

WP F – Process and Architecture

Monica Scannapieco
Istat

ESSnet Big Data II
Kickoff meeting of implementation track
Vienna, 3-4 December 2018

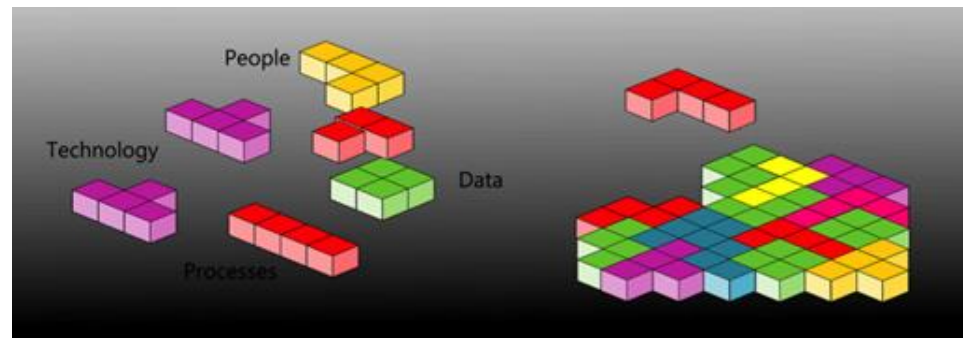
Outline

- Content
- Organization of the work
- Next steps and Conclusions

Content

WP F GENERAL OBJECTIVE

- Definition of **reference architectures** necessary to carry out big data production
 - At national level
 - At European level
- **Architecture** is the fundamental organisation of a system embodied in its **components**, their **relationships to each other**, and **to the environment**, and the **principle guiding its design and evolution**
 - [ISO/IEC 42010 Systems and software engineering - Architecture description]



PRELIMINARIES - 1

- To define and use architectures there exist several frameworks
- One framework very much used, is TOGAF - The Open Group Architecture Framework (TOGAF)
- Divides architectures into four layers:
 1. **Business architecture** - Describes the processes the business uses to meet its goals
 2. **Application architecture** - Describes how specific applications are designed and how they interact with each other
 3. **Data architecture** - Describes how the enterprise datastores are organized and accessed
 4. **Technical architecture** - Describes the hardware and software infrastructure that supports applications and their interactions

PRELIMINARIES - 2

- **Building Blocks** (BBs): represent (potentially re-usable) components of (IT) capability that can be combined with other building blocks to support business solutions
- TOGAF BB's characteristics:
 - package of functionalities defined to meet the business needs across an organization;
 - has a type, such as actor, application, or data entity;
 - has a defined boundary;
 - may interoperate with other, inter-dependent, BBs.

PRELIMINARIES - 3

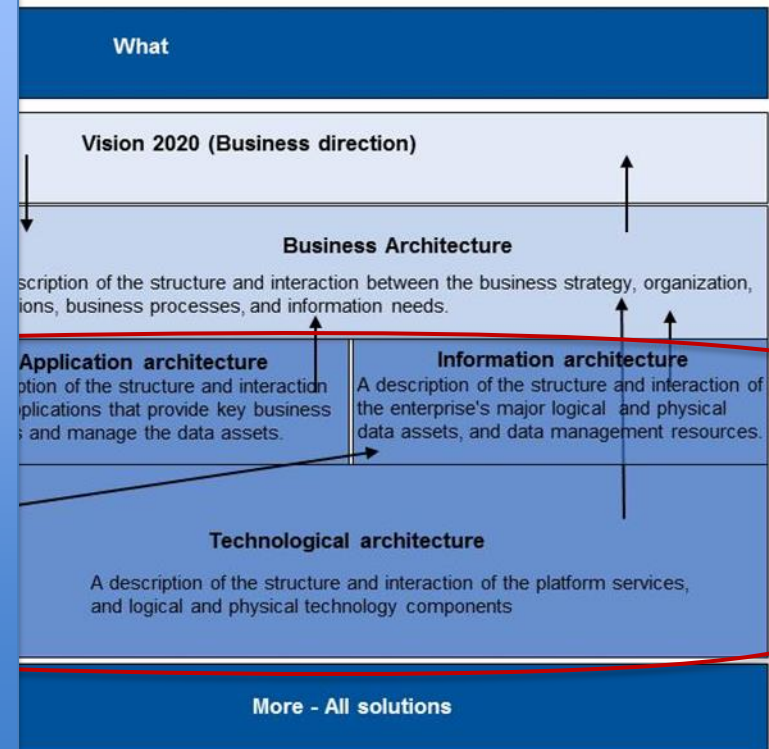
- **ESS Enterprise Architecture Reference Framework (EARF):**
 - Resulting from the work of a dedicated DIME/ITDG Task Force
 - Set of Key artifacts that can be used for governance and realization of Vision 2020 projects
 - Example: ESS EARF Building Blocks is an artefact proposing a set of components (Building Blocks) that can be combined to deliver the information systems needed in ESS .
 - E.g.: Metadata Management, Process Orchestrator, Primary Data Storage.
 - https://ec.europa.eu/eurostat/cros/content/ess-enterprise-architecture-reference-framework_en

WP F SCOPE: WHICH ARCHITECTURES?

Architectures necessary to carry out big national and European levels

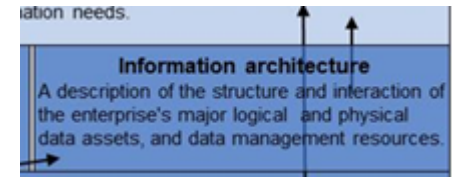
Why do we need a RA?

- Same common language
- Adoption of Standards
- Improving understanding of solutions
- Reusability, interoperability, portability, sustainability and maintenance



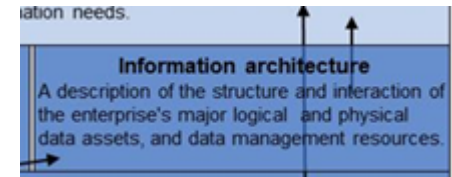
ESS EARF - Enterprise Architecture Reference Framework

WP F CONTENT: INFORMATION ARCHITECTURE - 1



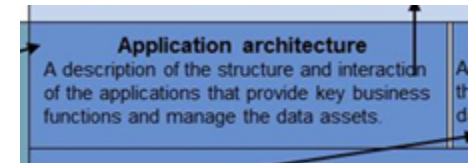
- **Information architecture:** description of major logical and physical data assets and data management resources
- For each implementation WP:
 - Data (e.g. enterprises anagraphical infos)
 - Structural metadata (e.g. data model of enterprises anagraphical infos)
 - Process metadata (e.g. URL of the website where enterprises anagraphical infos where found)
 - Quality metadata (e.g. accuracy of estimated economic activity)

WP F CONTENT: INFORMATION ARCHITECTURE - 2



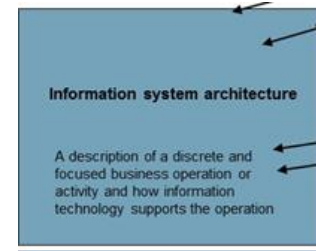
- Within the Information architecture, it will be useful to identify specific Building Blocks:
- Examples:
 - Document Base to store scraped Web sites
 - Graph base to store structural metadata

WP F CONTENT: APPLICATION ARCHITECTURE



- **Application Architecture**: description of the structure and interaction of the applications that provide key business functions and manage data assets
- For each implementation WP:
 - Application Services (e.g. Text Tokenizer)
 - Input
 - Output
 - Signature
 - Behaviour
 - Building Blocks (e.g. Dissemination platform)
 - Include application services but also
 - Coarser-grained components

WP F CONTENT: SOLUTION ARCHITECTURE



- **Information System Architecture:** description of a discrete and focused business operation or activity and how information technology supports the operation
- For each implementation WP:
 - Solution BBs
 - e.g. Elastic Search to implement a 'Document Base' BB
 - Deployment scenarios
 - e.g. National level deployment vs ESS level deployment

Reference Standards - 1

- Official Statistics Standards:
 - ESS EARF (Enterprise Architecture Reference Framework)
 - Information architecture:
 - GSIM (Generic Statistical Information Model)
 - CSDA (Common Statistical Data Architecture)
 - Application architecture:
 - GSBPM (Generic Statistical Business Process Model)
 - CSPA (Common Statistical Production Architecture)
- Technological standards
 - Web semantics standards (e.g. OWL)
 - SDMX

Reference Standards - 2

- Big Data Standardization Effort:
 - National Institutes of Standards and Technology (NIST)
 - NIST Big Data Public Working Group (NBD-PWG) prepared the NIST Big Data Interoperability Framework (NBDIF)
 - 9 volumes
 - Will be released in three versions
 - Currently version 2 is available at: <https://bigdatawg.nist.gov/>
 - Volume 6 is a Reference Architecture: NIST Big Data Reference Architecture (NBDRA)

Reference Standards: What is the role in WPF work?

- They are surely a starting point
- In the case of EARF:
 - Specialization effort needed
 - Can have a role of ‘frame’ for the work
- GSBPM:
 - Thought for ‘traditional’ survey-based process
 - Rethinking necessary for Big Data
 - Collect and Design Inversion
 - Phase Cycling
 - Etc.
- GSIM:
 - Where are unstructured data?

Reference Standards: What is the role in WPF work?

- CSPA
 - Non functional requirements need to be stressed: parallelism, security, etc.
- CSDA
 - Nice starting point for listing specific Big Data capabilities (like data ingestion, data transformation, provenance and lineage...)
- NBDRA
 - Huge effort
 - Useful to check completeness of our work
 - Not specialized to statistical domain

Organization of the work

WP F Participants

- 8 countries:
 - IT (coordinator)
 - BG
 - DE
 - DK
 - EE
 - FR
 - NL
 - PT

WORK BREAKDOWN STRUCTURE

- Task 1: Architectural Framework
- Task 2: Solution Architectures
- Task 3: Meetings for the Implementation Track
- Deliverables
 - F1 (M12): Data and Application Architecture Vs1
 - F2 (M25): Data and Application Architecture Vs2

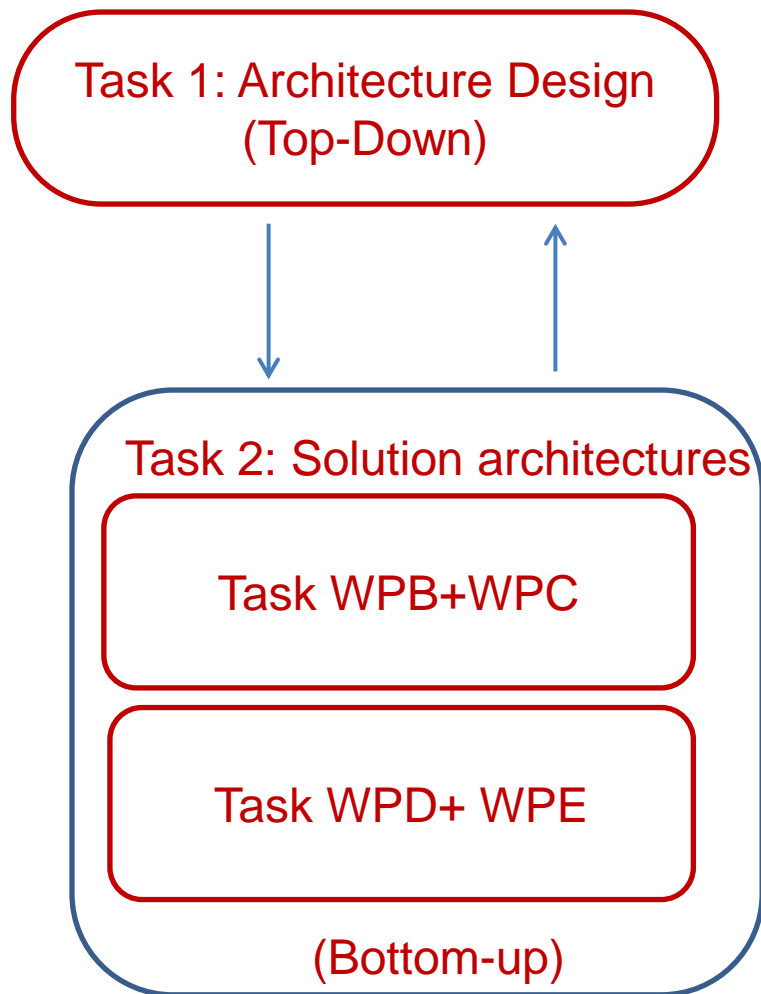
TASK 1: ARCHITECTURAL FRAMEWORK

- **Focus:** how data, processes and services need to be modeled within EU and national architectures to support big data production
- **Subtasks :**
 - Task 1.1: Setting up the tools (modelling language, architectural frameworks);
 - Task: 1.2: State of the art and sharing of BD Pilots I results;
 - Task 1.3: Identification of relevant dimensions (by sources, by data types, national-EU etc.);
 - Task 1.4: Data and metadata architecture design;
 - Task 1.5: Service and process architecture design.

TASK 2: SOLUTION ARCHITECTURES

- **Focus:** proposal of solution architectures for big data production. The proposal will result from either concrete participation or strict interaction with implementation activities carried out within WPs.
- **Subtasks :**
 - Task 2.1: Design of WPB+WPC solution architectures (IT, DE, FR, BG)
 - Task 2.2: Design of WPD+WPE solution architectures (IT, NL, EE, DK)

INTERACTIONS BTW TASK 1 AND TASK 2



- Participants to WPF will either take part or have strict interaction with implementation and piloting activities of the project.
 - Tasks/subtasks for coordination with WPF included in WPs
- They will abstract in a bottom-up way the concrete solutions emerging from the vertical lines of work inside the WPs.
- They will reason and share such solutions within WPF and evaluate the conformance to data and application architectures resulting from Task 1.
- They will provide feedback and guide to the vertical working lines inside the WPs.

TASK 3: MEETINGS OF THE IMPLEMENTATION TRACK

- Two meetings for two full days will be organised bringing together members of WPF with members of all implementation WPs (WPB-WPE)
 - a kick-off meeting in M2 (actually M1)
 - intermediate meeting in M14 (Dec 2019)
 - The objective of these meetings is to coordinate and organise the work of the whole implementation track together
- WP F internal meetings:
 - Mid 2019 (M9 June-July 2019)
 - Mid 2020 (M20 –May 2020)

WORKPLAN, DELIVERABLES AND MILESTONES



Workplan: Process and Architecture	M1-M3	M4-M6	M7-M9	M10-M12	M13-M15	M16-M18	M19-M21	M21-25
TASK F1: Architectural Framework								
Design of a data and application architecture for Big data production								
1.1 Setting up the tools (modelling language, architectural frameworks)								
1.2 State of the art and sharing of BD Pilots I efforts								
1.3 Identification of relevant dimensions (by sources, by data types, national-EU etc.)								
1.4 Data and Metadata architecture design								
1.5 Service and process architecture design								
Task F2: IT architectures								
Design of IT architecture for Big data production								
2.1 WPA+WPB IT architectures								
2.2 WPC+WPD IT architectures								

Deliverables

- **M12** Deliverable on Data and Application Architecture vs1
- **M25** Deliverable on Data and Application Architecture vs2

Milestones

- **M3** Report on the kick-off meeting for the implementation track
- **M9** Report on the WP meeting mid 2019
- **M14** Report on the mid-term meeting for the implementation track
- **M20** Report on the WP meeting mid 2020

NEXT STEPS AND CONCLUSIONS

- Very next steps:
 - Planning of virtual and physical meetings
 - Assignment of (specific) responsibilities on tasks and subtasks
 - Intermediate milestones