

DigLogs

Analyses of the most attractive BIG DATA tools and practices already on-going and directly transferable in the ITA-CRO area

Deliverable 3.2.2

Responsible partner: Elevante s.r.l. and University of Trieste			
Involved partners: n/a			
Version	Status	Date	Author
0.1	Draft	30.08.2019.	Elevante and UNITS
0.2	Draft	02.09.2019	PFRI
0.3	Draft	30.08.2019.	Elevante and UNITS
Notes:			
DISCLAIMER <i>This document reflects the author's views; the Programme authorities are not liable for any use that may be made of the information contained therein.</i>			

Table of Contents

Introduction.....	2
1. Big-data processes: best Practice Analyses of already on-going innovations.....	4
1.1. Data standardisation: the upcoming role of Digital Container Shipping Association (PP2) 5	
1.2. Machine Learning applications for logistics (PP2, Elevante)	7
1.3. Big Data Analytics (LP)	9
2. Identifying most attractive directly transferable innovations: the Questionnaire	11
2.1. Results.....	13
2.1.1. Global Analysis.....	13
2.1.2. Individual partner results	18
Conclusions.....	26

Annexes: Diglogs Questionnaire Part 2

Introduction

The purpose of this deliverable is to select and extensively analyse the most promising innovations related to big-data trends among the ones gathered in D3.1.2. The activity has been carried out by Elevante Srl and UniTS involving all the other PPs.

In this Deliverable the most attractive innovations already on-going and directly transferable in the Italy-Croatia area in the prospective three years are analyzed. To achieve overall goals, a straightforward methodological approach is developed, which is composed by two streams of activities.

First, a best practice analysis is conducted on a number of already-on-going innovations. The results are shown in the form of “cards”, reporting an in-depth analysis of each selected innovation. All PPs have been asked to collaborate and carry out at least one analysis each, in order to define the state of the art and the most recent developments regarding the innovations that can be applied in the next three years in the project area.

Secondly, in order to get an effective assessment about the transferability of available innovations, PPs are involved to identify priorities. The overall idea is to move from the level of disruptive innovations to pilot actions in a smooth and coherent manner. In particular, an on-line questionnaire was developed, in which each PP is asked to identify priorities by ranking the innovations identified in the previous Deliverables on the basis of two criteria, e.g., their importance and easiness of implementation.

The questionnaire was jointly prepared by the WPL and the University of Trieste in coordination with the LP. It was submitted to PPs in July and results were collected and analysed at the end of August in order to map the importance of the innovations described in D3.1.2 and the easiness/difficulty expected in their deployment.

Overall results are presented for each trend both in aggregate – e.g., for all the ITA-CRO area – and with respect to each PP, reporting the final rank of priorities based on the above mentioned criteria, thus, the identification of directly transferable innovations. A both descriptive and graphical representation is produced.

The results from the questionnaire have been analysed by Elevante S.r.l. and UniTS providing a global analysis and a single partner analysis. The former offers some insights about the big-data exploitation in Adriatic Region, while the latter provides results highlighting the priorities of each partner involved in the Diglogs project, some of which will be tested in subsequent pilot actions.

1. Big-data processes: best Practice Analyses of already on-going innovations

In the present section best practices concerning some of the most important innovations connected to big data are discussed. The best practice analysis is essential to provide a starting point for innovation implementation within the Adriatic Region, i.e. the project area. All PPs have been asked to collaborate and carry out at least one analysis each, in order to define the state of the art and the most recent developments regarding the innovations that can be applied in the next three years in the project area. In addition to those provided by PPs – among all 3 innovation trends, namely: Informatisation, Big Data/Data Analysis and Automation systems – WPL (Elevante Srl, PP2) has carried out 5 analyses, PP4 (UniTS) has done 3, while the LP has provided a document that also included several analyses. Within the Big Data/Data Analysis innovation trend, a total of 3 already on-going innovations have been investigated and presented in the following paragraphs.

1.1. Data standardisation: the upcoming role of Digital Container Shipping Association (PP2)

It is evident that rising levels of digitization worldwide have prompted the significant, and in some cases exponential, growth of global industries, and the maritime sector is one of many business areas which can stand to benefit from this development. There are port authorities, terminal operators and shipping companies which are already leveraging data to intelligently plan their operations, meet demand and ensure the security of their staff and cargo.

The wave of innovation technology development in shipping and logistics industry over the recent years has brought good opportunity to the whole industry for digital transformation. But, at the same time, individual companies are cautious about adopting new technologies since there is no common standard in the market.

Digital standards are now a priority in order to create value quickly and overcome the lack of a common foundation for technical interfaces and data.

On this account, in April 2019 four of the world's largest carriers (Maersk, Hapag-Lloyd, Ocean Network Express and MSC Mediterranean Shipping Company), have established the Digital Container Shipping Association (DCSA) with the aim of creating common information technology standards that will improve the overall efficiency of the shipping sector. The association is not for profit and has no intent of developing or operating any digital platform, but aims to ensure interoperability through standardization. Similarly, the association will not discuss any commercial or operational matters

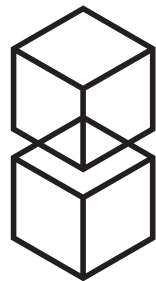
Later this year five more ocean carriers have joined DCSA as members: CMA CGM, Evergreen Line, Hyundai Merchant Marine, Yang Ming Marine Transport Corporation and ZIM Integrated Shipping Services.

DCSA is helping the industry define the baselined process standards from Booking to Return, including sub-processes, milestones, events and messages.

These process standards support a common view across the industry in relation to processes, milestones, events and messages, facilitating industry standardization and digitization efforts.

The results are publicly available, open-source, thus everybody has access to the process standards within the Industry Blueprint. The associations also fosters shared information technology standards relate to transmitting, receiving and exchanging data across the industry for standardizing data, message formats and interface specifications.

A previous case of major carrier joining together for similar reasons is INTTRA, a neutral digital platform that currently handles one out of every four container bookings, started years ago as a joint industry effort.



**DIGITAL
CONTAINER
SHIPPING**
ASSOCIATION

1.2. Machine Learning applications for logistics (PP2, Elevante)

Forecasting is vital in supply chain logistics. Predictions start when a vessel leaves Port A for Port B. At what date and time will it call the arrival port? How long will it take for custom duties to be cleared? How many trucks should be arranged for picking up loads? At what time?

All the questions and more are affected by decisions made earlier on and each decision is based on data about the individual logistic tasks, that can be represented as links in the chain.

In the past, predictions were made by professionals on the fly. Today, that is not feasible because too many variables are at play. Artificial Intelligence and machine learning can recognize patterns in the logistics chain that humans would not. The ships are larger and more frequent, the cargo exponentially increased in quantity and movements. There are many terminals and many ports to correctly remember and account for variations that can lead to delays.

We have lots of data: we have tagged and networked everything possible, from container weights and dispatch dates to vessel speed and weather conditions – all in real time. This raw data is then cleaned up and processed, which is where machine learning and artificial intelligence can take up from where human mind cannot bear the information load.

ClearMetal is a predictive logistics company that uses AI to unlock supply chain efficiency. They describe *predictive logistics* as using sophisticated technology to solve the most complex operational problems in shipping and logistics. The objective is to help agents in the supply chain – brokers, truck companies, terminals – cope and learn effectively from the complexity of the system.

One of ClearMetal's tools is Predictive Risk calculation. This helps freight operators understand the likelihood of a shipment moving from "On Time" to "Late" by using simulations. Hundreds of simulation models are generated per shipment to provide the probability of a status change. Once the data is fed in to the platform, the result might be: "87% of 130 simulations predict this shipment will arrive late".

Comparative Predictions of Container Dispatch Quantity



Figure 1: prediction via ClearMetal's machine learning vs forecasting models

PSA Marine, a marine services provider, awarded a contract Ernst & Young Solutions (EY) in Singapore to design and develop the optimization application called Blue 5.0. PSA Marine has claimed that Blue 5.0 will help vessel operators and vessel charterers estimate the total duration of voyage within the port of Singapore more accurately.

CargoSmart, the Hong Kong freight logistics service, accesses a database of every route offered by the world's 30 biggest carriers. It also has five years' worth of automatic identification system (AIS) vessel movement data. The system can predict at any moment if a vessel is likely to arrive on schedule.

Maana, an American analytics software company, is mathematically modelling the process that goes into re-routing a vessel in the event of a port closure or congestion. Using traditional methods, this could take several hours to evaluate the potential of an alternative port. With ML, it can be done in minutes.

1.3. Big Data Analytics (LP)

New cloud computing solutions will make data instantly and simultaneously accessible in many locations and across many devices. This massive amount of data requires the collection, curation, analysis and storage of large and complex datasets. This is often defined as Data Analytics.

Large amount of data are direct consequence of the Internet of Things (IoT), the upcoming paradigm relating to networking of various applicative physical devices ("things"), as opposed to a more traditional arrangement situation where networking refers primarily to computer and network devices and peripherals. "Things" are embedded with electronics, software, sensors and connectivity that enable them to achieve functional value and exchange data with other devices and systems.

Typical applications of IoT are various sensors or transponders used to ensure quality of shipment conditions (monitoring of vibrations, container openings or cold chain maintenance for insurance purposes), item location (search of individual items in big surfaces like warehouses or harbors), storage incompatibility detection (warning emission on containers storing inflammable goods closed to others containing explosive material) and fleet tracking.

The ever more rapid development of cheap low consumption sensors has resulted in ever more "items" being equipped with such sensors. This effectively means all such items can be tracked and that any activity such item is engaged in, or any circumstances it is exposed to, can be "measured". Thus, the item "senses" an activity, event or an environmental factor. Such item is also capable of receiving information from other "sense-like" items.

Effectively, the IoT refers to a wide and increasingly large range of physical objects ("things"), that are connected to a system and that can send and receive data.

After discussing the "sensing" of data, collecting and storing them, any use that can be made is understood as Data Analytics.

Data will be processed to facilitate new decision-making processes or processes. It can be used in port operations such as preventive maintenance schedules of either infrastructure or equipment, create intelligent inspections systems, sensor track data on speed, direction and

driving performance of large numbers of vehicles in order to optimize future routes, or support resilience management tools in order to adjust routing of supply chains in real time.

The possibilities are almost endless and consequently, the evolution of IoT and the Data Analytics creates the prospect of logistics becoming a data-centric industry where information takes precedence in logistics services' value propositions over the actual ability to move cargo.

Presently, there are over 23 billion IoT connected devices worldwide. This number is expected to reach 30 billion by 2020. and over 60 billion by the end of 2025. Gartner's report predicts that by 2020. more than 65% of enterprises will adopt IoT products. However, the report also stats that a lack of data science specialists will inhibit 75% of organizations from achieving the full potential of IoT.

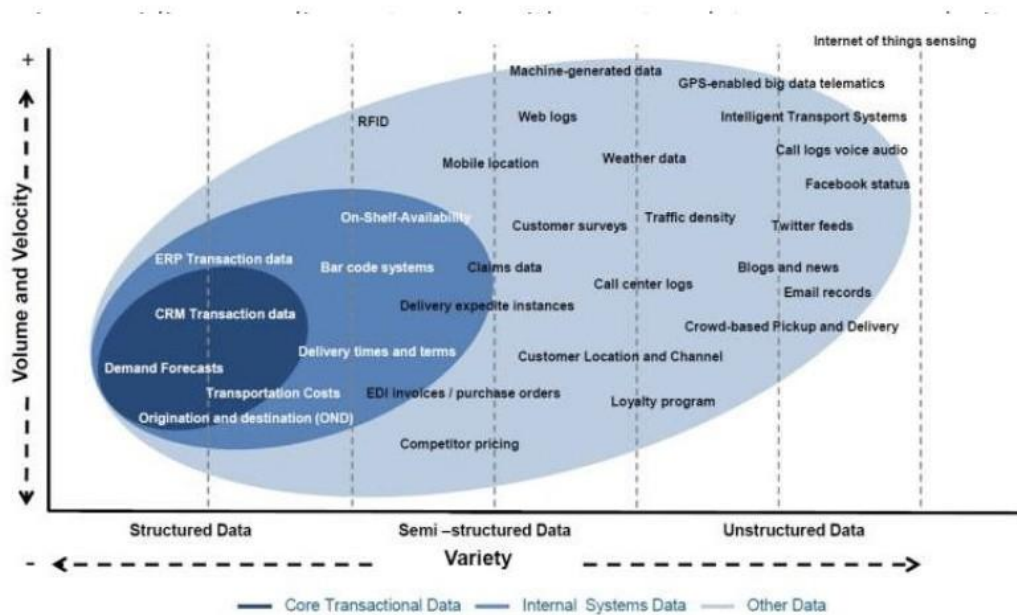


Figure 2: SCM Data Volume and Velocity vs. Variety

2. Identifying most attractive directly transferable innovations: the Questionnaire

A questionnaire has been designed and implemented as an online tool by means of the Google Forms platform. It has been submitted to all the PPs in order to collect data devoted to rank the innovations defined in previous deliverables by importance and analyse the issues related to each innovation deployment in the project area. The questionnaire is composed of three sections:

- General information
- Informatisation processes
- Big-Data
- Automation

In the general section, some information is collected regarding the respondent who is compiling the questionnaire. This information is useful to properly qualify the subject and his expertise. In detail, the following data have been collected:

- Name
- Surname
- Age
- Project Partner (the affiliation of the subject)
- Position (the position of the subject within the project partner)
- Seniority (years of service in the current position)

The other three sections deal with a specific innovation trend. Each subject was required to select the innovations he/she deems more relevant to them in terms of appeal and applicability to their local needs. In detail, each innovation was ranked per importance on a five steps scale:

1. not at all relevant;
2. not very relevant;
3. no opinion;

4. relevant;
5. extremely relevant.

Moreover, efforts required to implement the innovations in the current environment were also ranked according to the following scale:

1. very difficult to implement;
2. somehow difficult to implement;
3. no opinion;
4. somehow easy to implement;
5. very easy to implement.

PPs were required to briefly explain the reasons behind their choices.. They were also asked to suggest up to two additional innovations for each innovation trend. Finally, each subject was required to rate only innovations that apply to his context and field of expertise, in order to prevent the collection of inconsistent data.

2.1. Results

In this section, the results related to big-data innovation trends are provided. First, a global analysis coming out from the outcomes of all the collected questionnaires is provided. Then, the outcomes and ranking of the innovations are provided for every single partner.

2.1.1. Global Analysis

Overall, 24 responses have been collected from partners (see Figure 3).

The distribution of age and seniority of the subjects involved in the completion of the questionnaire are provided in Figures 2 and respectively. Since most of the partners are engaged in research activities, is not surprising the concentration on 0-5 years seniority, which is due to the dynamic labour market typical of such a sector. However, most of the low-seniority subjects are scholars or other specialists with strong experience. Moreover, all the other subjects involved in the questionnaire completion have a strong experience in their field since they cover the same position for a relevant time period (greater than 5 years).

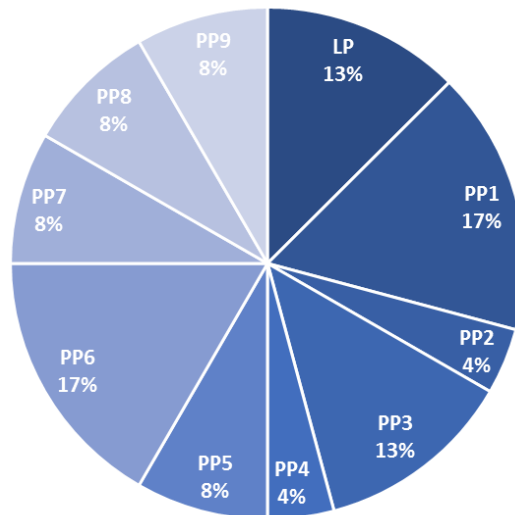


Figure 3 - Responses collected from PPs

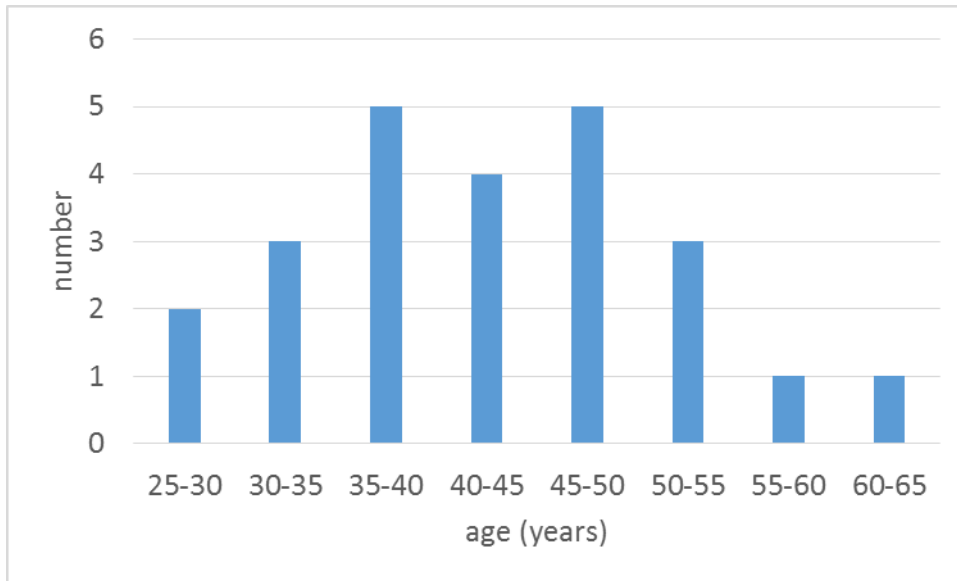


Figure 4 - Age of the subjects

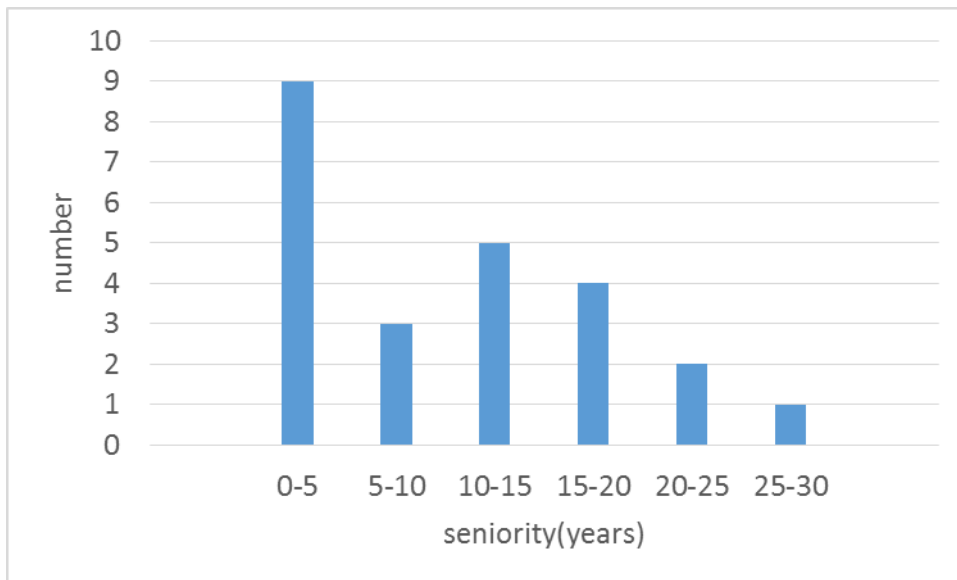


Figure 5 - Seniority (in current position) of the subjects

The innovations, ranked by importance I are shown in Table 1, together with their deployment difficulty/easiness D . Moreover, the standard deviation σ are provided for both quantities:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Table 1 - Global Rank of innovations

Rank	Innovation	I	σ_I	D	σ_D
1	Big data/data management	4.60	0.58	3.20	0.81
2	Data standardization	4.36	0.64	2.86	1.14
3	Big Data Analytics	4.30	0.64	3.00	0.77
4	Port traffic management	4.09	0.67	3.05	1.02
5	Loading/unloading optimisation	3.95	0.80	3.30	1.05
6	Automatic detection of logistics' level of service	3.89	0.81	2.68	0.80
7	Energy Efficiency	3.88	0.76	2.69	0.92
8	Anomalies detection & predictive maintenance	3.86	0.77	2.81	0.73
9	Machine Learning	3.75	0.62	2.60	0.92
10	Passengers flow analysis	3.70	0.71	2.95	0.86
11	Digital twin	3.52	0.91	2.35	0.91
12	Georeferenced data	3.48	1.10	2.85	0.85

Figure 5 provides a graphical representation of the results on a scatter diagram. It can be noticed that all the innovations are located in a restricted area in the upper part of the diagram. Figure 6 provides a more detailed view of such an area of the scatter diagram. Figures showed in Figure 6 refer to innovation ranking (Tab.1)., it can be concluded that no clear correlation exists between importance and deployment's easiness/difficulty. Moreover, the concentration in the upper part of the scatter diagram highlights how all the partners assigned a medium/high importance to all the innovations defined in D3.1.2, thus, confirming that the selection process was very effective.

However, the quite high value of standard deviations related to innovations' importance reduces the significance of the ranking. Nevertheless, a group of more interesting innovation

can be defined including *Big data/data management*, *Data standardization* and *Big Data Analytics*. Thus, the PPs assign globally a clear preference for data storing (database design/data management) and data standardization issues that are a prerequisite for the efficient application of data analytics. This preference suggests that the application of big-data technologies is just at the beginning in the project area, however there is a strong interest in the development of the infrastructures required for their exploitation.

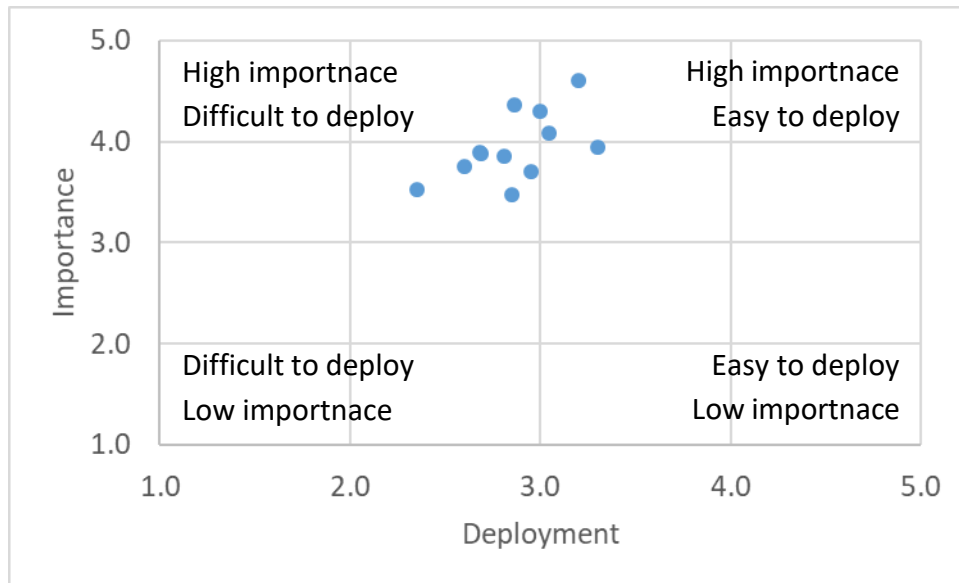


Figure 6 - Importance vs Deployment Easiness/Difficulty

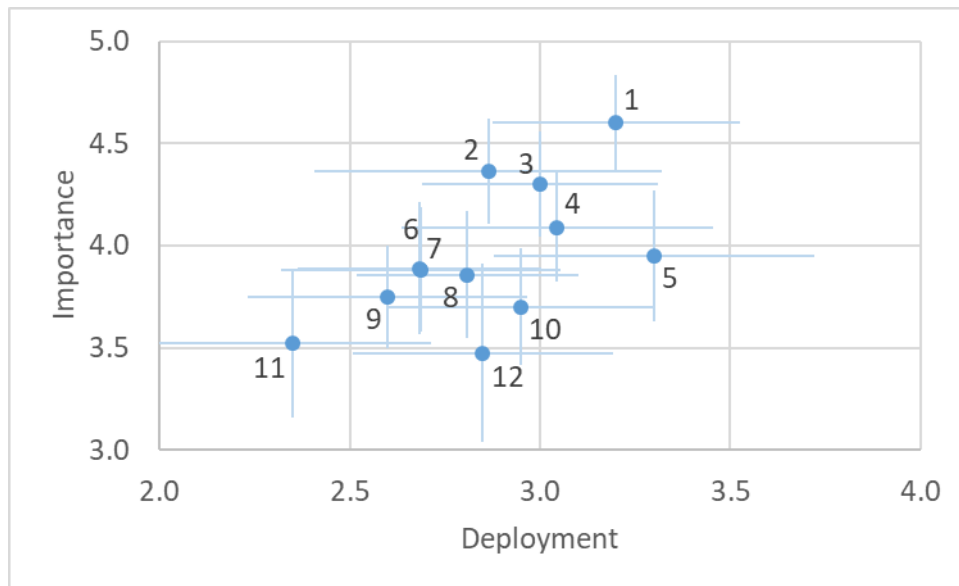


Figure 7 - Detail of Importance vs Deployment Easiness/Difficulty; numbers refer to ranking (Tab. 1); bars refer to 95% confidence interval

The least innovations by importance are *Digital twin* and *Georeferenced data*. The first was not considered very significant for improving the logistics chain although, according to PPs, digital twin could be a powerful simulation tool. Regarding georeferenced data, they have been judged useful for optimization purposes, but similar results could be easily obtained with the adoption of other technologies.

A wider spread was obtained on the innovations' implementation difficulty in the Adriatic Region. However, also larger standard deviations were spotted in this case, leading to greater uncertainty. innovations having the easiest deployment are *Loading/unloading optimization* and *Big data/data management*, since for both well-established procedures and applications are already available on the market. On the other hand, the most difficult to deploy innovation is *Digital twin* since it requires a considerable effort in terms of system integration, collection of historical data and processing algorithms.

2.1.2. Individual partner results

The following charts show the ranking of innovations deemed by each partner as the most interesting.

In the first chart the length of the horizontal bar represents the easiness of implementation for each innovation: the longer the bar, the easiest the implementation.

In the second chart, the lower part (in *blue*) of the histogram represents the importance of each innovation, while the upper part (in *orange*) shows the feasibility (easiness) of its implementation.

In our methodology, the importance and easiness ranking criteria are not meant to be summed to obtain a total value: therefore, this is simply a graphical representation to help us understand how those values “move” for each innovation and for each surveyed partner.

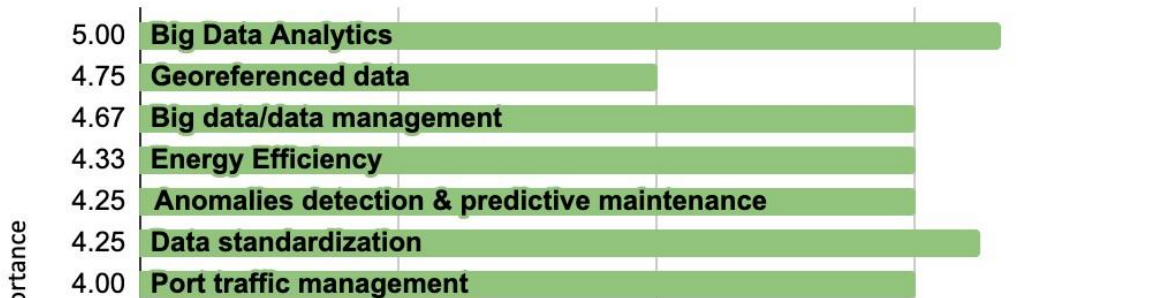
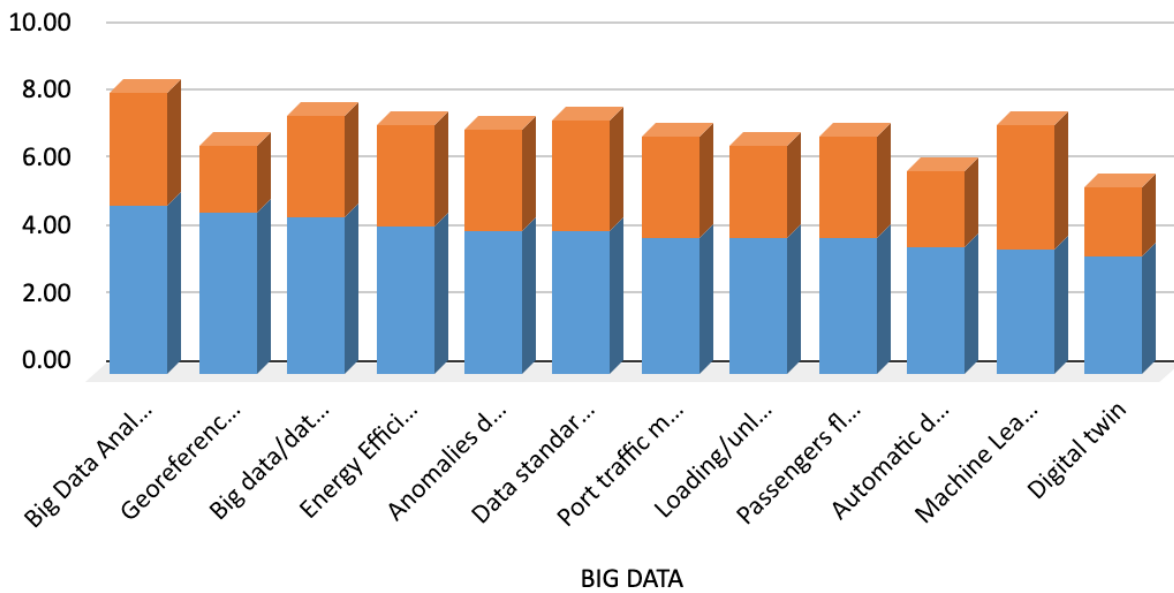
PPs interested in adopting an innovation were expected to be already somehow knowledgeable about the innovation itself and about what its implementation entails.

Please also note the these judgements, are as perceived by partner participants – and they do not refer to an objective evaluation/scoring.

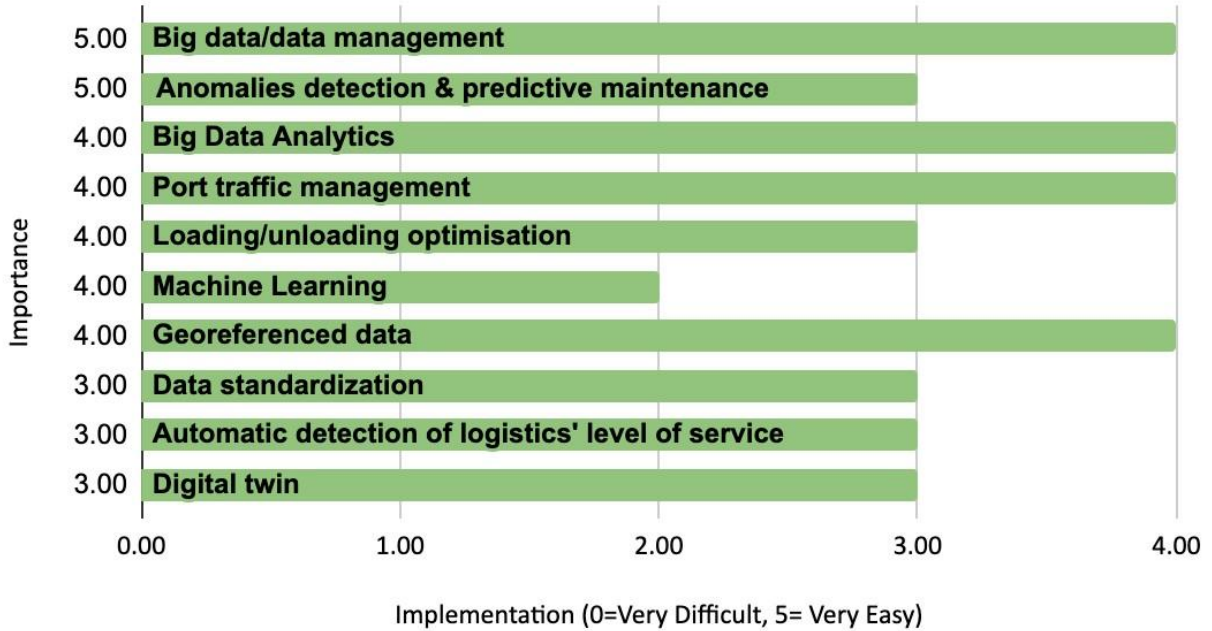
From those innovations considered important by the partner, we can see that a few were also seen as the most transferable (as perceived to be somewhat easy to implement). However, from the chart it is clear that many innovations we ranked as having mid to high importance and they could be be feasible for implementation.

Please not that results for LP and PP3 are not presented here, since those two partners will not be involved in pilot actions and therefore do not need to identify specific innovations within the context of the Diglogs project.

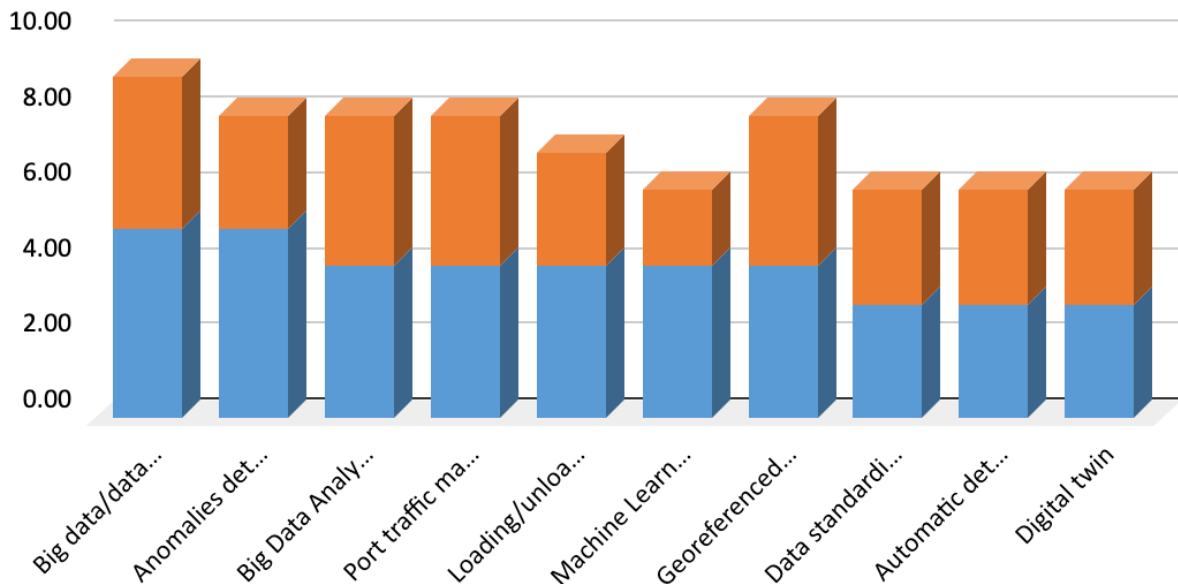
PP1 – combination of perceived Importance [blue] and implementation easiness/difficulty [orange]



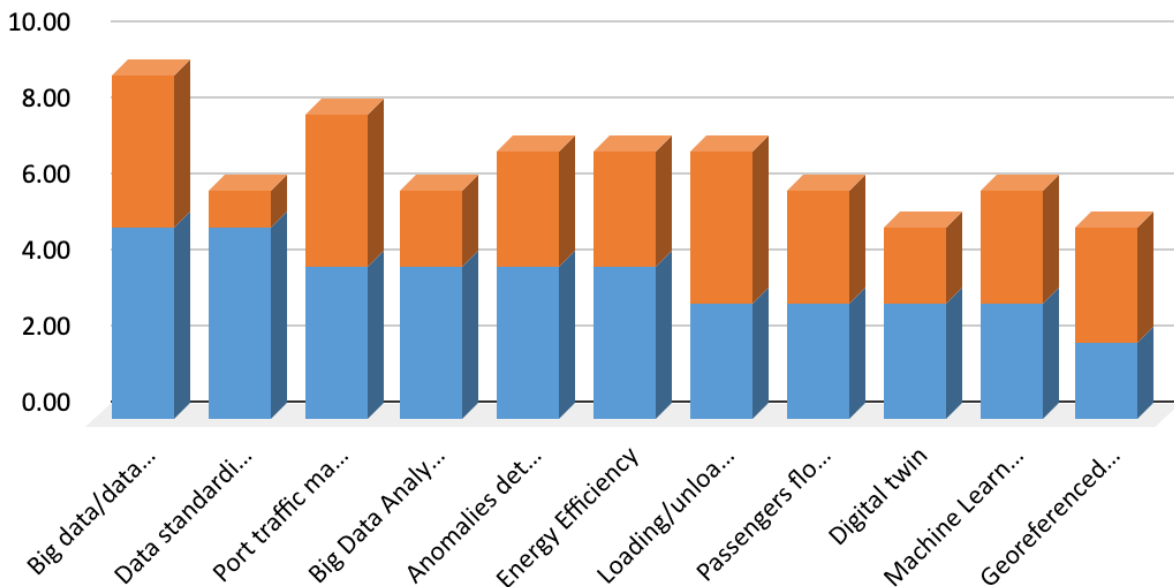
PP2 - Innovations in Big Data/Analysis



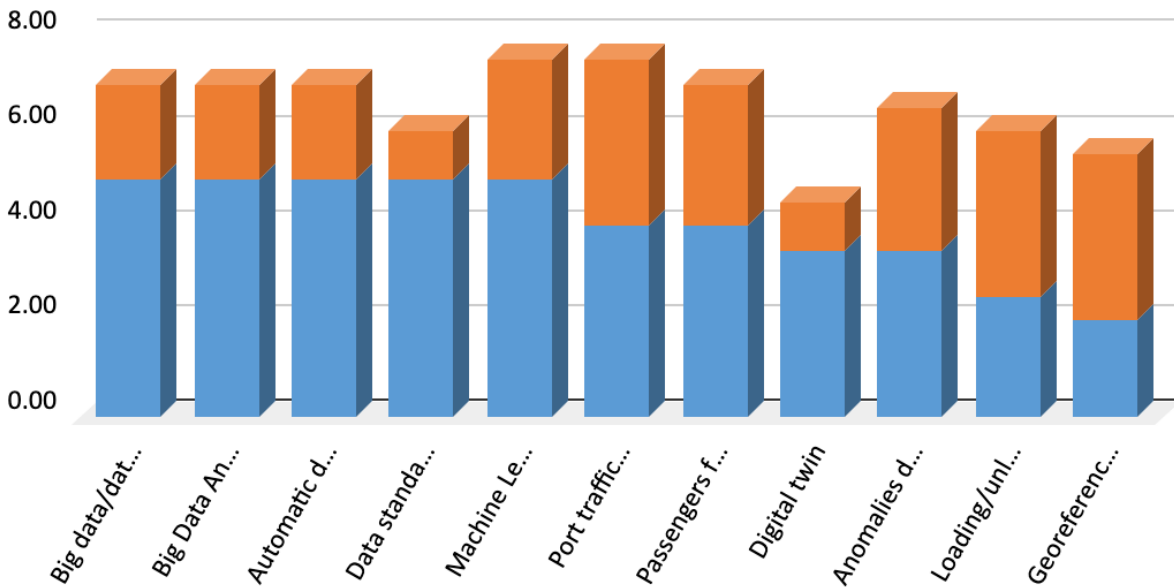
PP2 – combination of perceived Importance [blue] and implementation easiness/difficulty [orange]



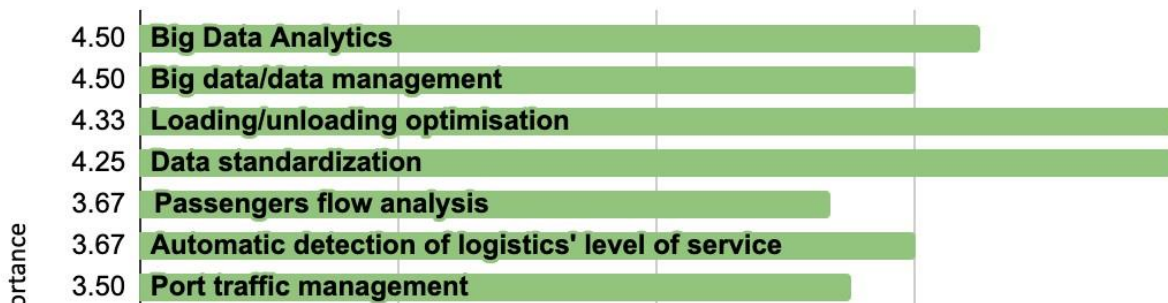
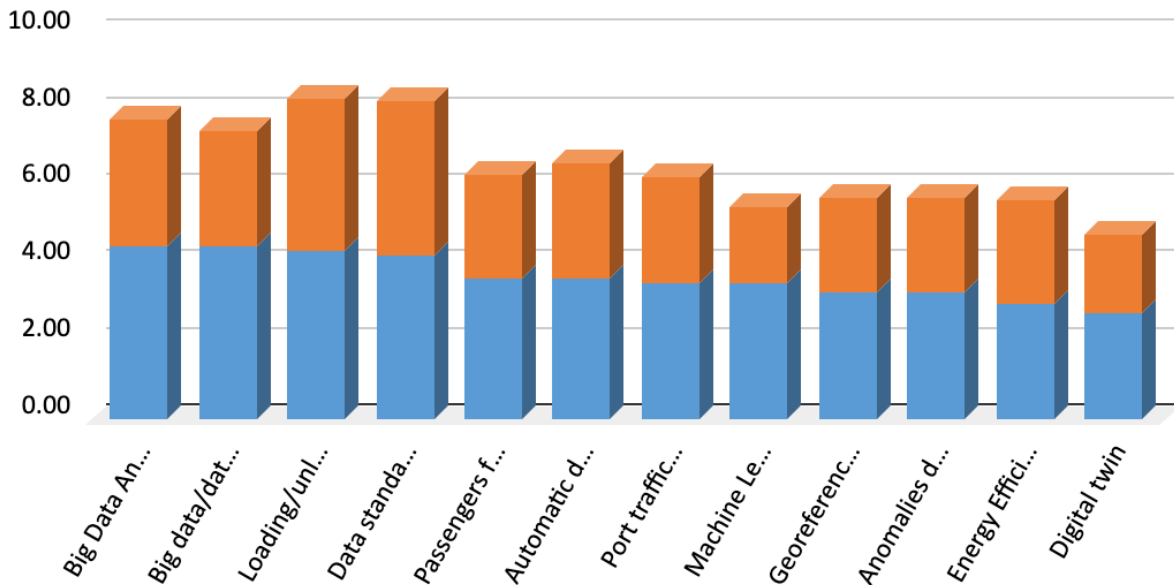
PP4 – combination of perceived Importance [blue] and implementation easiness/difficulty [orange]



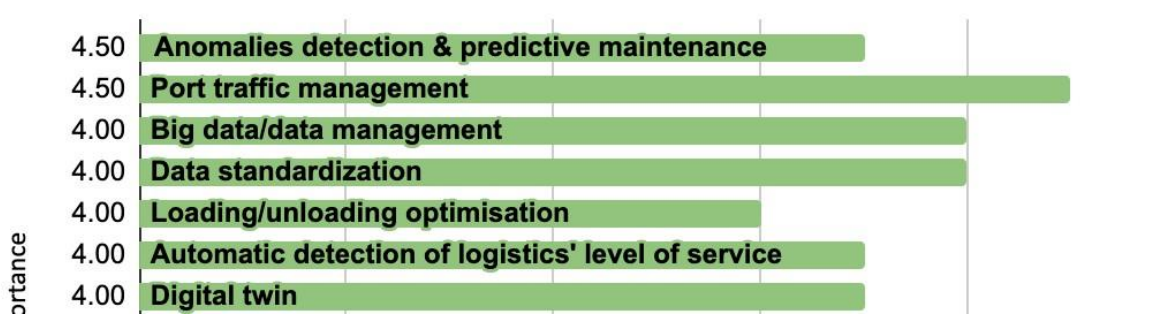
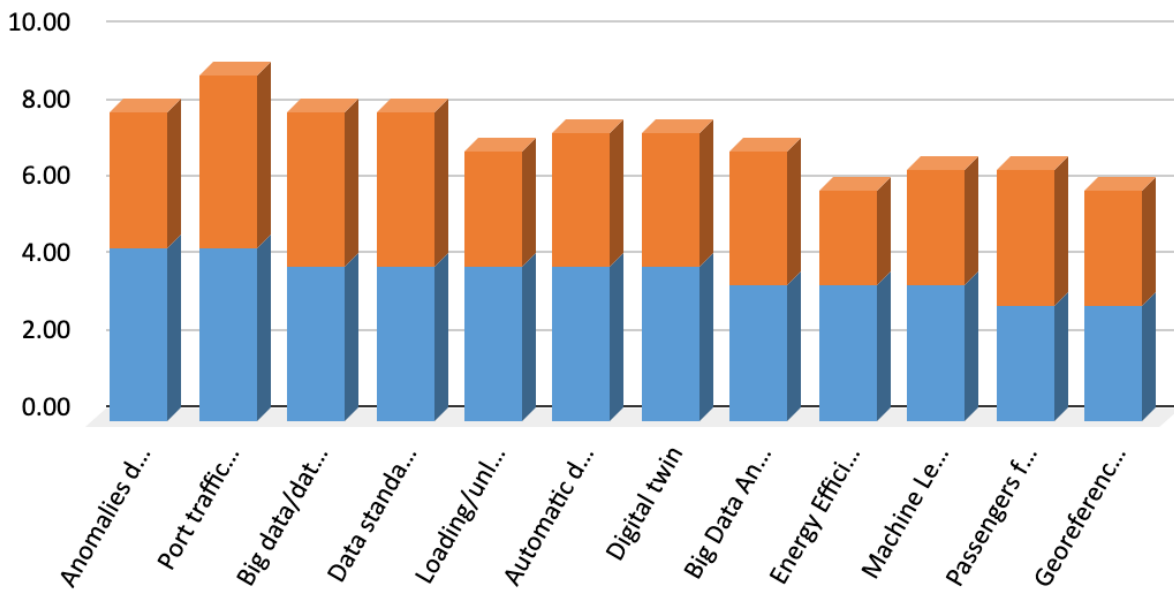
PP5 – combination of perceived Importance [blue] and implementation easiness/difficulty [orange]



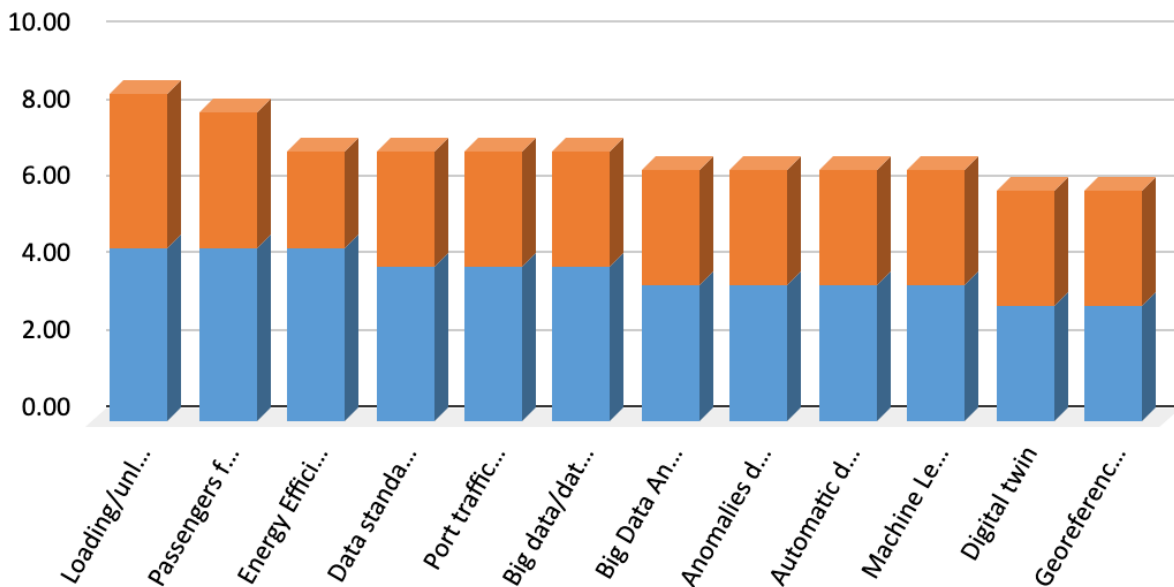
PP6 – combination of perceived Importance [blue] and implementation easiness/difficulty [orange]



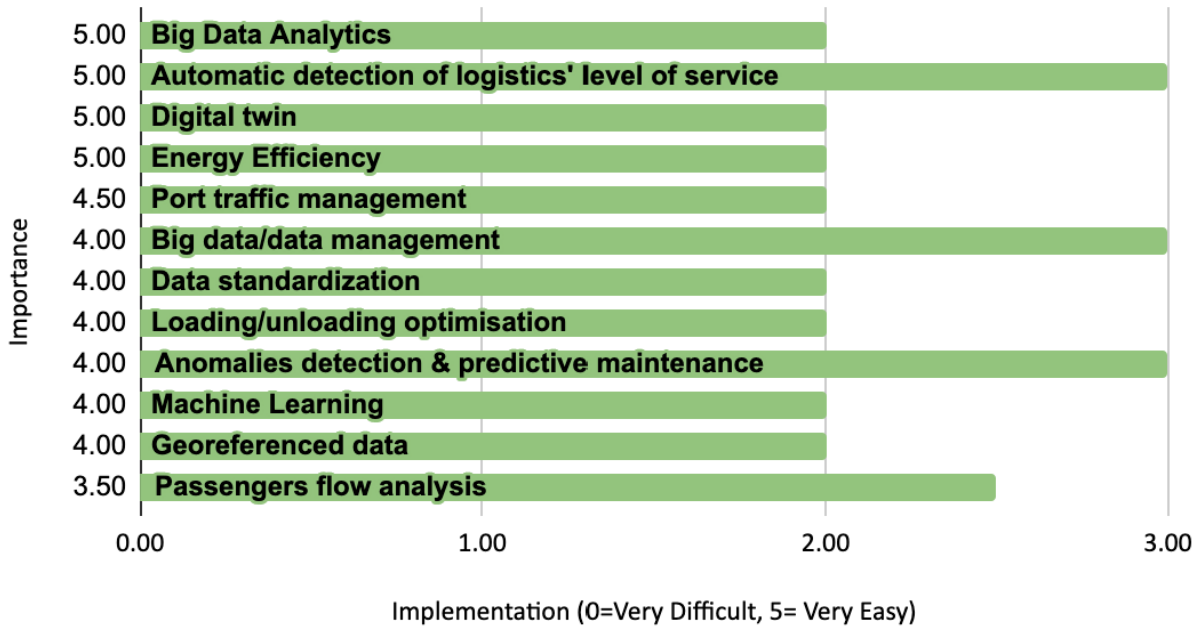
PP7 – combination of perceived Importance [blue] and implementation easiness/difficulty [orange]



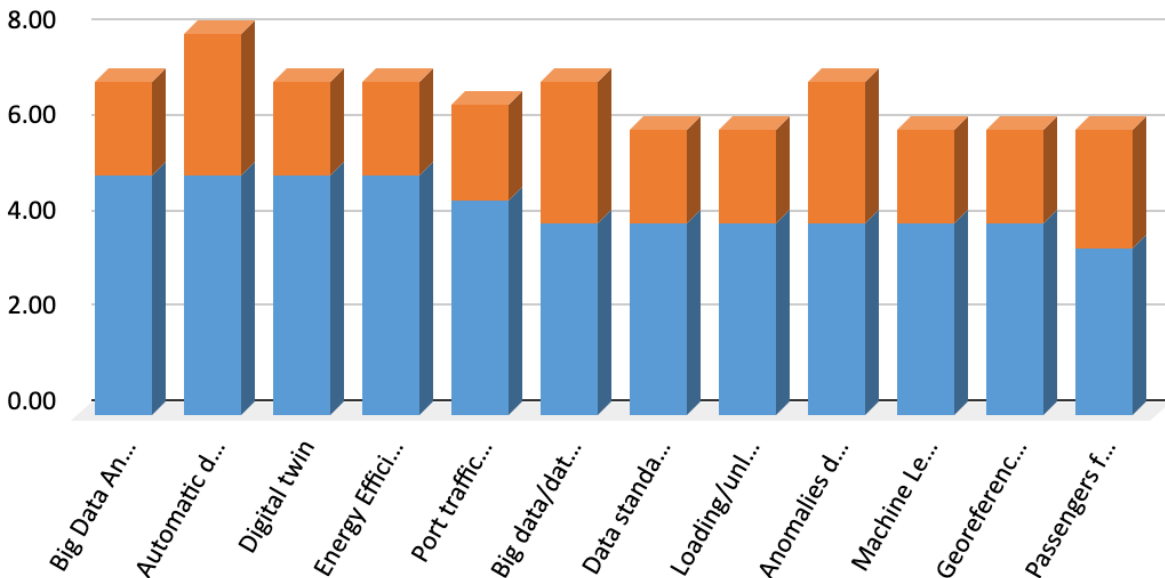
PP8 – combination of perceived Importance [blue] and implementation easiness/difficulty [orange]



PP9 - Innovations in Big Data/Analysis



PP9 – combination of perceived Importance [blue] and implementation easiness/difficulty [orange]



Conclusions

Some conclusions can be drawn from overall results.

From the previously presented graphical representations of data, it can be noticed that all the innovations are located in a restricted area in the upper part of the scatter diagrams. In general, it can be concluded that no clear correlation exists between importance and deployment's easiness/difficulty. It also appears that all the partners assigned a medium/high importance to all the innovations defined in D3.1.2, confirming that the selection process carried out by WPL and PP4 in the previous deliverable was effective.

However, the quite high value of standard deviations related to innovations' importance reduces the significance of the ranking. Nevertheless, a group of more interesting innovation can be defined including *Big data/data management*, *Data standardization* and *Big Data Analytics*. Thus, the PPs assign globally a clear preference for data storing (database design/data management) and data standardization issues that are a prerequisite for the efficient application of data analytics. This preference suggested that, in the Italy-Croatian program area, the application of big-data technologies is just at its starting phase, but there is a strong interest in the development of the infrastructures required for their exploitation.

The innovations ranked lowest by importance were *Digital twin* and *Georeferenced data*. The first was probably not considered very significant for improving the logistics chain although, according to some PPs, digital twin could be a powerful simulation tool. Regarding georeferenced data, partners have judged it useful for optimization purposes, however similar results could also be reached by adoption other technologies.

A wider spectrum was obtained on the innovations' implementation difficulty in the Italy-Croatia Adriatic program area. However, also larger standard deviations were spotted in this case, leading to greater uncertainty. Innovations appearing as the easiest to deploy were *Loading/unloading optimization* and *Big data/data management*, probably because well-established procedures and applications are already available on the market and directly transferable to PPs' contexts. The most difficult to deploy innovation is *Digital twin* since it requires a considerable effort in terms of system integration, collection of historical data and processing algorithms.

Remarkably, the whole “big data issue” does represent a significant focus in the area. In particular various aspects turned out to be important, ranging from data management to analytics and data standardization.