DELIVERABLE

D2.1 - Pilot Coordination Methodology Handbook

Project Acronym	TT
Project Title	Transforming Transport
Grant Agreement number	731932
Call and topic identifier	ICT-15-2016
Funding Scheme	Innovation Action (IA)
Project duration	30 Months [1 January 2017 – 30 June 2019]
Coordinator	Mr. Rodrigo Castiñeira (INDRA)
Website	www.transformingtransport.eu

Funded by the European Union's H2020 GA - 731932

Document fiche	
Authors:	Andreas Metzger [UDE] – editor; Oscar Corcho [UPM], Florian Flocke [Fraunhofer], Víctor Rodríguez [UPM], Roberto Spigolon [CEFRIEL], Francisco Yedro [UPM] – contributors
Internal reviewers:	Zoheir Sabeur [IT Innovation] Marion Sanlaville [CINTRA]
Work Package:	WP2
Task:	T2.1
Nature:	R
Dissemination:	PU

Document History			
Version	**Date**	**Contributor(s)**	**Description**
0.5	20/01/2017	UDE	Initial document draft based on Madrid KOM presentations and discussions
0.7	03/02/2017	CEFRIEL, Fraunhofer, UDE, UPM	Refinement of document and addition of annexes 1–3, including Data ID Card
0.8	08/02/2017	UDE	Updated document, following TT template for review by TT PCC
0.9	09/02/2017	UDE	Updated document based on PCC feedback for internal review
1.0	22/02/2017	UDE	Final document for delivery to EC

Keywords:	Coordination, Methodology, Pilots, Pilot Domain, Incremental Approach, Design, Timing and Milestones, Project Structure, Replication and Sustainability, Validation and Scale-up
Abstract:	This deliverable constitutes the TT Methodology Handbook, which serves as reference for supporting the technical coordination of TT. This deliverable also provides a structured template for the common description of the pilot designs (provided as part of deliverables D*.1 in WP4–WP10).

ACKNOWLEDGEMENT

This document is a deliverable of TT project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 731932

Table of Contents

Table of Figures

Table of Tables

Definitions, Acronyms and Abbreviations

Acronym	Title
CO	Confidential, only for members of the consortium (including Commission Services)
CR	Change Request
D	Demonstrator
DL	Deliverable Leader
DoA	Description of Action
Dx	Deliverable (where x defines the deliverable identification number e.g. D1.1.1)
EU	European Union
MSx	project Milestone (where x defines a project milestone e.g. MS3)
Mx	Month (where x defines a project month e.g. M10)
O	Other
P	Prototype
PC	Project Coordinator
PO	Project Officer
PP	Restricted to other programme participants (including the Commission Services)
PU	Public
QA	Quality Assurance
R	Report
TC	Technical Coordinator
TL	Task Leader
TT	TransformingTransport
WP	Work Package
WPL	Work Package Leader
WPS	Work Package Structure

Executive Summary

This deliverable constitutes the **TT Methodology Handbook**, which serves as reference for supporting the technical coordination of TT. To this end, this deliverable describes the structure and organization of the TT pilots, as well as the key elements of the TT piloting methodology (its terminology, incremental approach, phasing and timing). Thereby, this deliverable may also serve as a primer for on-boarding new pilot members in order for them to understand the embedding of their respective pilot as part of the overall TT project. If needed, the TT Methodology Handbook will be updated during the course of the project to take into account lessons learned from actual piloting activities. Finally, this deliverable provides a structured template for the common description of the pilot designs (provided as part of deliverables D*.1 in WP4–WP10).

1 Introduction

The TT Methodology Handbook provides key information about the TT approach and methodology for piloting. All 13 TT pilots will comply with the TT approach and methodology. As such, this deliverable serves as a common set of guidelines and background information to all pilot members.

The remainder of this document is structured as follows:

- Section 2 explains the overall work breakdown structure of TT and the embedding of the pilots as part of this structure.
- Section 3 describes the key motivation and aspects of the TT piloting methodology
- Section 4 describes the concrete milestones and deliverables, implementing the TT piloting methodology
- Section 5 concludes this deliverable
- Annexes 1–3 describe the template for the pilot designs (delivered as deliverables D4.1 – D10.1 respectively)

The TT Methodology Handbook complements the information provided in the Project Management Handbook (D-1.1), which includes information on overall administrative, quality assurance and project management issues.

2 TT Structure

TT covers **13 pilots** in **seven highly relevant pilot domains** (see Figure 1). For each pilot, TT will explore innovative use cases and will engage key players in the sector in order to demonstrate the transformative nature that existing and very-near-to market big data technologies can bring about.

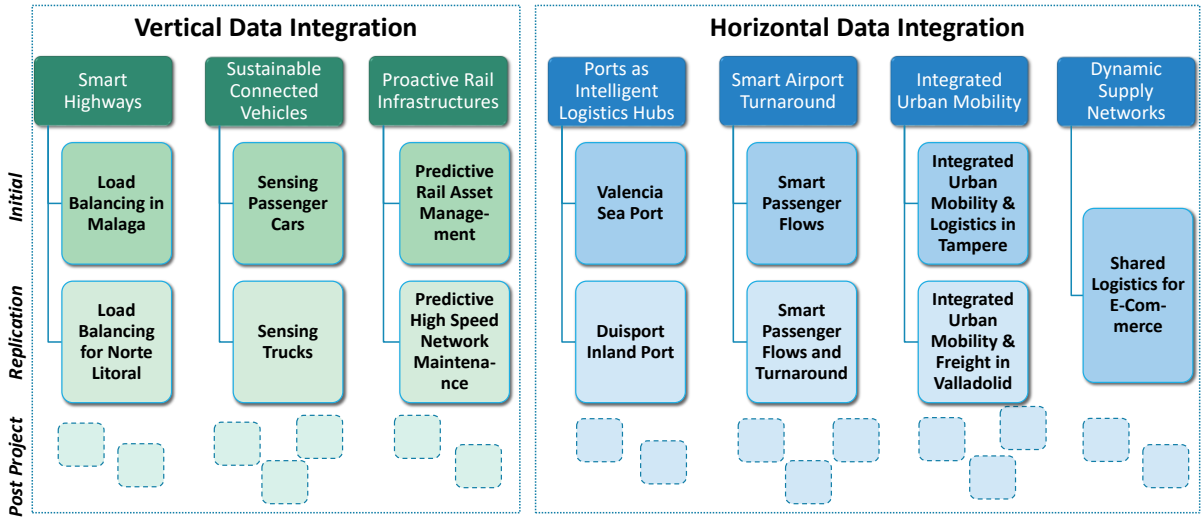

Figure 1: TT's 13 pilots in seven pilot domains

Each of the pilots will employ dedicated data analytics solutions best suited for the specific pilot requirements and datasets. In addition, each pilot will exploit big data infrastructures, which are best suited and closely linked with the data sources of the pilot. The underlying rationale for not prescribing a single data analytics solutions for all pilots, but allowing the pilots to define their specialized data analytics solutions is the "no free lunch" theorem that emerged in the field of machine learning and analytics[1]. As each data set, domain, and use case may be different, using a single data analytics solution will most probably not work (optimally) in all cases.

2.1 Synergies and Collaboration among Pilots

The pilots will not work in isolation but will strongly collaborate and cross-fertilize each other. Starting from common requirements of the pilot domains (such as data quality, data protection and scalability of data processing), experience in applying Big Data technologies and data assets will be shared among pilots.

In particular, pilot coordination and exploiting synergies will happen at three levels and will each be managed by dedicated and clear roles:

- **Within the individual pilots** to ensure vertical / horizontal data integration (managed by the respective pilot leader = task leader),

[1] e.g., see David Wolpert, William G. Macready: No free lunch theorems for optimization. IEEE Trans. Evolutionary Computation 1(1): 67-82 (1997)

- **Among the pilots within a pilot domain** to identify common requirements and solutions for the domain (managed by the respective pilot domain leader = work package leader),
- **Among and across all pilots and pilot domains** to distil cross-cutting requirements of the mobility and logistics sector and to demonstrate generic solutions.

For the third level of coordination, the following aspects of coordination and synergies will be given particular attention and each have their dedicated tasks (and task leaders) as part of the TT work plan:

- **Requirements analysis and lessons learned** (Task T2.2): Assessing the potential and the requirements for Big Data across the TT pilots facilitates effective knowledge exchange on common concerns, and allows merging lessons learned for post-project replication.
- **KPI achievement and impact generation** (Task 3.1): Based on a structured and shared KPI assessment framework, all pilots will follow a comparable approach to KPI definition and assessment.
- **Data Concerns:** Three in tandem working tasks within TT are in charge of data concerns, from three different, complementary points of view: legal / governance (management), integration and quality (technical), as well as open data and data sharing (impact).
    - **Data governance and data protection** (Task T1.3): To ensure sharing and access to data within TT, pilots will be supported by having a clear data governance and data protection scheme in place, allowing pilot members to handle required data correctly and to support their decision making about how a data source may be exploited in a pilot use case.
    - **Data quality, integration and homogenization** (Task T2.3): Common methods and procedures for data management will be shared across pilots. Thereby, each pilot can build on project-wide expertise. This includes (i) data quality procedures to support the pilots to understand the current status of data and the data quality needs, (ii) data integration methodologies to facilitate lightweight data links and real-time data aggregation, and (iii) data homogenization standards (including semantics-based approaches) when it is not possible to avoid data schema mapping.
    - **Open data and shared data** (Task T3.3): Where the previous aspect was concerned with methodological and technical concerns of data integration and quality within pilots, data assets of TT may also be shared across pilots and beyond TT. To determine whether there may be an opportunity – on a case by case basis – for data sharing, meta-data about each of the TT data assets will be collected in structured form (cf. the Data ID Card defined in Appendix A1.2).

2.2 Benefits of TT Structure

The aforementioned TT structure has the following main benefits when considering the size and scale of the TT project (which involves 47 organizations and over 1800 PMs of technical work):

- **Clear responsibilities:** Each of the 13 pilots can be considered and managed as a small, individual project as part of the overall TT "programme". Each of these pilot "projects" has a manageable size and number of involved members (on average 140 PM and 4 members). Also, each pilot lead is responsible and also best experienced for the successful execution and management of his/her pilot "project".
- **Clear hierarchy:** Handling of risks and problems happens in a hierarchical fashion. Detailed risks and problems that cannot be handled within the pilot are escalated to the pilot domain (WP). In case risks and problems cannot be handled on the level of the pilot domain, they are escalated to the Project Coordination Committee, and for ultimate resolution to the TT Technical Coordinator (TC)
- **Clear clustering of competencies:** The work in the pilots and pilot domains has been clustered along the transport sectors domains and use cases to exhibit as few dependencies as possible, while still delivering their objectives. Specifically, many of the TT transport domain partners and data owners have been assigned to one pilot only and have a concise and clearly defined role in the pilot. This clear clustering facilitates reduction of communication overhead (not everybody has to be involved in everything), while still enabling intense alignment within the pilot domain, and less intense alignment across pilot domains.

3 TT Methodology

TT follows an incremental, iterative approach to design, develop and validate the big data solutions with the pilots. Such incremental, iterative approach has the benefit of incorporating learning effects, reacting to recent insights and technology trends, as well as exploit new outcomes and insights from relevant, related BDV PPP projects.

TT will follow a 3-phase replication approach within each pilot domain combined with a 3-stage methodology for validation and scale-out within the phases.

3.1 3-Phase Pilot Replication and Sustainability Approach

As shown in Figure 2, the pilot domains will start with an **initial pilot** at the beginning of the project, and then replicate the solutions by reusing the results as part of a **replication pilot**[2].

[2] Due to the more complex horizontal data integration to be performed for the E-Commerce pilot, we refrained from pushing for a replication pilot due to the time constraints of the project.

The replication pilot considers insights, findings and lessons learned from the initial pilot. This replication approach is one key means to demonstrate the reusability and generic nature of the TT solutions. A replication pilot addresses similar and related objectives as the initial pilot, but typically adds a further level of complexity; e.g., in terms of processes or data assets. In addition, as indicated in Figure 2, TT will prepare a **post-project replication phase** to ensure **sustainability of project results** (see Task T3.6).

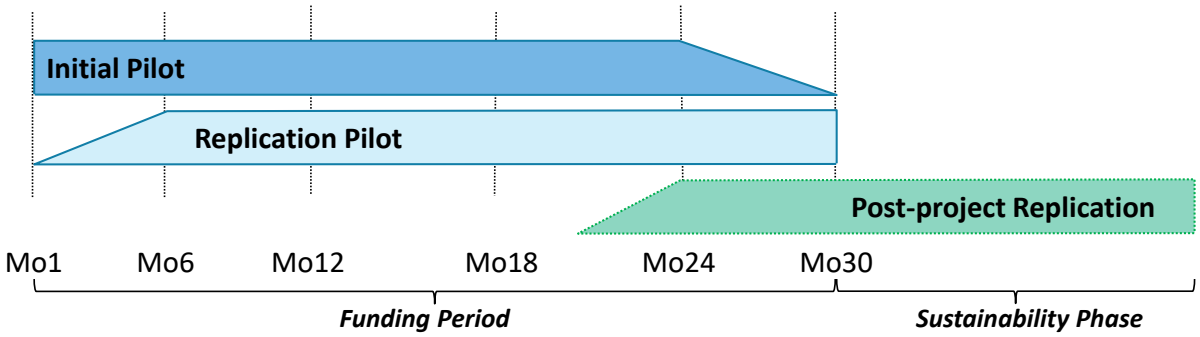

Figure 2: Initial and replication pilot approach

3.2 3-Stage Validation and Scale-up Methodology

Within each pilot, TT follows a **3-stage methodology** for validation and demonstrating the scalability of Big Data solutions. The three stages differ with respect to the embedding of the technology into the operational environment, the form of the big data infrastructure being used, as well as the scale of data that is exploited. Starting with a focussed technology validation, the 3-stage methodology will ultimately deliver insights from actual in-situ trials, i.e., a prototypical implementation of the solutions in the field, involving real end-user and actual production data. The three stages of the methodology along its three main aspects are explained in Table 1.

Table 1: TransformingTransport 3-Stage Validation and Scale-out Methodology

Stage	Embedding in Operational Environment	Big Data Infrastructure Used	Scale of Data
S1: Technology Validation	Understanding problems and potentials for Big Data by analysing concrete aspects and validating key solution ideas and concepts	Local, existing infrastructure small-scale infrastructure used for exploratory experiments	Selected (historic) data pinpointing problems and opportunities

Stage	Embedding in Operational Environment	Big Data Infrastructure Used	Scale of Data
S2: Large-scale experimentation and demonstration	Large scale experiments in (controlled) environments, i.e., decoupled from productive environment	Dedicated large-scale data processing infrastructure for experimental purposes	Large scale historic and real/live data, possibly anonymized or simulated (for controlled experimental conditions)
S3: In-situ (on site) trials	Trials in the field, involving actual end-users	Actual data processing infrastructure of pilot partners	Real, live production data complemented by large scale historic data

4 Pilot Milestones and Deliverables

The timing and outcomes of the pilots follow the aforementioned structure and methodology. During the execution of TT, the pilots will deliver incrementally more refined and concrete outcomes.

4.1 Key Types of Pilot Outcomes

The main types of outcomes delivered by the pilots, and which will be reported in the pilot deliverables (see below), include:

- **Requirements** express the needs towards the TT big data solutions from the side of the (end-)users of the transport domain; e.g., *enable terminal operators to better manage terminal and port operations in real-time*.
- **Objectives** express how the pilots envision addressing and satisfying the requirements (i.e., describe the envisioned solutions); e.g., *develop a terminal productivity cockpit with predictive and prescriptive analytics capabilities*.
- **Use Cases and Scenarios** describe typical usage and application scenarios. They show in concrete terms ("story book" form) how the technical TT solutions will be employed by the relevant stakeholders to achieve their stated requirements.
- **KPIs (Key Performance Indicators)** provide metrics and target values to assess in how far requirements are met by the pilot solutions; e.g., *10% improvement in terminal efficiency*.

- **Demonstrators** are "prototypes" and components that facilitate experimentation, showcasing, measuring and demonstrating the pilot solutions. Ideally, demonstrators exemplify and realize the typical use cases and scenarios identified for the pilot.
- **Lessons Learned** analyse and document the findings of applying big data technology to the pilot requirements. They serve as reference for post-project replication to understand the benefits but also the limitations of big data technologies in the target sector of the pilot domain.

4.2 Timing of Pilots

Table 2 shows the milestones and deliverables relevant for the TT pilots. Column #1 gives the milestone number from the DoA, column #2 gives the project month of delivery. Columns #3–4 describe the main outcomes of the initial and replication pilot respectively, while column #5 gives the deliverable which subsumes these outcomes (this flow of outcomes into deliverables is also indicated by the arrows). Finally, column #6 describes which types of outcomes (according to Section 4.1) will be delivered for the given milestones.

Table 2: Milestones and how they lead up to pilot deliverables

Milest.	Mo	Initial Pilot	Replication Pilot	Deliverables	Type of Outcomes
MS2	3	Pilot Design	Pilot Design	**Pilot Designs [D*.1]**	**Requirements, Objectives, Use Cases & Scenarios**
MS3	6	Results Stage S1			
MS4	9		Results Stage S1	**Performc. Assess. Plan [D*.2]**	**KPIs**
MS5	12	Results Stage S2			
MS6	15		Results Stage S2		
MS7	**18**			**Pilots Release 1 [D*.3] - DEM**	**Demonstrators**
MS9	24	Results Stage S3			
MS10	27		Results Stage S3	**Pilots Release 2 [D*.4] - DEM**	**Demonstrators**
MS11	**30**	Lessons Learned	Lessons Learned	**Big Data Impact [D*.5]**	**Lessons Learned**

MS7 (M18) and MS11 (M30) represent the end of the respective project reporting periods and thus constitute the timing of the two project reviews.

A more detailed description of the structure and contents of the Pilot Design deliverables (first pilot deliverable in M3 of the project) is provided in Annex A1.1 of this document.

4.3 Evaluating Pilot Performance (Flow of KPI-related Information)

KPIs are a key instrument in TT to demonstrate in measureable terms the impact and contributions of the TT solutions, i.e. to evaluate pilot performance. To this end, defining and measuring KPIs follow a clearly-defined flow along the duration of the project and involving the key relevant WPs and tasks.

At the time of writing the overall process and flow of KPI-related information is clearly defined. As part of this process, concrete guidelines will be defined to guide pilots on how to evaluate pilot performance. In particular, the definition and assessment of KPIs will be guided by cross-cutting tasks of TT, in particular Task T2.2 ("Pilots Requirements Analysis & Lessons Learned") and Task T3.1 ("Joint Exploitation, Market Impact & KPI Achievement"). Thereby, KPIs common to all pilots will be identified early on, thereby also contributing to finding commonality among pilots.

The process and flow of KPI-related information is expressed below based on the order in which the KPI-related information is provided in the TT deliverables. Steps 1–3 contribute to the definition of KPIs, whilst steps 4–7 contribute to the measuring and assessment of the defined KPIs:

1. **D*.1 - Pilot Designs (for each pilot domain) [M3]** provide the definition of pilot requirements (e.g., improvement in efficiency, usability, performance, customer satisfaction, enabling new business models, …) and the specification of the pilot objectives to address these requirements.
2. **D2.2 – Pilot Requirements Analysis [M4]** define domain-specific KPIs and their respective baselines for each of the different pilots as mechanisms to assess in how far requirements are met by the pilots.
3. **D3.8 – KPI Assessment Framework [M6]** Complements the pilot KPIs with project KPIs (such as impact, leverage, community) and also defines the overall approach for KPI measurement and assessment to be followed by the pilots.
4. **D*.2 – Performance Assessment Plan (for each pilot domain) [M9]** explains concretely how pilot-specific KPIs will be measured and assessed (making use of the KPI assessment framework methodology from D3.8).
5. **D3.11 - Period 2 Impact & KPI Achievement Report and Plan [M18]** includes a first report on pilot KPI measurement and assessment.
6. **D3.12 - Period 3 Impact & KPI Achievement Report and Plan [M30]** includes a second report on pilot KPI measurement and assessment.
7. **D*.5 – Big Data Impact in Pilot Domain Report [M30]** includes a final report of lessons learned and assessment of pilot KPI measurement and assessment.

5 Conclusions

This document provides the key concepts, methodology and timing relevant for TT pilot execution and coordination. It thus serves as a concise reference for TT pilot leads, members and new personnel joining the TT project.

Appendix 1. Template for Pilot Design Deliverables (D*.1)

The sections below provide a common template across all pilot domains for describing the pilot designs, which are delivered as D*.1 in M3 of the project.

A1.1 Deliverable Structure

Each pilot design deliverable will provide the following key sections:

1. **Motivation and Ambition**

   – provide a general motivation for the pilot domain (e.g., in terms of needs and opportunities for efficiency improvement etc.) and sketch the overall envisioned ambition (situation at the end of TT) of the pilot domain

2. **Design of Initial Pilot**

   – provide a brief overview of the pilot and a brief explanation of what follows

   2.1. **Requirements** – *see definition in Section 4.1*

   2.2. **Objectives** – *see definition in Section 4.1*

   2.3. **Use Cases and Scenarios** – *see definition in Section 4.1*

   2.4. **Data Assets** – *please use the below table; also see template for defining data assets (Data ID Card) as defined in Appendix A1.2*

Name of Data Asset	Short Description	Initial Availability Date	Data Type	Link to Data ID Card (in basecamp)
Name 1				
Name 2				
…				

   2.5. **Big Data Technology, Techniques and Algorithms** – *provide a detailed, in-depth description of the technology planned to be developed and deployed in your pilot (going beyond the level of detail provided thus far in DoA-Appendix A.7) – also indicate whether open source or commercial*

   2.6. **Positioning of Pilot Solutions in BDVA Reference Model** – *see the explanation in Appendix A1.3*

2.7. **Big Data Infrastructure –** *provide a detailed explanation of the infrastructure (middleware, storage and processing capacity, etc.) that will be used and its characteristics (e.g., also regarding scalability and data protection)*

2.8. **Roadmap** – *define what will be available as results of each of the three Stages 1-3 – please use the below table (see Section 3.2 for an explanation)*

Stage	Delivery Date (Project Month)	Features / Objectives Addressed	Embedding in Operational Environment	Big Data Infrastructure Used	Scale of Data
S1: Technology Validation					
S2: Large-scale experimentation and demonstration					
S3: In-situ trials					

3. **Design of Replication Pilot** – *same substructure as Pilot 1*

4. Commonalities and Replication

4.1. Common Requirements and Aspects – *here would be the place to indicate general aspects among the pilots; e.g., common requirements are requirements which are addressed by both the initial and replication pilots*

4.2. Replication – *explain what and how of the initial pilot will be replicated*

5. Conclusions

– conclude the document with a general summary and provide some brief perspectives (e.g., indication of forthcoming deliverables)

A1.2 Data Assets and API ID Card

Each data asset together with the APIs for accessing the respective data asset will be described following a structured template, the "**Data Assets and API ID Card**" (**Data ID Card**[3] for short).

The Data ID Card is a XLS document that will be created for ***each*** data asset and may be updated and maintained during the course of TT. The Data ID Card will also serve as central information source for the project-wide data management and governance activities.

The pilot design deliverables (D*.1) should describe in Sections 2.4 and 3.4 respectively the data assets of the pilot, in a brief table (see above). This table should then give for each data asset a link to the concrete Data ID Card XLS file in basecamp. Thereby, the confidential information about the data assets may not be visible in the public pilot design document.

The Data ID Card consists of the following sections with the respective information elements, relevant for the pilot designs:

- **General Information**
  - Name of the Data Asset / API
  - Internal name of the Data Asset / Data Asset / API
  - Name of the Data Asset / API Provider
  - Short Description
  - Extended Description
  - Version
  - Initial Availability Date
  - Data Type
- **Legal Information**
  - Personal data
  - License
  - Rightsholder
  - Other Rights Information
- **Contacts**
  - Data Asset/API Owner/Responsible
  - Data Asset/API Owner/Responsible contacts
- **Technical Documentation**
  - Technology
  - Name of the System
  - Data Asset data model/ API Interface

[3] Available to TT members from https://3.basecamp.com/3320520/buckets/1429164/vaults/338367019

  - Data Model - Standard Glossaries
  - Data Identifier - Standard used
  - Data Model - Specific Data Model
  - Data volume
  - Update frequency
  - Data Archiving and preservation
- **Access and Security**
  - Access Level
  - Access Mechanism
- **Data Quality KPIs**
  - (This will be defined later and will be used for T1.3, T2.3 and T3.3)

A1.3 BDVA Reference Model

TT is part of the Big Data Value PPP (e.g., see http://www.bdva.eu/) and as such will benefit and contribute to collaboration among related PPP projects. To provide a common framework for understanding potential points of common interests among PPP projects, the so called Big Data Value Reference Model will be employed as part of the technical collaboration activities.

To this end, each TT pilot is asked to position its solutions in the Big Data Value Reference Model. Ideally, each of the solutions developed is given a unique number in the deliverable, which is referred to in the description of the solution (i.e., Section 2.5 or 3.5 in D*.1 respectively). This number should then also be used for positioning the solution graphically in the below visualization of the Big Data Value Reference Model[4]. The BDVA Reference Model covers the most important Big Data technical areas (shown horizontally). The BDVA Reference Model also covers key cross-cutting concerns, such as data protection, cyber security, development and operations, and standards (shown vertically).

Figure 3 depicts the Big Data Value Reference Model and illustrates an abstract pilot contributing to 2D visualisation of predictive analytics solutions based on streaming analytics, where data linking is done using existing solutions.

[4] Available as editable version (PPT) to TT members from https://3.basecamp.com/3320520/buckets/1429164/vaults/338367019

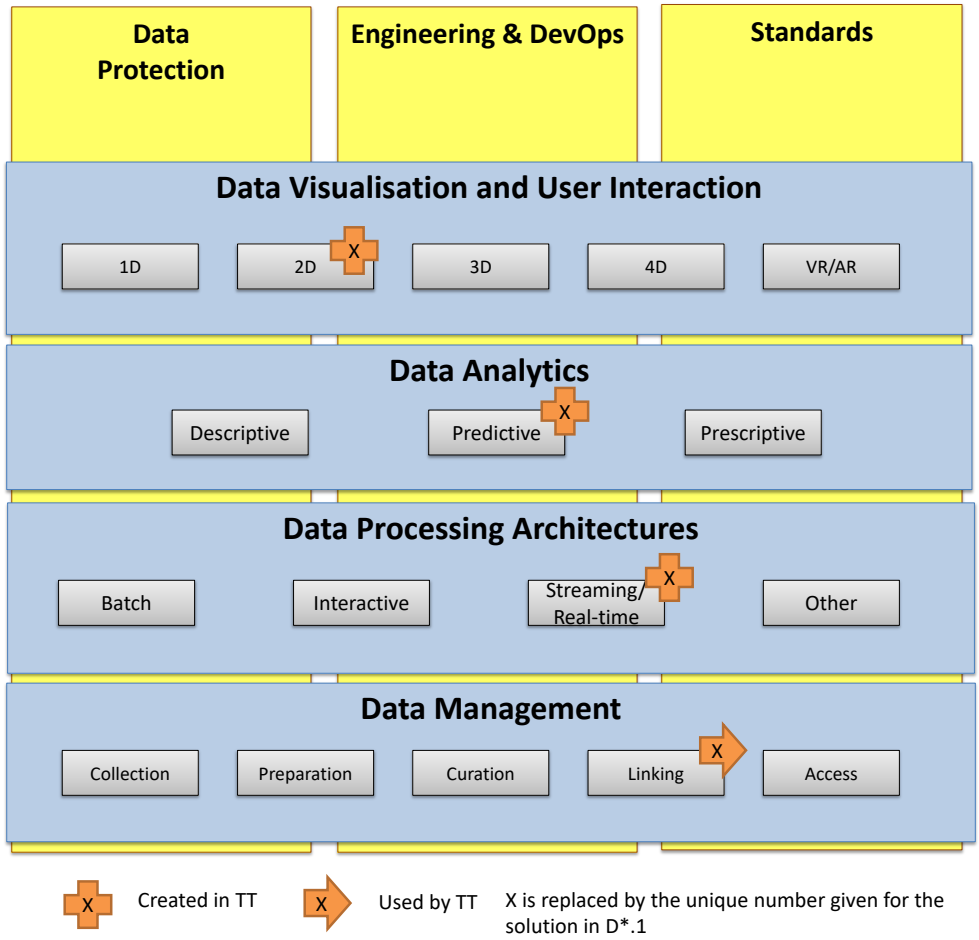

Figure 3: Big Data Value Reference Model