- European and international initiatives
- Modernisation strategy
- Putting users in the driver's seat
- Linked Open Data (LOD) in official statistics
- Future Challenges







Istat

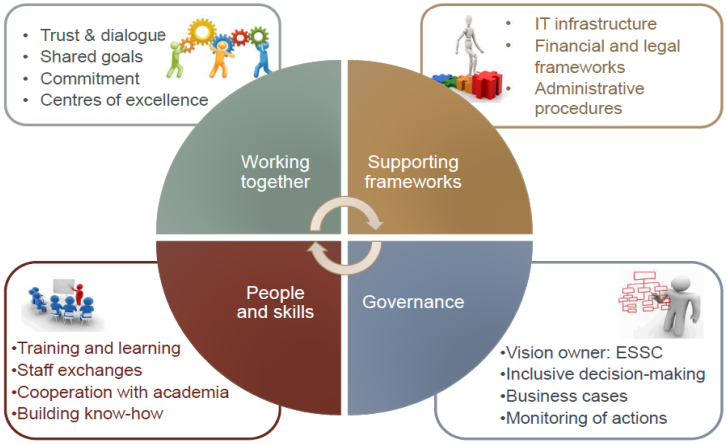
"*Times they are-a-changing*": drivers, threats and opportunities

- Budget constraints
- IT trends and growth of *prosumers*
- Scaled up demand for decision-oriented information
- Competitive pressures: data commoditisation
- Increasing quality concerns (e.g. fiscal data, timeliness)

ESS *Vision 2020*: main challenges and key areas of intervention

Main Challenges

- ✓ Data revolution
- ✓ New metrics
- ✓ Price of statistics
- ✓ Future of Europe



Key areas

- ✓ Identifying user needs and cooperation with stakeholders
- ✓ Enhancing quality of European statistics
- ✓ Harnessing new data sources, including Big Data for integrated data collection and processing
- ✓ Developing efficient and robust statistical processes based on process industrialisation
- ✓ Developing ESS common dissemination and communication systems

Web-COSI

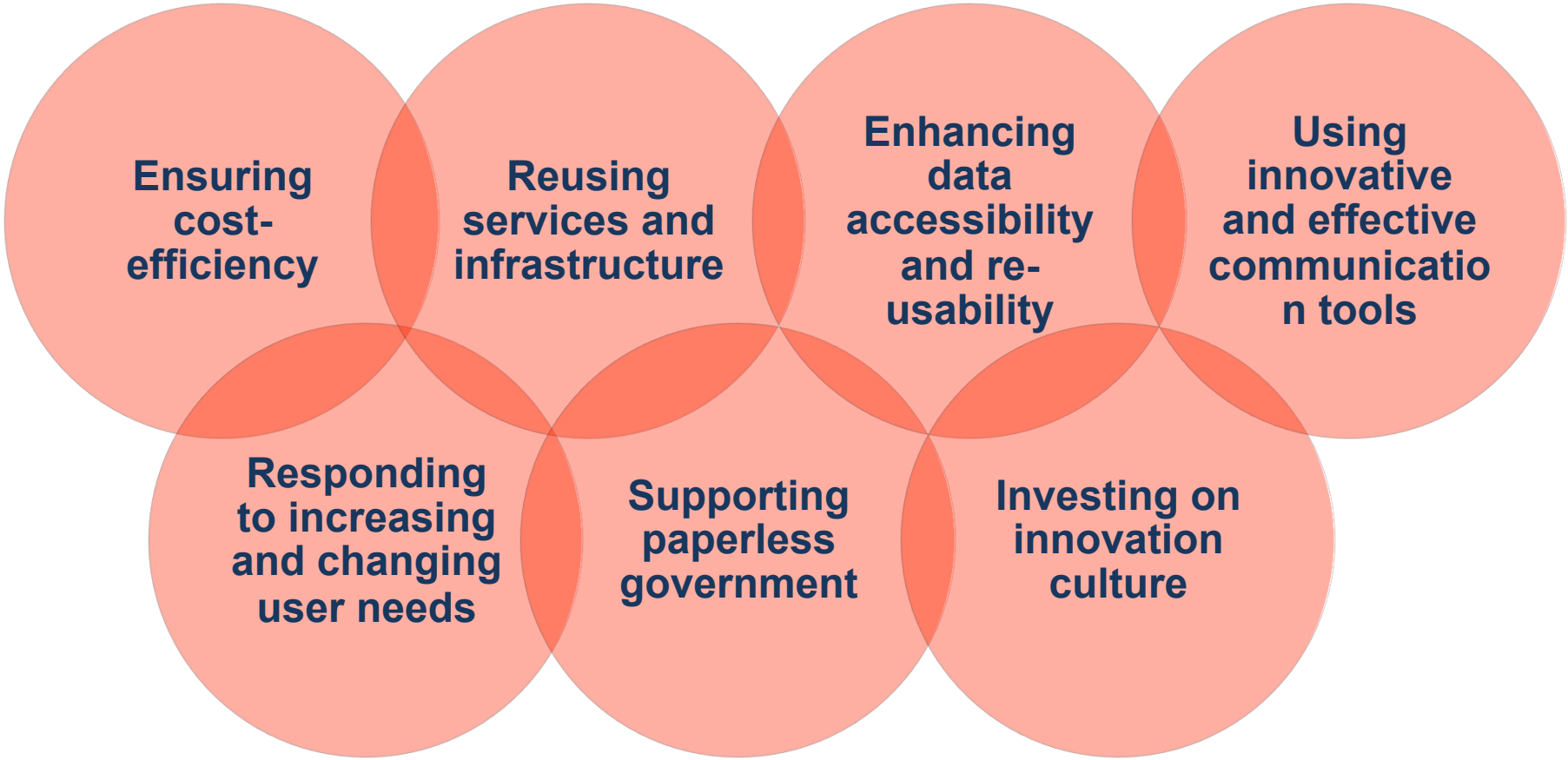
Web-COSI
Web Communities for
Statistics for Social Innovation
INCREASING TRUST IN COLLECTIVELY GENERATED STATISTICS

It is a two-year project (1/2014 -12/2015) funded by the EC-DG CONNECT, FP7 Work Programme 2013, based on a Consortium led by **Istat** with the partnership of **OECD** and two NGOs: **Lunaria** and **i-genius**

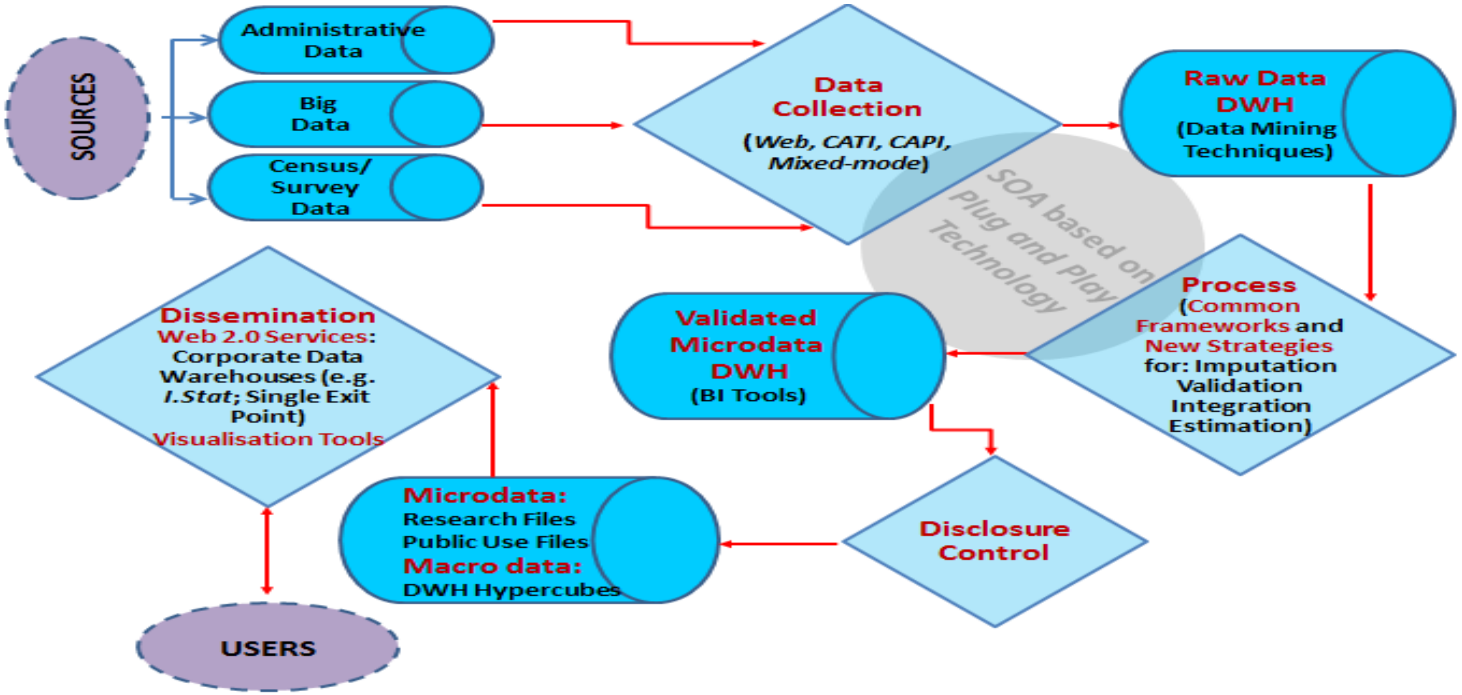
- It uses Web 2.0 technologies to foster the engagement of citizens and society at large in the area of new measures of societal progress and well-being
- It implements Web 2.0 applications for collecting, producing and visualising locally generated information and data, empowering the usage of crowd sourced data
- It contributes to a better integration and complementarity of official and non-official statistics, increasing trust in collectively generated statistics

Modernisation Strategy

Common innovation goals:



Stat2015 stylised framework

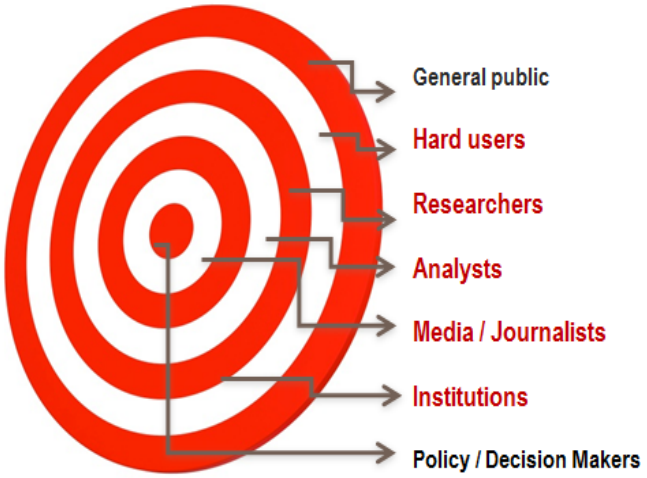


- NSI commitments towards common **Business Architecture** (BA) to **industrialise** and **standardise** the statistical production process
- **Service-Oriented Architecture** (SOA) based on **plug-and-play technology** to ensure also **interoperability** between different systems

- Shared **standard IT tools** and **methods**
- Support **data sharing** and **Open data**
- Explore **digital and big data**

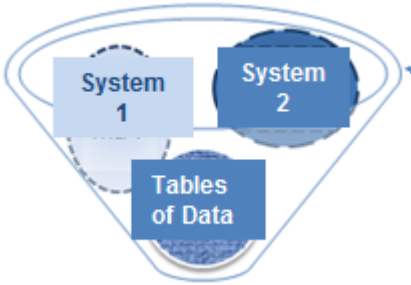
Putting users in the driver's seat

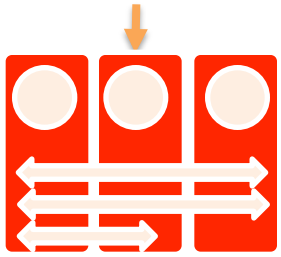
- From "naked" data to tailored services and tools
- From plain dissemination to Web 2.0 statistics (e.g. collaborative platforms, visualisation tools)
- From data description to analytics based on Linked Open Data
- Towards Open Data real time machine (impacts assessment and simulation)



USERS	TOOLS
Hard users	✓ Facilitate the re-distribution of statistics ✓ Easy to use download facilities
Researchers	✓ Promoting research based on Central Bank statistics ✓ Availability of micro data
Analysts	✓ Easy to select country comparison ✓ Ability to select charts, graphs and associated statistics directly from publications
Media /Journalists	✓ Workshop, training, web seminars ✓ Using descriptive statistics in press releases ✓ Adding deep links to series in press releases ✓ Communicate using a common terminology
Institutions	✓ Provide reporting agents with comparable statistics

I.Stat – The macrodata dissemination system





- Data **export** in different formats
- **Data** and **metadata** available **at the same time**
- **Logging in** to create a **customised Data Warehouse**
- **Creating** tables

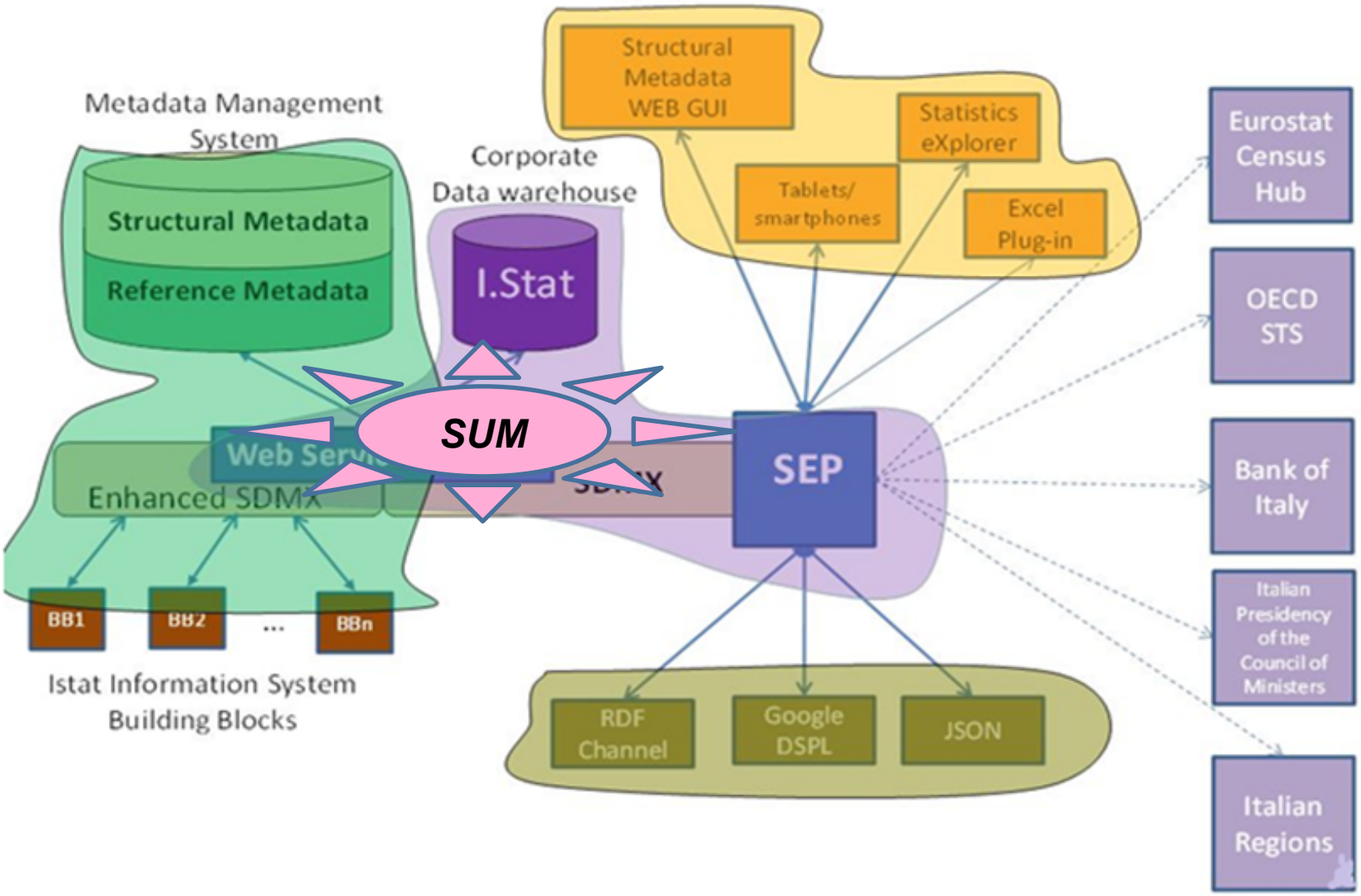
Implementation of a **Single Dissemination System where:**

- all Istat data are uploaded;
- there is constant interaction with users;
- a continuous process of improvement exists

Data are available by **thematic area and not by statistical production process**

Availability of a set of functions to improve **data surfing**

Single Exit Point – machine to machine

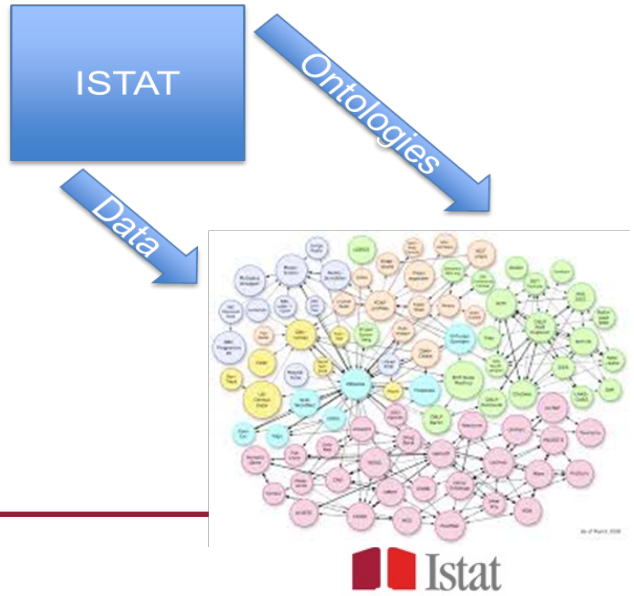


Linked Open Data (LOD) in official statistics: a turning point

- LOD-centred business strategies: adding value through integration

  ✓ Definition of common ontologies to drive **internal** data and metadata integration

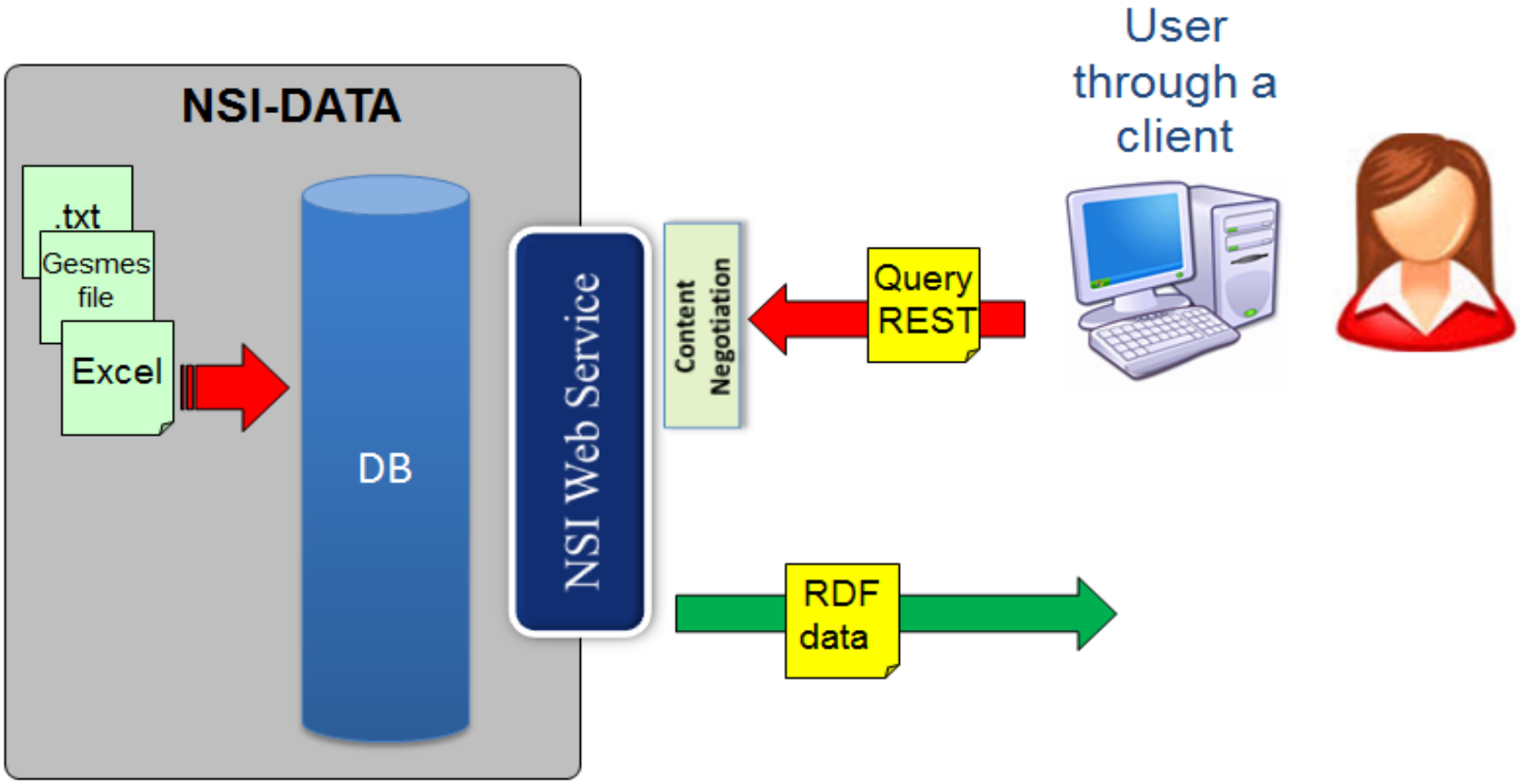
  ✓ Statistical ontologies (together with data) as a critical **link for LOD Cloud** (e.g. classifications)



Istat

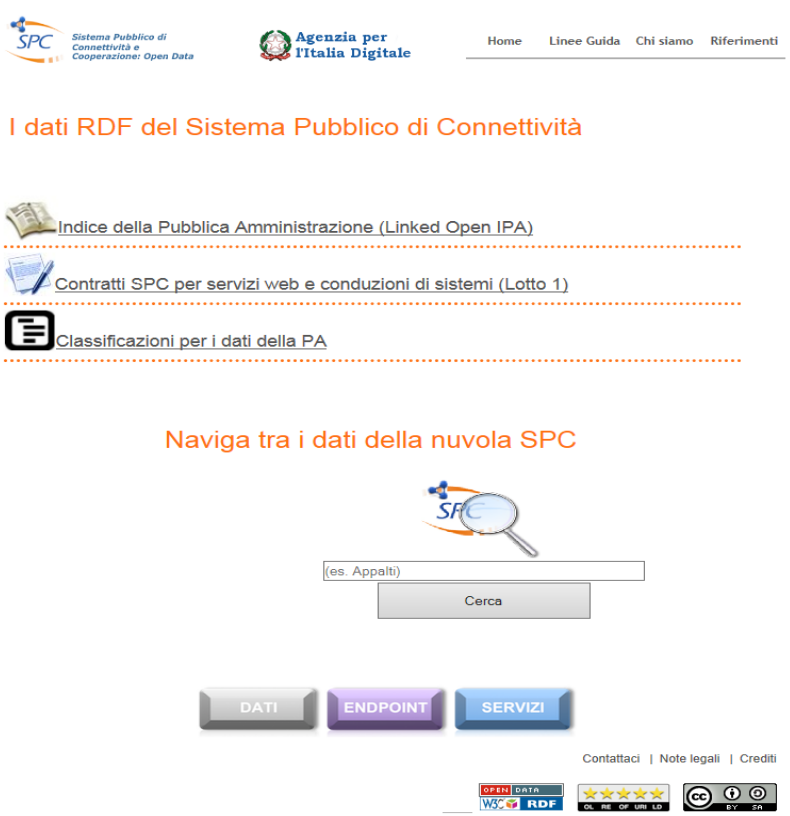
Project - Istat and Eurostat (*Grant SDMX*)

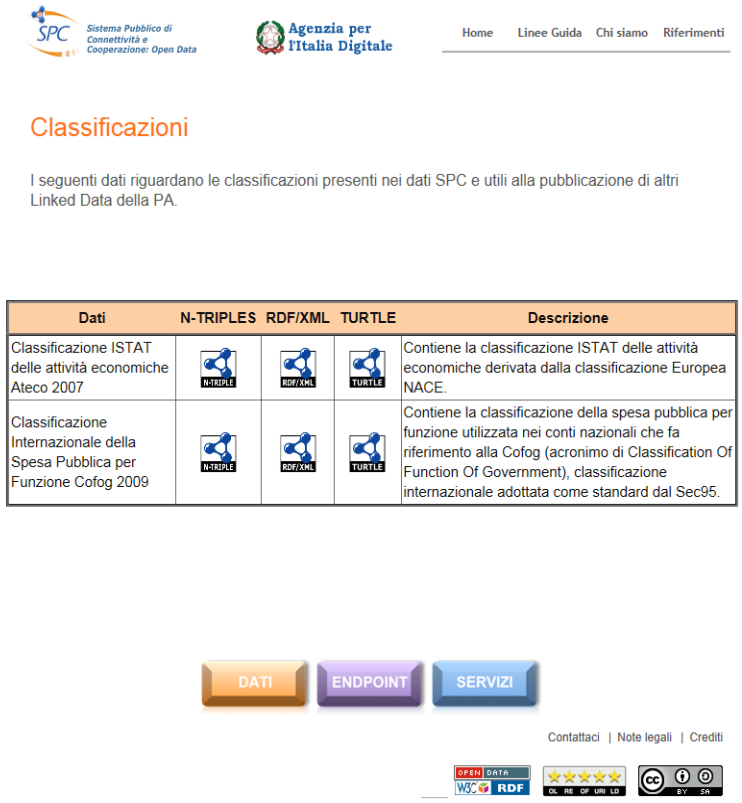
SDMX-based Dissemination



Project - Istat and Agid

- Shared design solutions to produce LOD
- Guidelines to illustrate LOD governance
- Some Istat Classifications available in LOD on Agid website





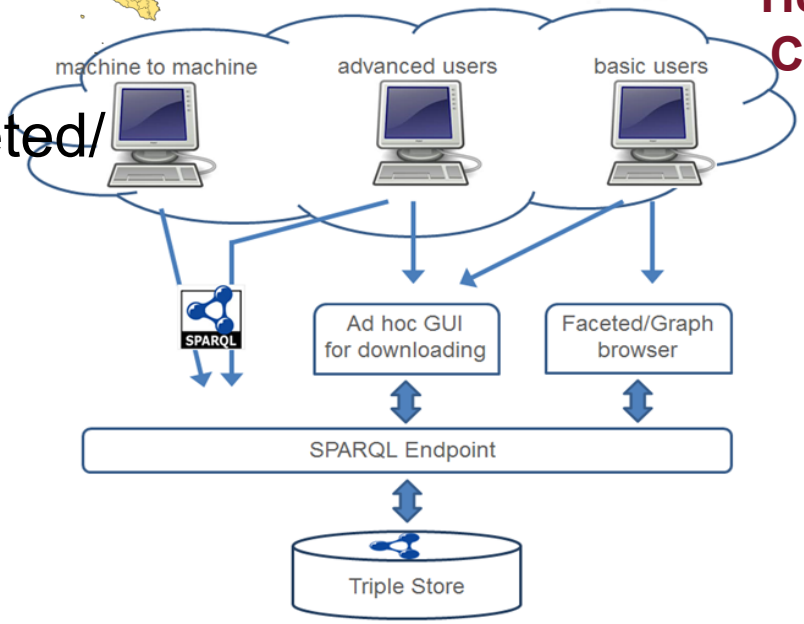
Istat LOD portal (I)

Three access points to cover the requirements of the different possible users:

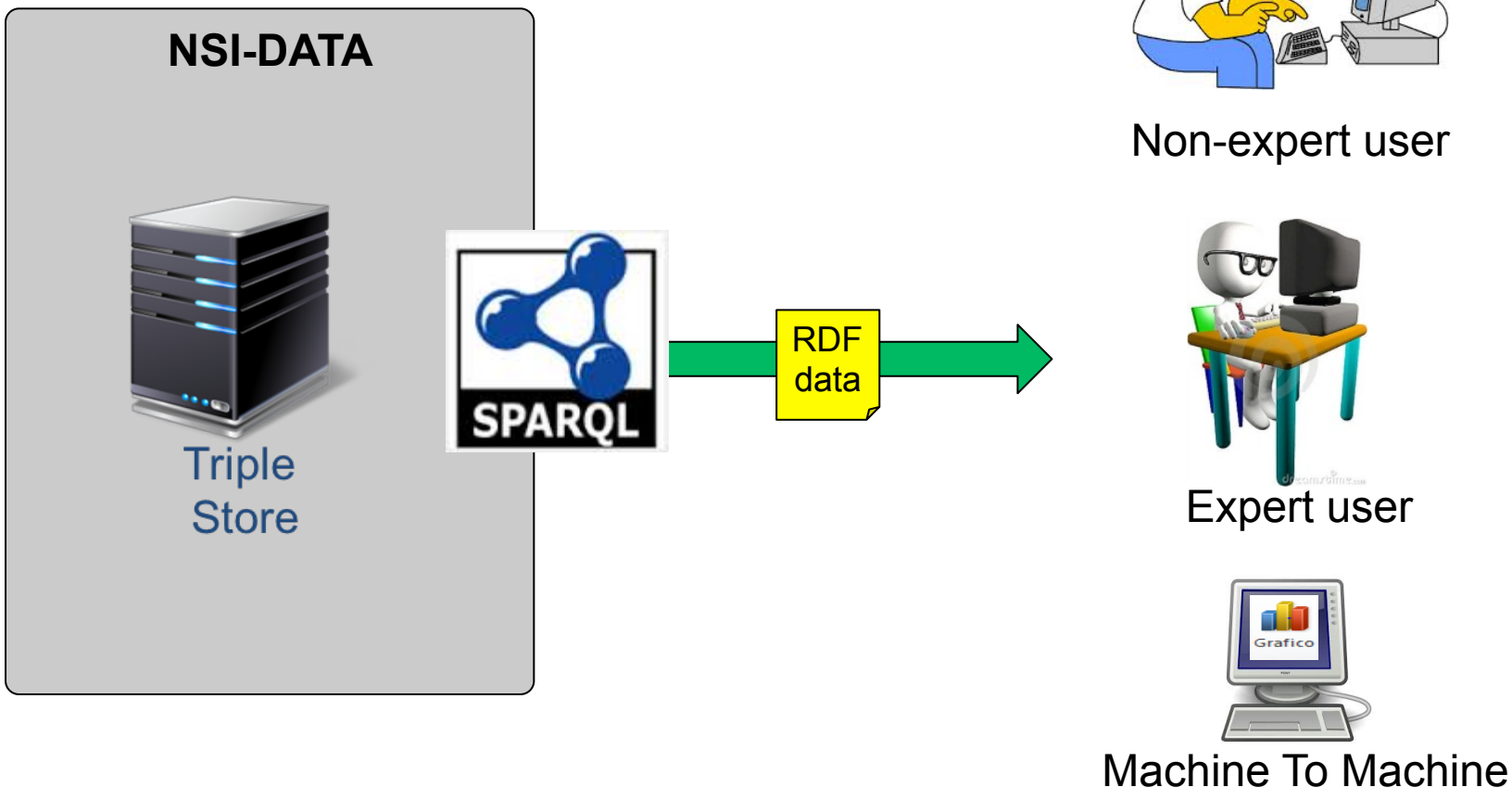
- SPARQL endpoint
  - ✓ Advanced users
  - ✓ Machine-to-machine communications
- Linked Data Interface (Faceted/ Graph browser)
  - ✓ Basic users
- *Ad hoc* GUI for datasets downloading
  - ✓ Basic users

Territorial Data

Italian Population and Housing Census

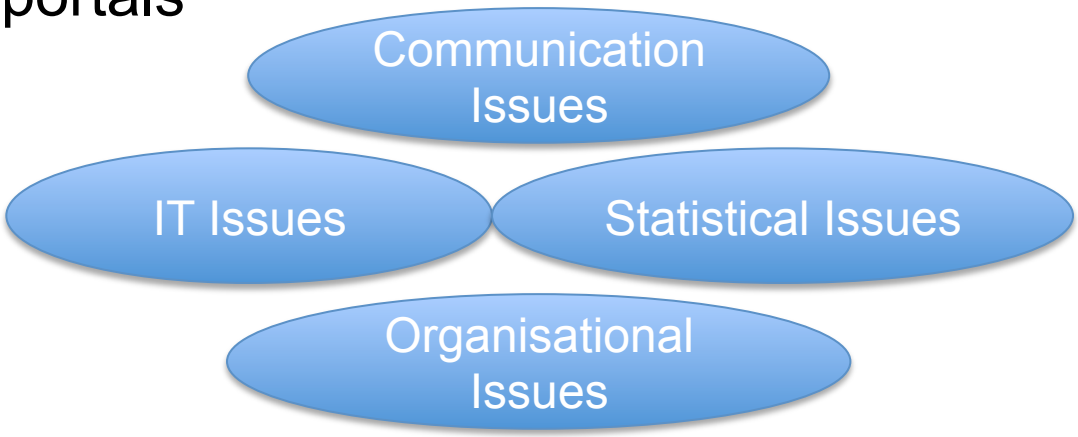


Istat LOD portal (II)



Linked Open Data (LOD) in official statistics: How?

- ✓ Organisation - Partnerships with private sector: beyond monopoly vs. competition
- ✓ Communication - Demand management and tailoring based on needs: new market assessment tools
- ✓ Statistics - Linking official statistics with other data
- ✓ IT - Rich data service portals



Research Agenda for NSIs (I)

- Making statistical standards LOD-AWARE: e.g. Generic Statistical Information Model (GSIM)

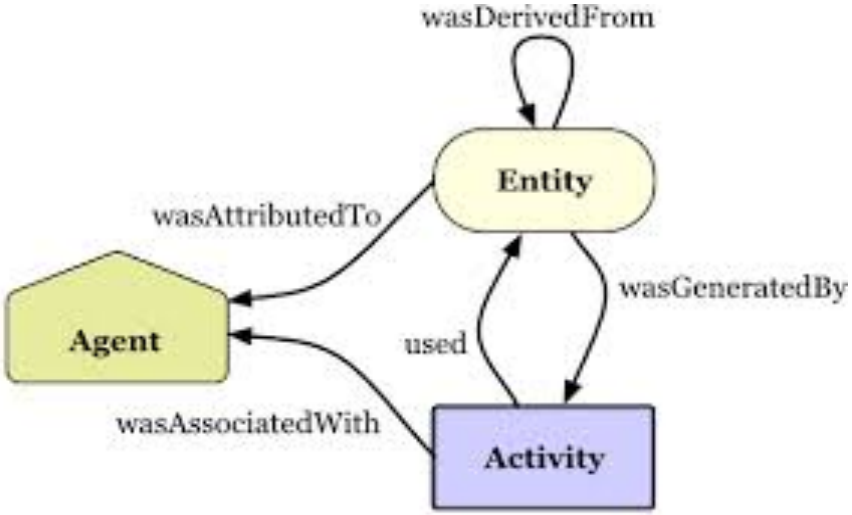
Modelled and Implemented by

SDMX + DDI + RDF

- Methods for LOD analytics:
  - ✓ Interactiveness
  - ✓ Discovery
  - ✓ Personalisation

Research Agenda for NSIs (II)

- New quality frameworks:
  - ✓ Provenance, both as publishers and as users of LOD



- IT infrastructure and tools

- New skills: LOD scientists, LOD designers, LOD builders

Future Challenges for NSIs

- Legislation, i.e. with respect to the **access and use of data**
- Privacy, i.e. managing **public trust and acceptance** of data re-use and its link to other sources
- Continuing to ensure **data quality/integrity** and **suitability of statistical methods** within this innovative framework
- Finding **new skills**
- Increasing **statistical culture** among users
- Fostering **international collaboration**
- Engaging **stakeholders in decisions**
- Building **partnerships with private sector**

Thank you for your attention

Contacts:
baldacci@istat.it

www.istat.it