**UNECE High-level Group for the  
Modernisation of Official Statistics**

**Project Proposal: Data Architecture 2**

|  |
| --- |
| This project proposal was prepared by Data Architecture 2017 project (coord. by Dick Woensdregt and Carlo Vaccari), and is submitted to the HLG-MOS for their approval. |

|  |  |
| --- | --- |
| **1 Purpose** | |
| To complete the development of a reference framework for data architectures, testing the reference architecture defined in 2017 against other use-cases.  Also, to Integrate the Data Architecture with outcomes from other groups like UN-GWG, Data Integration project and additionally to work on formal metadata specification through ontologies. | |
| **2 Project description** | |
| The project consists of the completion of a reference framework, to describe a standardized data environment, and to support the design, production and dissemination of official statistics.  In 2017 Data Architecture (DA) project three main deliverables were planned:   * Reference Architecture document * Use-cases * Guidelines   The Reference Architecture, including Capabilities and Building Blocks, has been defined in 2017, and five different use-cases have been described, but the drafting of the guidelines was not possible because of the late availability of use-cases.  In 2018, the following work-packages will be provided:  WP1: WP1: DA Guidelines   * Draw up guidelines for the introduction of DA inside NSIs - the guidelines should clearly identify the suggestions and the steps for the introduction of the reference Data Architecture in our institutes   WP2: Revised Data Architecture that includes consideration of:   * Data Integration, Metadata and DA – the Data Integration project in 2017 listed many use-cases that must be checked against DA; at the same time it’s important to design the integration of DA with standard metadata systems * DA impact on new data sources – in 2017 we started to coordinate the activities on DA with those of UN Global Working Group on Big Data (<https://unstats.un.org/bigdata/>). We defined a common template for use-case definition and we shared some use-cases between the two groups. We think that the collaboration can be very fruitful for both sides, leading to a new definition of a DA to be used to facilitate the usage of new data sources like Big Data in Official Statistics * DA and Statistical Ontologies – many groups started defining shared ontologies to improve the semantic definition of statistical data; also these activities can be framed in a global Data Architecture project. In addition, a coordination activity will be done with the ESSnet on Linked Open Statistics that is going to start.   WP3: Survey   * Survey on Building Blocks in NSIs - an interesting way to verify the compliance of Data Architecture with NSIs data structure could be a web survey by which we would ask NSIs to list the software tools used for each Logical Building Blocks identified in Data Architecture   The outputs of the new DA project would be:   * Guidelines for the introduction of Data Architecture in NSIs including:   + Description of data artefacts, like Statistical Data Dictionaries, to drive the definition of data structures and metadata   + Relations of the DA with the data-life cycle management   + Best practices to ensure data quality and to share technical solutions (like CSPA services)   + Maturity Model and Roadmaps for the adoption and implementation of Data Architecture * A more complete list of use-cases that can be used as Proof of Concept for the Reference Architecture; use-cases should cover a wide range of real systems to test completeness and adherence of DA to the needs of NSIs * A revised Reference Data Architecture, complete with Capabilities and Building Blocks, with more details about:   + the influence of new data sources on DA   + the connections between Metadata systems and DA   + the relationships between DA and semantic data structures * Report from the survey, listing the software tools used by NSIs to manage data for each Capability and Building Block defined in our DA | |
| **3 Alternatives considered** | |
| 1. Try to complete Data Architecture outside of UNECE project: the alternative is not valid for the lack of a sponsorship that guarantees the necessary staff commitments and for the need for co-ordination of the project to ensure continuity and coordination with other initiatives. 2. Doing nothing. This option would damage NSIs, because it is important to be able to share integrated data definitions and because an Enterprise Architecture is based in four layers (business, **data**, applications and technology), all of which are needed. | |
| **4 Expected Benefits** | |
| ☐ | Reduced costs |
| ☒ | Increased efficiency |
| ☐ | Reduced risks |
| ☒ | New capabilities to meet user needs |
| Justification:  Some expected benefits from the project are:   * Enhancing governance of data inside an organization * Helping to guide the data strategy of the statistical institution, by aligning the efforts of its different units * Better understanding of data and metadata managed by an institution * Defining how to incorporate and integrate traditional and emerging data sources * Driving the definition of data structures and metadata, for data and processes, including receipt and exchange * Enhancing the ability of statistical offices to share information and knowledge, enhancing the comparability and transparency of the information * Helping to adopt common CSPA statistical services, by providing a common data context where they can operate * Making it easier to source data from different domains, for new products and services to satisfy new users’ needs * Addressing the information management challenges that prevent national statistical agencies from quickly leveraging new data sources * Positioning national statistical agencies to leverage opportunities, based on technological development in linked data and metadata, next generation data storage, analysis and visualisation, and new methods and algorithms | |
| **5 Which key priorities in the HLG-MOS Strategic Framework does the proposed project relate to?** | |
| ☒ | Take cost out of our organisations to reinvest in more value added areas |
| ☒ | Explore new areas collectively and leverage each other’s' research investments in specific areas |
| ☒ | Provide whole of government data ecosystems based on international standards, for better estimates in key policy areas |
| ☒ | Renew our governance and operating processes |
| Justification:   * It reduces re-work needed to deliver the information to different applications. * It helps in the adoption of different standards, by providing a common data context where they can be implemented. * It supports the development and implementation of the data governance needed to support the alignment of efforts around a data strategy. * It reduces costs, by helping to share information, applications and knowledge with less effort, and using common tools. | |
| **6 How does the proposed project relate to other activities under the HLG-MOS?** | |
| It complements Data Integration project results and ensures coordination with other initiatives like UN-GWG and Ontologies.  It complements and consolidates the CSPA model, by taking care of the most important assets for statistical organizations, data and metadata.  It applies the business framework provided by GSBPM, and makes concrete use of GSIM standard.  It covers the layer between business process and applications defined by an enterprise architecture. | |
| **7 Proposed timetable** | |
| Project to be completed by November 2018.  At least two face-to-face meetings between March and September 2018 | |
| **8 Expected resources and costs** | |
| Coordination of the project  Travel costs for two face-to-face meetings | |