

INTEROP-VLab Task Force – TFX: Big Data

This document should not exceed 1 page

Subgroups should be listed and if necessary document on 2nd page

Title of Activity	Duration in Years or Permanent	New TF / Within X TF / Other Activity	Status
Big Data	Permanent	New	APPROVED
Proposer/s and Sponsors (Organisation/Person name/Email and lead Proposer)			
Organisation	Name	Email	Lead
ITI	Daniel Sáez	dsaez@iti.es	X
Current Members (To be updated regularly)			
ITI,			
Description of the activity (content) and expectations			
<p>Enterprise networks are becoming complex, tightly connected, and very dynamic, in response to changing demands, fluctuating markets, and low margins. These networks highly rely on data interoperability networks, which in several cases need to deal with large volumes of data (Volume), coming from different sources and formats (Variety), and/or characterised by massive and continuous data flows (Velocity). The intersection of these three dimensions (Volume, Variety, and Velocity, among others) at large scales is challenging, and thus within active research in the Big Data field.</p> <p>The objective of this task force is to deal with the challenges of Big Data (Volume, Variety, Velocity) in the field of enterprise data interoperability and data integration. Powerful tools can be applied to collect, store, analyse, process, and visualize huge amounts of data, in order to improve resource allocation, quality of service, maintenance, marketing, decision making, ... along a global value chain and create a significant advantage for European Industry. Data coming from different sources and from all the actors in the value chain will influence the decision making process. Interoperability and integration of datasets are essential for a wide adoption and impact within and across sectors.</p> <p>This task force expects to gather knowledge and experience from different fields, join efforts, and lead the advance of data interoperability and integration techniques in line with current and emerging industry needs. That is, dealing with large volumes of data, from different sources and different formats, and requiring fast processing times. This would require an effective and efficient management of interacting components from different systems, the automation of data transformations, processing and analysis, as well as driving the adoption of standardized data bridging models for different industry sectors to enable an easy and rapid connection and processing. These data interoperability and integration layer will provide a base for the application of analysis techniques to increase efficiency and sustainability along a value chain.</p>			
Motivation and rationale for the Activity			
<p>Big Data is a big trend in Europe, with its own PPP (www.bdva.eu), impacting in every sector and domain. Data usage is growing but, in both industry and academia, it is treated and handled in a fragmented way. In order to ensure a coherent use of data, a wide range of stakeholders along the Data Value Chain should be brought together to facilitate cooperation in every domain/sector. The cross-fertilisation of stakeholders and datasets is a key element for the advance of the Big Data economy.</p> <p>In this sense, there is an urgent need to build an interoperability layer across all different systems taking advantage of transformation and semantic integration techniques.</p>			
State of Play (context of the activity and work already done, if any)			
<ul style="list-style-type: none"> • Launching step 			
Targeted audience and stakeholders of the work – who is it for			
<p>The stakeholders that will form the basis for interoperable data-driven ecosystems as a source for new businesses and innovations on Big Data are:</p> <ul style="list-style-type: none"> • ICT Vendors (Large and SME) • Users across different industrial and application sectors (private and public) • New Big Data Value companies that do not yet exist but will emerge in the near future 			

<ul style="list-style-type: none"> • Researchers and academics who can provide knowledge and thought leadership (data interoperability, data management, pattern recognition, model based techniques) 		
Sub groups/task forces (if applicable). If valid document each on second page		
ID	Subgroup Name / Activity	
	Research and Standardisation	
	Demonstrators	
	Experimentation	
Plan and Expected outcomes (translate into deliverables or other measurable outcomes)		
ID	Title + 1 Line Description	Date
	See subgroups	
Other data of interest		
This task force is related to BDV PPP priorities		

Subgroups

Sub Groups ID/Name (Repeat as necessary)			
TFX.SG1	Research and Standardisation (ITI/Rubén de Juan)		
Description of the activity (content) and expectations			
<p>This subgroup will identify research topics dealing with data interoperability and integration issues at Big Data scales (i.e. featuring high Volume, Variety, Velocity and Validity requirements), and within different application domains. The research will be oriented towards the design of a middleware layer providing different processing tools, methods and algorithms for data integration, analysis and knowledge extraction at different abstraction levels (operational, tactic, and strategic). This middleware will be used by higher level layers and their applicability demonstrated in real-world settings.</p> <p>This subgroup will also implement activities oriented towards the standardisation and adoption of a common data model for data integration and exchange to be used in Big Data settings (high Variety, high Velocity, and/or strict Validity requirements).</p> <p>Topics will be horizontal and will be tested in several use cases.</p> <p>Tasks to be performed:</p> <ul style="list-style-type: none"> • Definition of a medium-term (3-years) roadmap with the strategic activities to perform and a short-term (annual) action plans with the implementation activities • State-of-the-art and state-of-the-market: <ul style="list-style-type: none"> • Mapping of Big Data 4Vs challenges in the field of Enterprise Interoperability. • Identification of current data standards for enterprise interoperability and degree of industrial adoption. • Analysis of missing gaps and open issues concerning Big Data challenges. Taxonomy of methods, techniques and technologies for data models integration and interoperability support • Definition of industrial use case scenarios from different sectors reflecting the challenges posed by Big Data integration and interoperability issues. • Elaboration (or adoption) of a technical proposal for the standardisation of a common data metamodel for facilitating rapid data integration and interoperability from different sources and formats. • Promotion activities with relevant standardisation bodies, committees and working groups to gather feedback, influence upcoming standards and propose results 			
Plan and Expected outcomes (translate into deliverables or other measurable outcomes)			
ID	Title + 1 Line Description	Date	Status

	Prioritized list of research topics for Big Data Interoperability and Integration		
	Description of industrial use cases representative of the Big Data interoperability and integration challenges		
	Technical proposal for the standardisation and promotion activities		

Sub Groups ID/Name (Repeat as necessary)			
TFX.SG2	Demonstrators (XX)		
Description of the activity (content) and expectations			
<p>This subgroup will identify sectors where data integration and interoperability is an open issue in terms of Big Data VVVs (Volume-Variety-Velocity), and on which Big Data techniques and methods could produce a high impact. Within the sectors identified, its relevant actors will be identified looking to set up lighthouse projects to show the impact of Big Data solutions on their challenges.</p> <p>Tasks to be performed:</p> <ul style="list-style-type: none"> • Definition of an action plan • State-of-practice: (1) Identification of sectors with Big Data integration and interoperability issues; (2) Characterisation of challenges in terms of Big Data 4Vs (Volume, Variety, Velocity, Validity). • Demonstrators building: (1) Analysis and prioritization of representative sectors; (2) Identification and contact with relevant actors/roles; (3) Structure and setting of (3) Big Data demonstrators 			
Plan and Expected outcomes (translate into deliverables or other measurable outcomes)			
ID	Title + 1 Line Description	Date	Status
	Prioritized sectors and challenges identified		
	List and structure of demonstrator projects		

Sub Groups ID/Name (Repeat as necessary)			
TFX.SG3	Experimentation (XX)		
Description of the activity (content) and expectations			
<p>This Subgroup will define the components/tools, data and techniques to be integrated in an experimentation space where SMEs and entrepreneurs can test solutions making use of an integrated ecosystem for developing big data applications.</p> <p>Tasks to be performed:</p> <ul style="list-style-type: none"> • Definition of an action plan • State-of-market and practice: (1) Characterization of SMEs needs and requirements for big data applications; (2) Identification and mapping of Big Data methods, techniques, and tools suitable for SMEs. • Experimentation spaces building: Definition and setting up of (3) experimentation spaces 			
Plan and Expected outcomes (translate into deliverables or other measurable outcomes)			
ID	Title + 1 Line Description	Date	Status
	Characterization of SME needs and mapping of Big Data resources		
	List of experimentation spaces		

