

BREAL- Big Data REference Architecture and Layers

Making Big Data Real in Statistical Processes

Monica Scannapieco
Istat, Italy

In collaboration with:

Frederik Bogdanovits, Tauno Tamm (Statistics Estonia, Estonia), Arnaud Degorre, Frederic Gallois (INSEE, France), Bernhard Fischer (DESTATIS, Germany), Kostadin Georgiev (BNSI, Bulgaria), Remco Paulussen (CBS, Netherlands), Sónia Quaresma (INE, Portugal), Donato Summa (ISTAT, Italy), Peter Stoltze (Statistics Denmark, Denmark)

GENERAL OBJECTIVE

- ESSnet Big Data Pilots II:
 - Project funded by Eurostat
 - Dates: 2018-2020
 - 22 EU partners
 - Following ESSnet Big Data Pilots I (2016-2018)
- Workpackage “Process and Architecture” for the definition of **reference architectures** necessary to carry out Big Data-based production of Official Statistics at National and European levels
 - Participants: IT(coordinator), BG, DE, DK, EE, FR, NL, PT

WHY ARCHITECTURAL WORK?



NSIs have recently start building their Big Data management information systems

WHY ARCHITECTURAL WORK?



Building 'similar' information systems by reuse

WHY ARCHITECTURAL WORK?

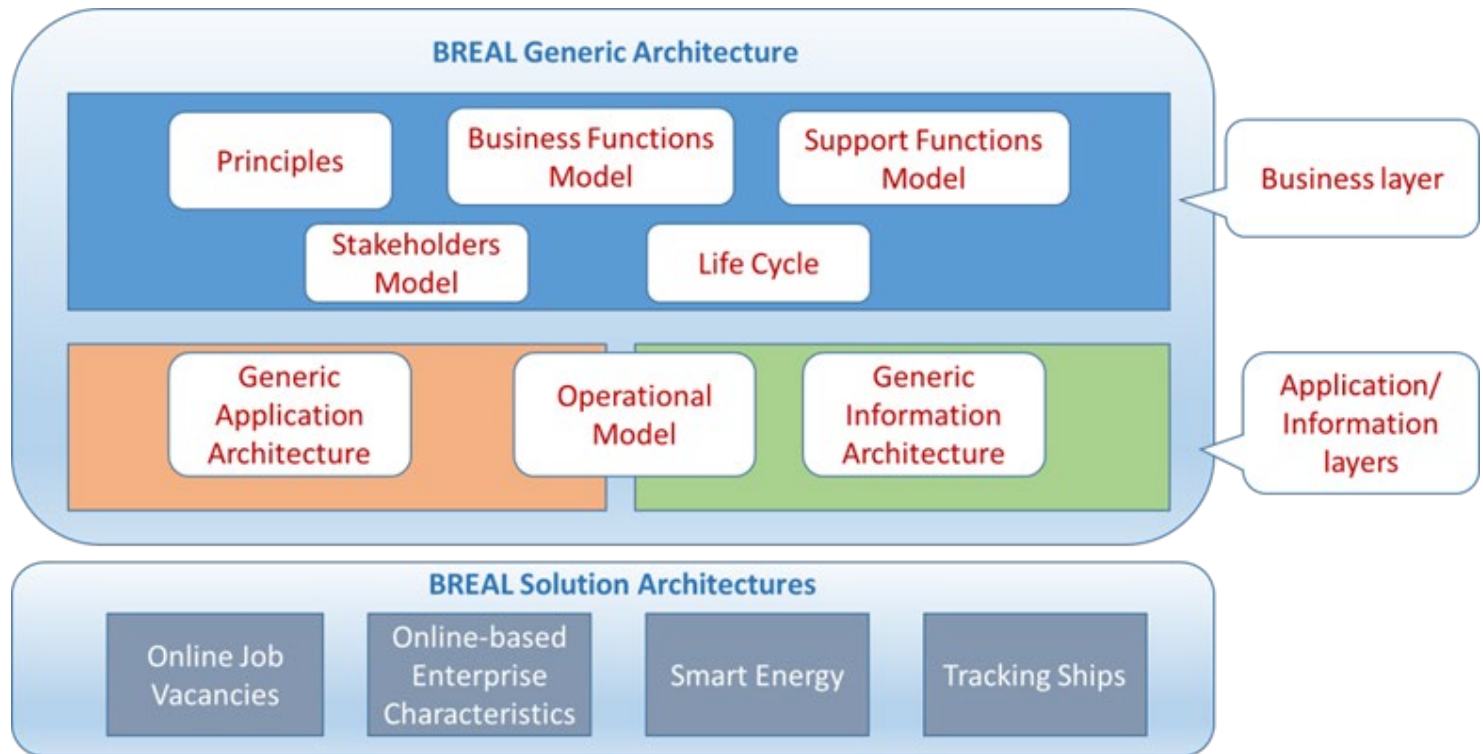


Having common infrastructures at EU level

BREAL

- Definition of **BREAL** (**B**ig data **RE**ference **A**rchitecture and **L**ayers)
- BREAL Scope
 - Within ESSnet Big Data II project: standardize and orient implementation WPs' work
 - Beyond the project: serve as a reference for National and European investments on Big Data-based Official Statistics

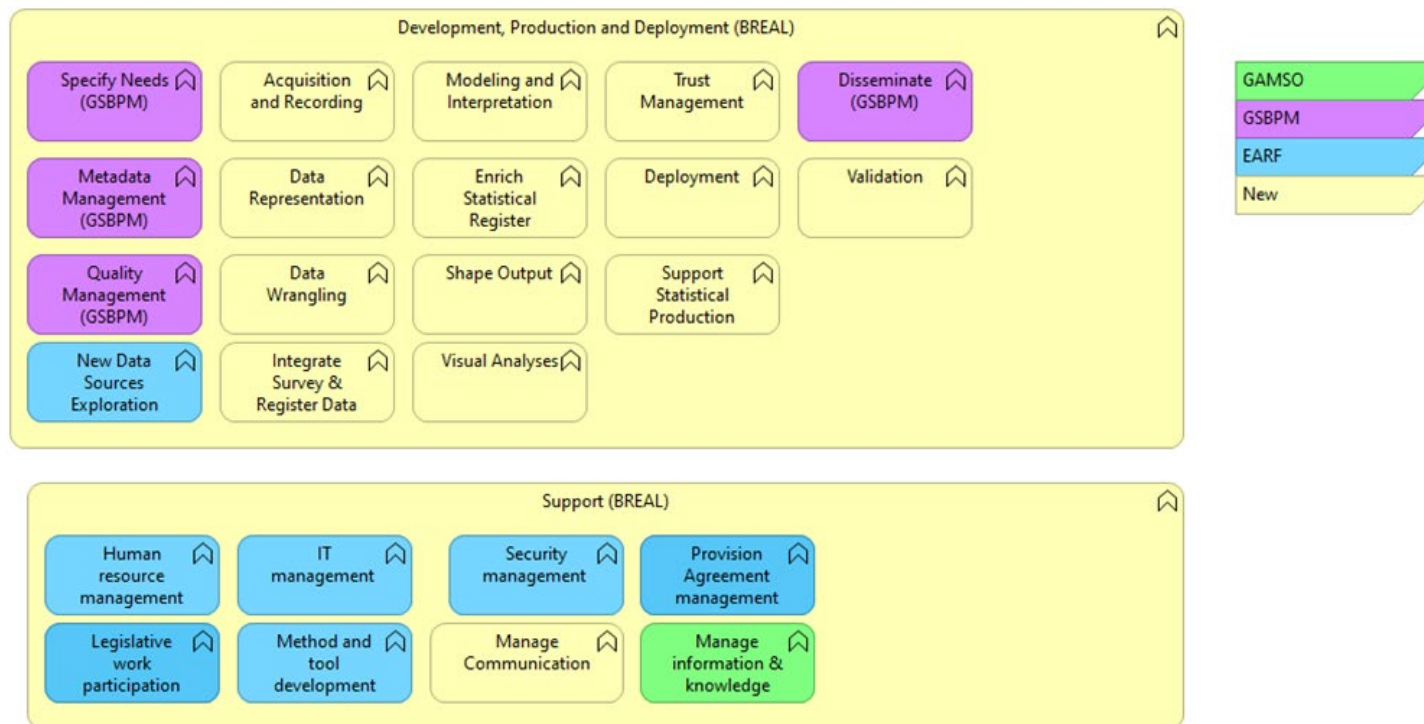
BREAL at a Glance



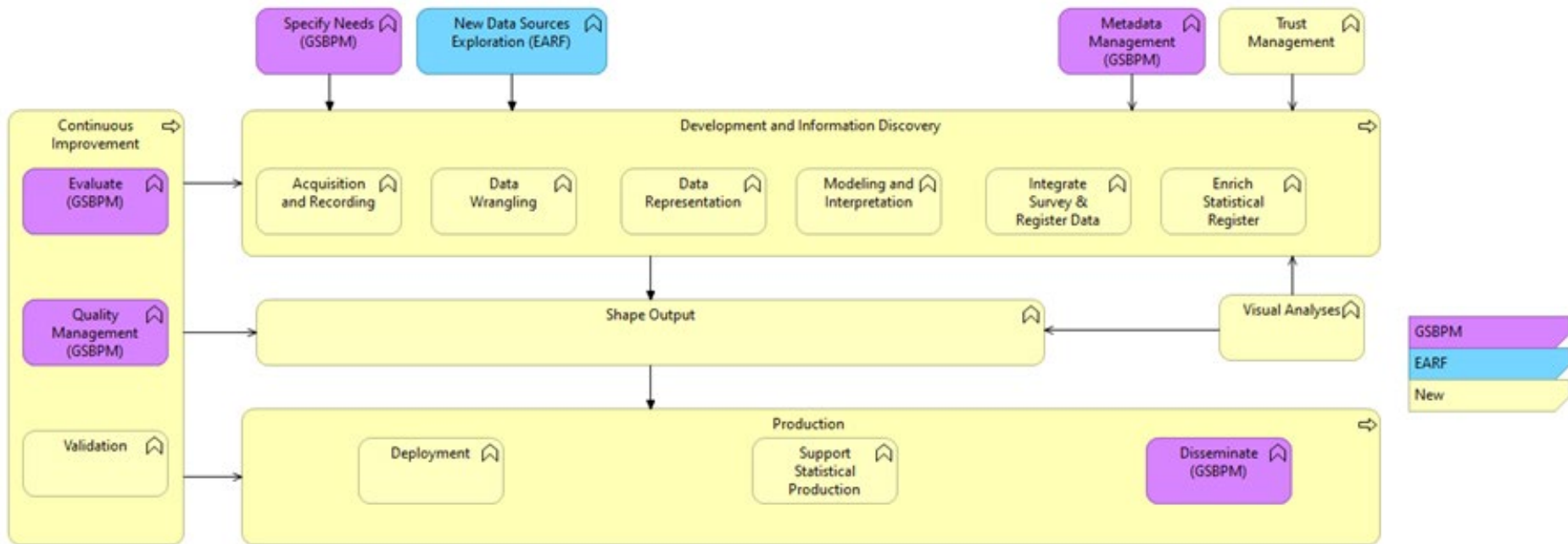
BREAL Business Layer (BL)

- List of principles
- Business and Support functions model
 - Relying onto GSBPM and GAMSO
- Life Cycle model
- Stakeholder model

BL: Business and Support Functions Models



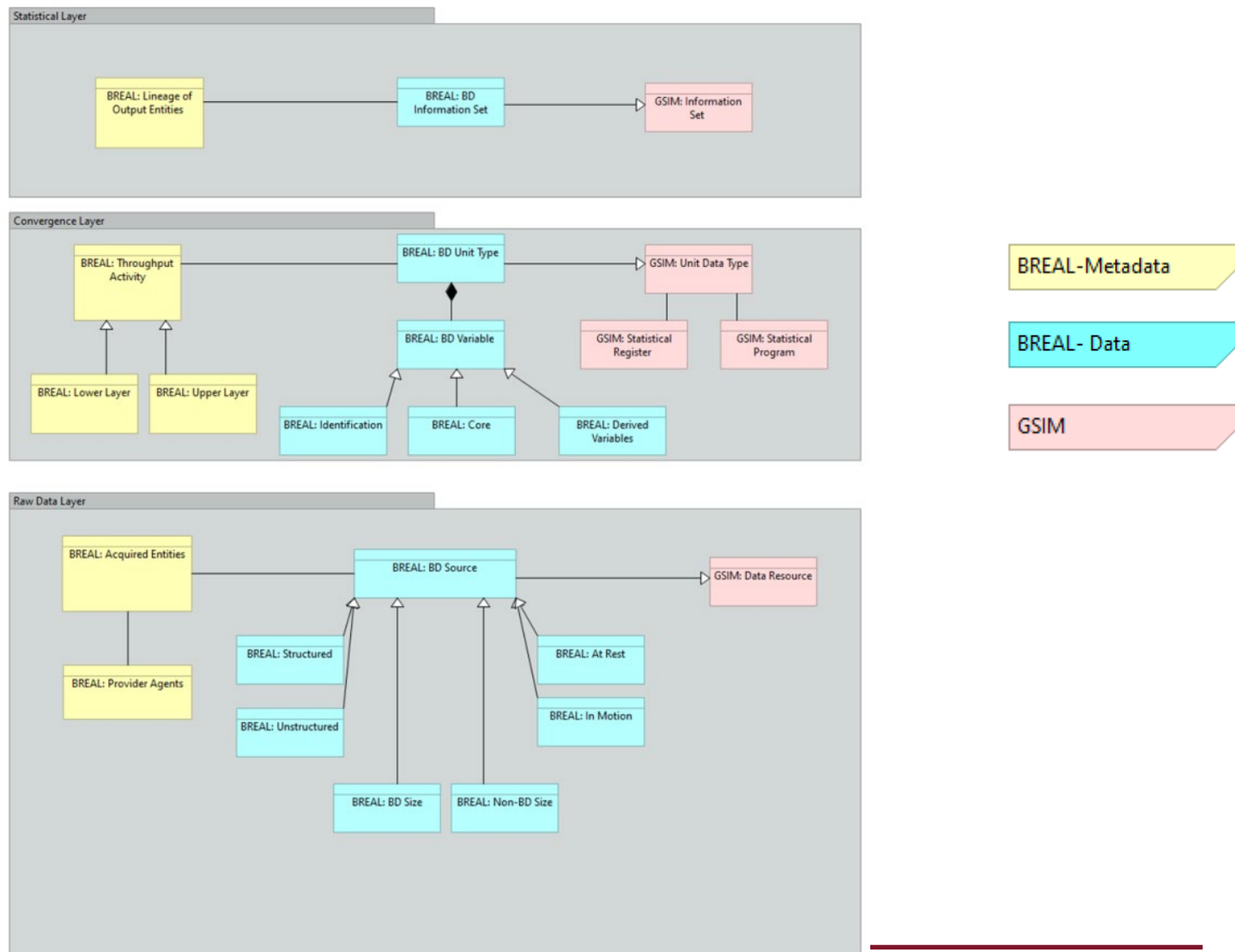
BL: BREAL Life Cycle



BREAL Application/Information Layer

- Generic Application Architecture, with a set of generic application services, proposed to show how the identified business functions can be implemented
- Operational Model, describing how data and services can be deployed in a Big Data solution
- Generic Information Architecture
 - Relying onto GSIM

Generic Information Architecture for Big Data - 1



Generic Information Architecture for Big Data - 2

- The BREAL Generic Information Architecture for Big data consists of three layers:
 - Raw Data Layer, including data that are acquired and stored
 - Convergence Layer, containing data represented as units of interest for the analyses.
 - Statistical Layer, including those concepts that are the targets of the analysis.

Conclusions

- BREAL is a reference architecture for Big Data-based statistical production
- It relies on ModernStats standards (GSIM, GSBPM, GAMSO)
- Main users of BREAL are:
 - NSIs that aim to introduce the use of Big Data in their production processes
 - Public and private organizations that would like to follow a defined and controlled way of producing Big Data-based statistics guided by the Official Statistics expertise