This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

Verification through Accelerated testing Leading to Improved wave energy Designs

Deliverable 2.5

Specification for data management for the VALID Hybrid Test Platform

Version 1.0

2022-11-30

Lead participant: TECNALIA

Dissemination level: Public

DOCUMENT STATUS

Authors

Name	Organisation
Vincenzo Nava	TECNALIA
Pablo Ruiz Minguela	TECNALIA
Günter Lang	AVL
Daniel Krems	AVL
Matthias Seidl	AVL

Approval

Name	Organisation	Date
Natalie Williams Portal	RISE	2022-11-30
Fabio Santandrea	RISE	2022-11-30

Document History

Version	Description	Reviewer	Date
0.1	Outline	Günter Lang	2022-06-15
0.2	Methodology, Results and Annexes	Pablo Ruiz-Minguela	2022-09-27
0.3	Internal revision at TECNALIA- drafting guidelines	Günter Lang	2022-10-11
0.4	Revision and extensions by AVL	Pablo Ruiz-Minguela	2022-11-15
0.6	Full draft for internal peer review	Fabio Santandrea	2022-11-28
0.7	Implmentation of changes after peer review	Vincenzo Nava	2022-11-29
1.0	Final version for EC review	Project Officer	2022-11-30

Dissemination level

Short	Type	
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

EU PROJECT NO: 101006927

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

Executive Summary

Deliverable D2.5 of the VALID project analyses the specification for data management of the VALID Hybrid Test Platform during the experimental campaigns for the three User Cases of the project. As part of the work in WP2 – Task 2.4, the deliverable collects useful information for the development and adaptation of AVL's Integrated and Open Development Platform (IODP) to serve as an intermediate hybrid platform for the experimental testing of the User Cases.

In this report, first the concept of Product Data Management (PDM) is introduced as the data handling during product development and then described in detail (Section 2). The experience in the automotive sector regarding the overall data handling requirements in complex product development and hybrid testing settings are therefore described as example and guidance for the User Cases.

The methodology adopted for the identification of the data needs is introduced in Section 3. A survey internal to the consortium members has been carried out. The respondents to the questionnaires were grouped into two classes: the wave energy developers and test rig managers involved directly into the three User Cases and other users.

The survey covered a wide spectrum of areas inherent to the requirements in terms of data intelligence (the relevance of automated data workflows in their tasks, the robustness of the data storage approaches used in their institutions and the occurrence of failures in data treatment) as well as in terms of data management cycle (pre-process and operations with data, data quality, and storage and data amount).

The answers were collected and analysed in detail in Section 4 and therefore they were used to draw recommendations (Section 5) to be adopted during the hybrid testing phase.

The importance of a robust, trackable, and advanced data flow system has been clearly identified via the procedure adopted. Similarly, the infrastructure for data management should be easily accessible, guaranteeing a multiuser access and being able to cope with a high amount of data collected during the experimental campaigns.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

Project partner names

- RISE Research Institutes of Sweden AB
- Fundacion Tecnalia Research and Innovation
- Corpower Ocean AB
- RINA Consulting S.p.A.
- Biscay Marine Energy Platform SA
- IDOM Consulting, Engineering, Architecture, S.A.U.
- Aalborg University
- AVL List GMBH
- Wavepiston AS
- Delft University of Technology
- Aquatera Sustainability Ireland LTD
- Julia F. Chozas, Consulting Engineer
- Yavin Four Consultants, Unipessoal LDA
- Aquatera Atlantico
- Technical University of Denmark

Table of Contents

Tables

Figures

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

1 Introduction

1.1 Background

A recurrent problem in the wave energy sector is that reliability testing at low TRLs (which corresponds to early development stages) is too costly, and there are many uncertainties and design-decisions that need to be considered. In contrast, when the development process reaches high TRLs, the design is too rigid to make changes, and associated costs can be prohibitive.

The VALID project aims to develop, implement and enforce a new test procedure based on accelerated hybrid testing techniques. Accelerated hybrid testing allows to integrate knowledge from a real environment (ocean, uncontrolled testing), a simplified lab environment (physical test rigs, controlled testing) and a virtually enhanced environment (numerical models, controlled testing). Once implemented, it will enable the industry to scale-up simulated lab conditions and test a virtual model of the existing structure, and hence reducing uncertainties, increasing confidence in results, empowering informed decision-making, and thus, largely assisting in the design and development process of WECs, especially at low TRLs.

WP2 is tasked with the development and adaptation of AVL's Integrated and Open Development Platform (IODP) to serve as an intermediate hybrid platform for the wave energy sector. The foundation of this work is the interface between all the modules in the WEC device development process, independent of tools or suppliers. For this purpose, the requirements for the VALID Hybrid Test Platform must be established to align the wave energy development process, to create and integrate the different numerical models and testbeds, as well as to manage data from different sources.

In particular, Deliverable D2.5 analyses the specifications for data management within the VALID Hybrid Test Platform.

1.2 Structure of the Report

This deliverable is organised in five main sections:

Section 1: Introduction, where the scope and structure of the deliverable are presented.

Section 2: Overall Data Handling Requirements, which introduces the data management cycle and presents the experience in the automotive sector regarding the overall data handling requirements in complex product development and hybrid testing settings.

Section 3: Aim and Methodology, describing the survey approach adopted to assess the current status of the data management within the consortium and to identify needs for a more robust and reliable data management cycle in hybrid testing in each of the three User Cases.

Section 4: Results, which analyses the responses in relevant areas of the data intelligence and data maintenance cycle.

Section 5: Data Management Recommendation for the Hybrid Testing Methodology, which summarises the key recommendations for the overall data handling, data intelligence and data management cycle.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

2 Overall Data Handling Requirements

The accelerated hybrid testing campaigns to be carried out within the VALID project present significant challenges in terms of data handling, as suggested by the historical trends observed in the practice of more mature industrial sectors such as the automotive (see Section 2.1 for further details) Task T2.4 in the VALID project was specifically designed to analyse the role of Product Data Management (PDM) in the industrial development of wave energy systems. More generally, PDM includes all organization-based tasks for the identification, supply, storage, and archival of data generated or collected during the product development process [1].

The scope of PDM and its relation to Product Lifecycle Management (PLM) is shown in Figure 7.

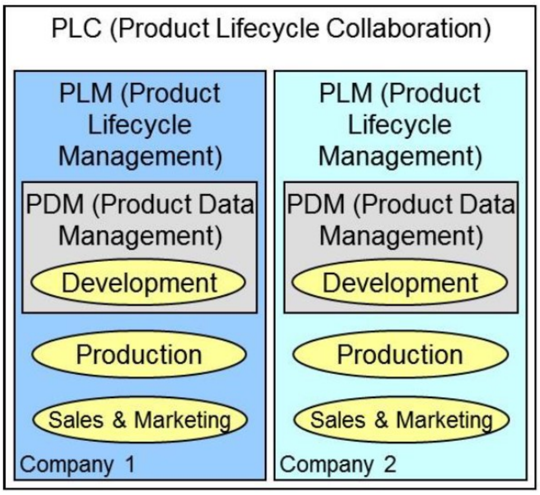


Figure 1: Nomenclature and application areas of PLC, PLM, and PDM [1]

The main benefits of a PDM system, according to *NPD Solutions*, a product development consulting company, are [2]:

- **Time-to-Market:** Data is instantly available to all with access. There is no waiting for paper documents to be distributed nor time wasted while documents sit in an in-basket waiting for review. Time spent searching for component and product data is reduced. Collaboration features also speed and improve the process.
- **Improved Productivity:** Studies have shown that engineers spend 25% to 35% of their time searching for, retrieving, handling, filing, and storing documents and information. This time can be reduced with a PDM system and its single repository, its classification and information structuring capabilities. The classification and search capabilities aid design retrieval, provide the opportunity to avoid "reinventing the wheel", and, as a result, reduce the related development effort.
- **Improved Control:** Because PDM enables better management of different product configurations and it assures that everyone is working on the most recent data, thus limiting the risk of working with obsolete data. Access control features assure that only authorized parties can access or change proprietary information. Control over engineering changes is improved with less manual effort.

Another source [5] lists following key advantages of introducing a PDM system:

1. Shorter time to get the first insights: Enterprise-level data management and analysis for characterization and validation

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

2. Simplified generation of reports: Faster result correlation, creation of data sheets and protocol generation
3. Management of all test data: Collect and manage data independently of the test provider over the entire life cycle

2.1 A Look at the Automotive Industry

In the automotive Product Development Process (PDP), several standards for data management are established. These standards shall guarantee consistent tracking of data for e.g., safety critical systems or fulfillment of regulatory requirements, such as those regarding greenhouse gas emissions. Further standards exist to support operations during the service life and after-market phases of vehicle lifecycle. The main driver for vehicle developers (e.g., OEM, TIER1 or TIER2 supplier etc.) to adopt data management standards is the possibility to make the development process more efficient.

As of today, a large part of the automotive industry maintains a Bill of Materials (BoM) or product-structure-driven Vehicle Development Process (VDP). For many decades, PLM approaches have been based on a static PDM approach. The top of Figure 2 schematically shows different artefacts involved in PDM. However, current industry requirements need more flexible approaches based on the integration of multiple tools that must be able to exchange information objects efficiently and seamlessly.

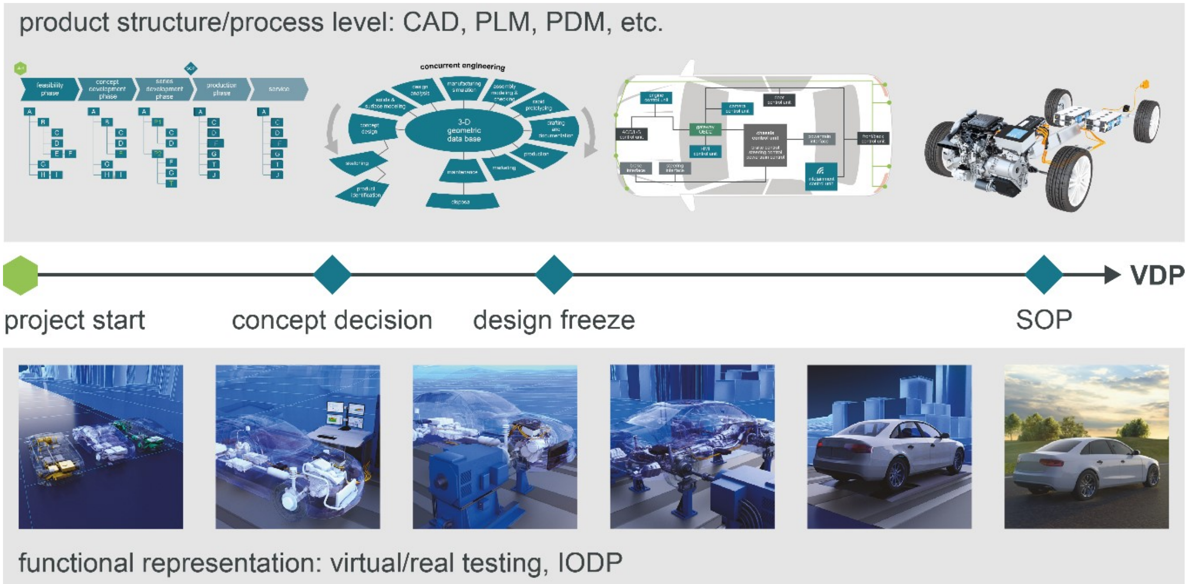


Figure 2: The vehicle development process is based both on a product structure and a functional product representation [3]

By tracking characteristic values over the entire development process, deviations can be identified earlier, and changes of the product design are cheaper and easier to apply in an early stage (so-called frontloading), Figure 3.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

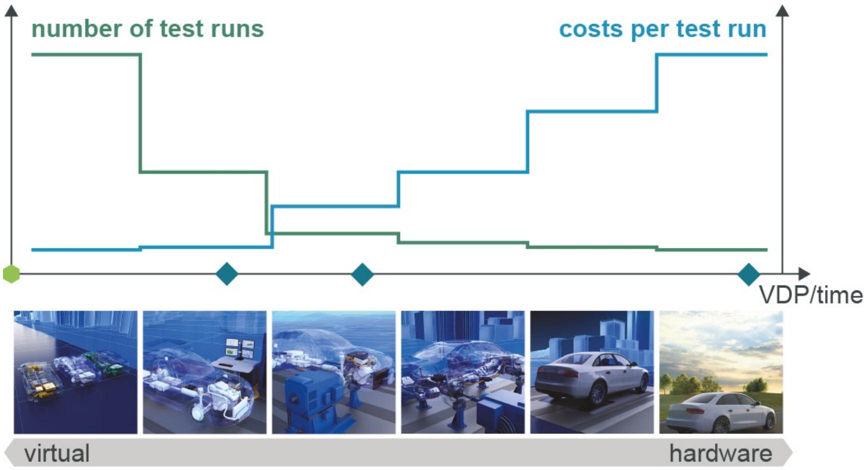


Figure 3: Evolution of the number of test runs and cost per test run along the vehicle development process [3]

Such a process is only possible when system parameters are up-to-date and traceable at any stage of the development process. Otherwise, the result of tests is neither representative, nor reproducible. Therefore, a PDM system seems mandatory. A possible approach (to PDM?), called Data Backbone, was proposed by AVL [3].

Over the course of the VDP, a massive amount of data is generated, e.g., Computer-Aided Design (CAD) models, simulation models, parameters, and test results. These different kinds of data are typically stored in highly optimized, but isolated and decentralized data management systems.

The Data Backbone enables consistent management of data across the entire VDP by combining data from multiple sources. Utilizing data virtualization technologies, data can be set into relation with each other while physically remaining in its original location: Instead of duplicating the data, a repository of links captures relationships amongst the data. The interconnection of data is crucial in data management as it enables *traceability of data.* On the other hand, users are required to comply with a strict format for their data to enable *big data or massively structured data* technologies.

For example, by interlinking different data elements, it is possible to capture which calibration data set was used to generate a specific test result on an e-drive test bed. Another example is shown in Figure 4: specific versions of calibration sets and embedded software in a car's control units can be linked together into so-called *flashsets*. Flashsets help calibration engineers determine the proper combination of software and calibration before starting their work.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

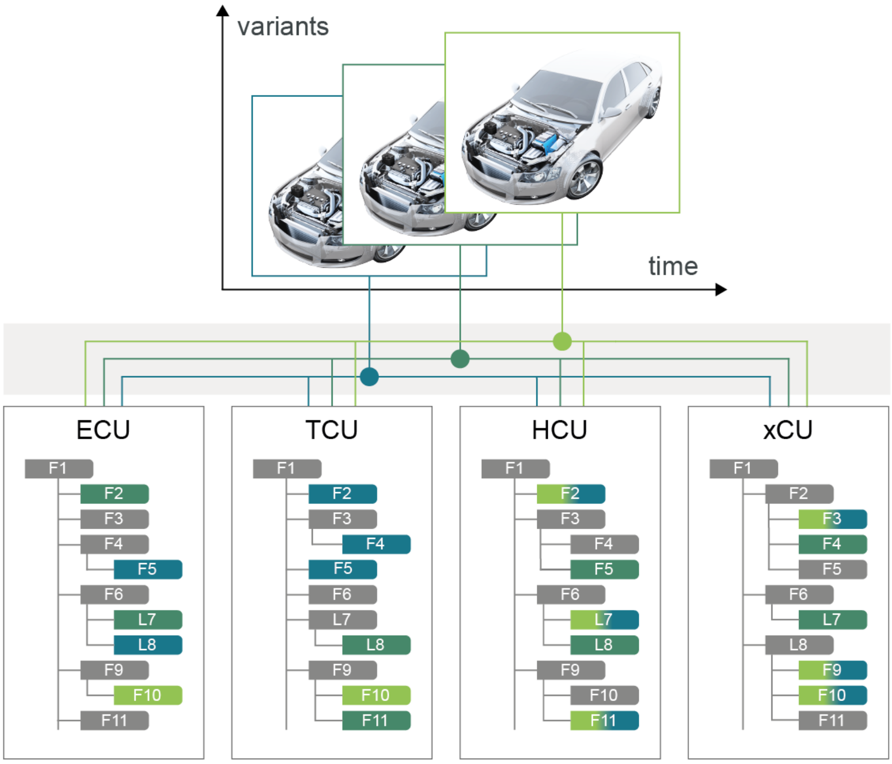


Figure 4: The flashset manager as an example for the data backbone [3]

2.2 Development and co-operation standards

A White Paper, published by the Interest Group "Digital Plant" of ProSTEP-iViP [4], gives an overview of how to use data standards in order to 'accelerating simultaneous development of products and manufacturing resources.

In short, the goals of the Interest Group are:

- Preparing solutions for the integrated development of products and manufacturing resources
- Enabling and accelerating the collaboration between vehicle manufacturers and manufacturing equipment suppliers
- Providing access to actual product, resource and process data in a heterogeneous system environment (i.e., PDM, CAx, process and manufacturing resources planning systems)
- Enabling the coupling of systems
- Driving vendors to develop interfaces
- Establishing flexible and comprehensive data communication based on international standards

The interlinkage of in-house and supplier process communication is shown in Figure 5. ISO 10303 is an ISO standard for the computer-interpretable representation and exchange of product manufacturing information. Thus, it is not directly linked to hybrid testing, but the figure is cited, as it clearly shows the multiple dependencies on product data, and how a standard was implemented in order to enhance the efficiency of product development and manufacturing by supporting data sharing between different parties.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

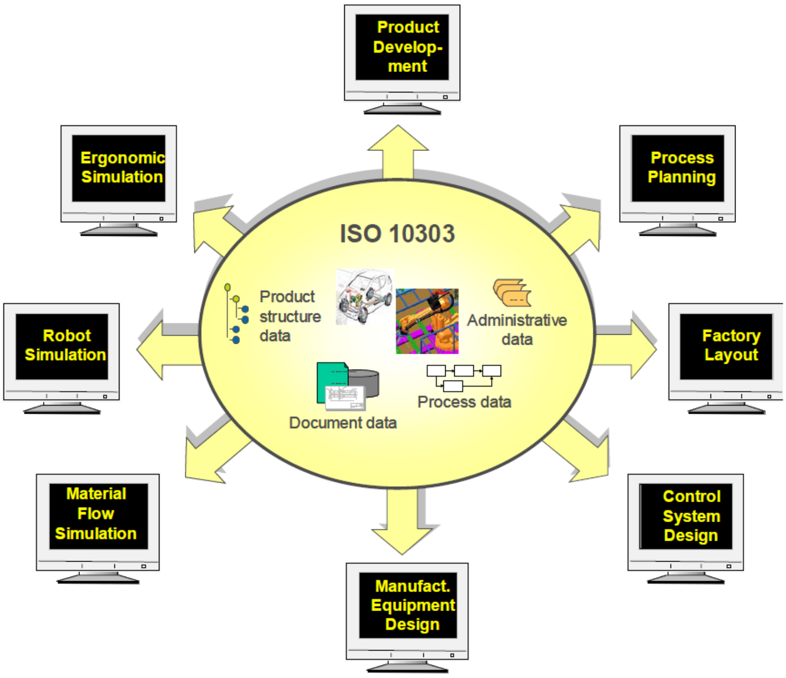


Figure 5: OEM - Supplier and in-house downstream process communication [4]

To reach the goals stated about, the following approach was suggested:

1. Definition of application and system scenarios for the exchange of product and resource data (resource data: product data of manufacturing equipment)
2. Definition of user cases
3. Definition of recommendations for the use of ISO 10303 and related standards for the exchange of product and resource data
4. Definition of implementation guidelines

2.3 Evaluation of most important data management features in the automotive industry

A survey with both project managers and development engineers in the automotive industry, conducted by AVL, gave the following results:

Most important features (view of project managers)

- Product maturity dashboard & validation process, featuring frequent updates
- Mapping output of simulation results directly to product KPIs
- Traceability of simulation results
- Standardized process for trust in simulation/test results

Most important features (view of development engineers)

- Context information (model- and hardware-version, parameters, in- and outputs) stored with results
- Harmonized post-processing approach
- Continuous and integrated development environment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

The outcomes of the survey can be translated into a concise list of prerequisites for efficient data management in product development:

- capability to constantly monitor product maturity (in terms of meeting the requirements)
- support for storage and traceability of component- and model-data, as well as their links to test and simulation results
- capability of extracting Key Performance Indicators (KPIs) from the data.

There are several PDM tools on the market that fulfil these requirements, such as.:

- Siemens Teamcenter
- Dassault 3DEXPERIENCE
- Autodesk PLM360
- PTC Windchill
- AVL CRETA (focusing on calibration data management)
- AVL CRETA for Simulation (model parameter management)

Cost-free solutions for PDM are also available.

While most PDM systems can handle data of hardware (e.g., part numbers), hybrid testing also requires the management of simulation data (e.g., model version, software version, and model parameters). The relationships among multiple data sources of heterogenous nature is illustrated in Figure 6, which represents a typical scenario for vehicle development in the automotive industry. The capability to handle data about physical and virtual artifacts is a strong driver in the selection of a PDM system to support product development with hybrid testing Figure 6.

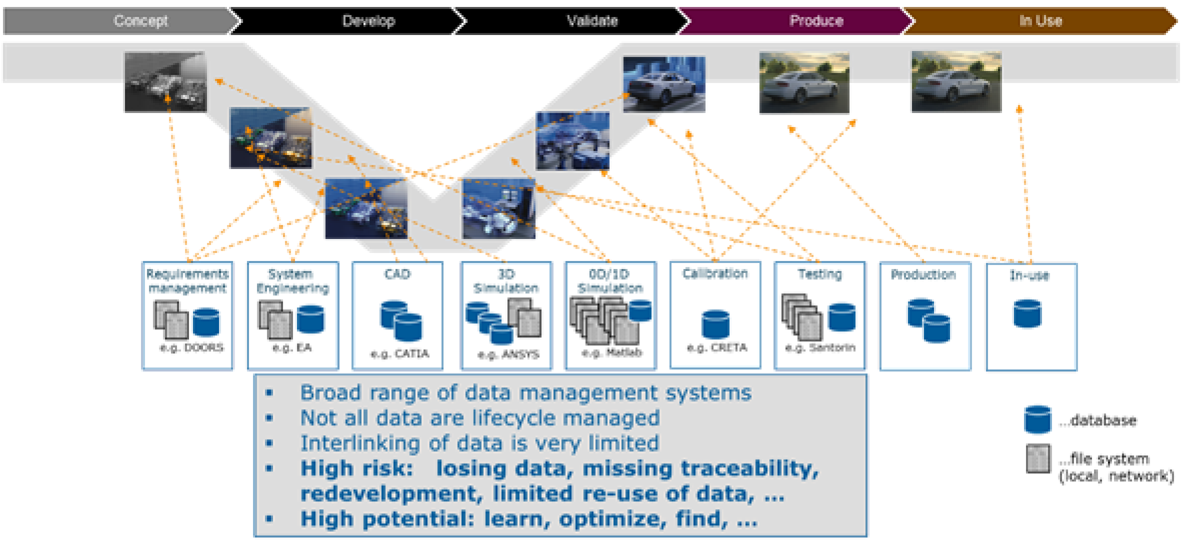


Figure 6: Interlinkage of data sources throughout the vehicle development process (source: AVL)

One of the objectives of T2.4 of the project VALID was to investigate if the experience matured in the automotive industry about data management could be transferred in the wave energy sector, thus leading eventually to the identification of suitable technical requirements for the users of the Hybrid Test Platform in VALID. The methodology implemented to pursue it are described in Section 3.

3 Aim and Methodology

Task T2.4 of the VALID project is intended to understand the requirements for data management in in the context of the VALID Hybrid Test Platform illustrated in the different User Cases.

An internal survey approach was conducted with a twofold aim:

1. To assess the current status of data management among the members of the VALID consortium.
2. To identify needs and improvements for a more robust and reliable data management to be applied in the hybrid test campaigns which will be carried out in the three User Cases of the project. UC1 is led by Corpower and it aims at testing the dynamic sealing failure; UC2 is led by IDOM and it involves the testing of electric generator failure; UC3 is led by Wavepiston and aims to testing the failure of the hydraulic pump of the PTO.

As an outcome of the survey, the results of the questionnaire are ultimately intended to identify criticalities and bottlenecks for data management in all phases of the testing stage: from planning the different accelerated test scenarios to controlling the test during operation and monitoring the progress of the test.

The survey was created using a Google Form and it was divided into three main sections:

- **Identification**: only for the records, asking for name of the person filling the survey, company and role inside the project
- **Data Intelligence**: this section collected information about the value assigned by each company to procedures increasing the automation process in data handling, the traceability, the storage and up to which level these measures have been implemented in each organization.
- **Data Management**: this section provided more insight into the specific case of application targeted in the VALID project, collecting information about data in terms of handling, sharing, formatting, integrity, safety of the sources and documenting.

The full content of the questionnaire as it looked like in the Google Form, as well as the multiple choices answers are included in the Annex 1. For the ease of the reader, in Table 1 the list of the questions for each section and their numbering is included.

Table 1: List and numbering of participants

#	Question
IDENTIFICATION	
1	Name
2	Institution
3	Type of institution
4	Main technical role of your institution in the VALID project (in case of widespread contribution, select which is most affected by the data management)
DATA INTELLIGENCE REQUIREMENTS	
1	How important is an automatic data management workflow?
2	How important is a robust storage system (file system, database) for your application
3	Which is the most used storage system in use in your company?

4	How would you rate the experience with the storage systems in use in your institution?
5	Did you experience errors in the past because of using outdated data?
6	In case you answered yes to the previous question, what was the problem (e.g. wrong design data for simulation, wrong boundary conditions for simulation or testing, etc...)
7	How important to track KPIs (key performance indicators) during the processes?
8	Do you currently track the main KPIs during the processes?
9	How important to visualise constantly the main KPI (key performance indicators) during the operation(s)?
10	Is the system that is currently in use in your institution able to visualise important KPIs real time?
11	How important is traceability of data throughout all the lifecycle of the project?
12	Which are the traceability systems adopted in your institution?
13	Are your product parameters managed in a central place?
14	How are your product parameters managed in a central place?
15	Are your models managed in a central place?
16	How are your models managed in a central place?
17	Is your measurement data managed in a central place?
18	How is your measurement data managed in a central place?
19	Are your commercial data linked to project data?
20	How are your commercial data linked to project data?
21	Do you extract metadata during postprocessing (e.g. automatic recognition of critical events)?
22	Which meta data do you extract during postprocessing (e.g. automatic recognition of critical events)?
DATA MAINTENANCE CYCLE REQUIREMENTS	
1	How important is to plan adequately the pre-process of data for reducing risks in terms of delays
2	How important is to plan adequately the postprocess of data for reducing risks in terms of delays
3	Which are the most usual pre-processing techniques you apply for your data?
4	How many users can have access to the data not simultaneously?
5	How many users can have access to the data in parallel at the same time?
6	Do you use any tool to manage metadata
7	Which tools do you use to manage metadata?
8	Are data standards available in your domain?
9	Are you following any standard for data formatting?
10	Which data standard are you using for formatting your data?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

11	How do you share data internally to the project?
12	How do you share data externally?
13	How important is to check for the quality of data?
14	Which is the current trade-off between computational time and accuracy of the pre-process of the data?
15	Do you use data from several sources
16	Which are the sources of data you are using?
17	Do you make a real time check of the consistency/robustness of the data
18	Which tools do you use to check consistency and robustness of the data?
19	Do you apply redundancy in data acquisition?
20	How do you create data redundancy?
21	Do you use any tool for versioning the data?
22	Which tools do you use for versioning the data?
23	Is your data organised in a way that the most recent/released data can be accessed easily?
24	Where is the data stored?
25	For each of the following examples of data, select the format(s) you have adopted in your domain
26	Do you save ..?
27	Which is the amount of data (input) you require for a simulation/hybrid experimental sample?
28	Which is the amount of data (output) you produce in a simulation/hybrid experimental sample?
29	How many channels are in your logging system?
30	In case you record time series, how long the time series are?
31	Which is the sampling frequency of your time series?

The questions were agreed between TECNALIA and AVL after receiving feedback from the other partners in the consortium.

The categories of questions were as follows:

- **Linear scale** from 1 to 10, when the respondents were asked to grade the importance of the requirements, on a scale from 1 (low importance) to 10 (highest relevance).
- **Checkboxes**, where the respondents were asked to choose one or more of predefined options in a list or to express an alternative answer in their own words.
- **Short-answer text, where** the respondents could insert a short free text based on their experience.
- **Multiple choice or tick box grids**, where the respondents can select the simultaneous selection of two requirements.

The survey was sent out to representatives of all the organizations that participate to the project; at least one answer was required from partners who lead the UCs, while the answers

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

from all the other partners were optional, primarily considered for statistical purposes. The respondents were recommended to answer exclusively with regard to the specific contribution of their organization to the project. This recommendation applied especially to universities, departments and business units of companies, and research centres with a wider business portfolio that can apply different data management protocols.

The survey was launched on June 15, 2022. It was initially conceived with all the questions requiring a mandatory answer. During the meeting hosted by AVL in Graz, Austria, on June 30th, 2022, it was discussed whether the highly technical content of the questions could limit the possibility to engage a larger sample of (qualified) respondents. For this reason, after the meeting, the requirement of mandatory answers was changed to optional. Furthermore, for most questions it was added the possibility to include personalised feedback ("Other" field) in order to customise the answers to the actual status/needs of the respondent. Given the low answer rate in the survey and the summer break in the middle, it was kept open until September 15, when the minimum number of answers was collected.

The list of the respondents is included in Annex 2.

4 Results

4.1 Analysis of Responses

Given the relatively low number of responses (11), it was decided to split the survey results into two main groups:

- **Group A**: Wave energy technology developers (IDOM, CorPower Ocean and Wavepiston) and respondents with a role in VALID as manager/owner of a physical test rig (BiMEP, TECNALIA)
- **Group B**: All the other respondents (AVL List GmbH, Yavin Four Consultants, Delft University of Technology, Aquatera Atlántico, Aalborg University, RISE).

DTU has not answered to the questionnaire even if they are test rig managers (Group A). This is due to their recent involvement into the project. Their needs and views are considered to be integrated into the contribution of Wavepiston.

Group A answers are mostly useful to identify needs and to characterize the common practice adopted in the institutions leading the UCs. Group B answers can be useful to identify approaches commonly used in their institutions for solving similar situations.

Quantitative results, i.e., to the questions using linear scales, have been postprocessed considering mean and median values, being the mean representative of the average answer and accounting for the overall answers. On the contrary, the median provides an insight on the most frequent answers in the community of the respondents. More qualitative answers have been visualised by means of bars diagrams and a critical narrative insight.

It must be pointed out that, due to the limited number of answers collected, the results have no statistical significance. They should just serve to identify trends and as a basis for recommendations regarding data management requirements for hybrid testing in VALID. Further investigation will be carried out under other project activities and finally reported in D1.5 Methodology for Critical Component/Subsystem Testing.

Analysis of survey results is described in Sections 4.2 and4.3.

4.2 Data Intelligence

4.2.1 Data Management Workflow

The first question in the Data Intelligence section concerns the requirements for data workflow and, in particular, the importance attributed by the respondents to automatic workflows for data management (Figure 7).

It is noteworthy that, while most of the respondents in Group A retained automation an important requirement (average 8.4, median 8), a lower rating was given by the respondents in Group B (average 6.2, median 7).

The difference between median and mean reveal different attitudes towards automation between the two groups of respondents: in Group A, some of the ratings were higher than the average, while the opposite was observed in Group B, where some respondents gave very low ratings to the importance of this requirement. Thus, <u>automated workflows for data management seems to be important for the community of the VALID partners directly involved into the hybrid testing, while the rest of the partners has lower expectations</u>**.**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

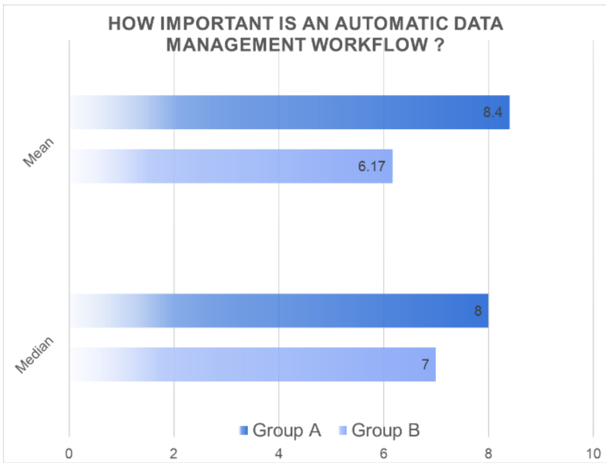


Figure 7: Question 1: How Important is an Automatic Data Management Workflow?

4.2.2 Data Storage

The second question was about the relevance of having a robust data storage system for the user cases (see Figure 8).

The analysis of the answers given to this question shows similar trends and ratings for both Group A and B. Medians and means are greater than 8 for both groups. Therefore, <u>the robustness of the data storage systems is seen as an important requirement by all the respondents to the survey and the VALID user cases</u>.

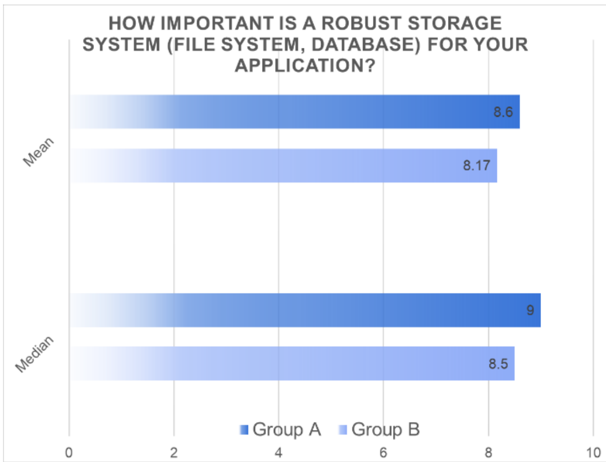


Figure 8: Question 2: How important is a robust storage system (file system, database) for your application?

<u>The respondents reported either file system or database as most used means for data storage</u> (Figure 9), with a higher preference for file system in the case of Group B. In this case, however, it must be kept in mind that some respondents gave more than one answer, when both systems were used at their organization.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

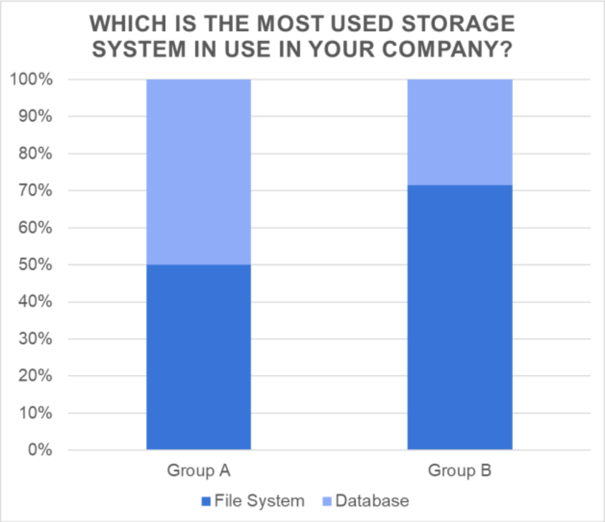


Figure 9: Question 3: Which is the most used storage system in use in your company?

Group A retained, on average that the current data storage system is just sufficiently robust (mean rating 6.4), an opinion that seems to be shared by most respondents in the group (median rating 7). while in Group B the average is higher than 8 (see Figure 10). This is confirmed. Therefore, the robustness of the currently used data storage systems is seen as just sufficient by the users of Group A and fully adequate by users of group B of the VALID hybrid platforms.

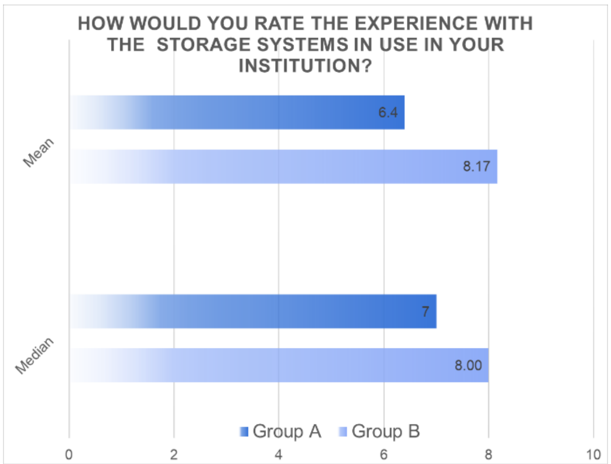


Figure 10: Question 4: How would you rate the experience with the storage systems in use in your institution?

4.2.3 Outdated Data

The majority of the respondents reported to have run into errors due to the usage of outdated data (up to 80% in Group A, as shown in Figure 11). The possible reasons for that were addressed in Question 6, which resulted in the following list:

- Corrupted data, related to data system robustness
- Inadequate sample rates
- Data overflows
- Wrong inputs
- Outdated interfaces and data formatting.

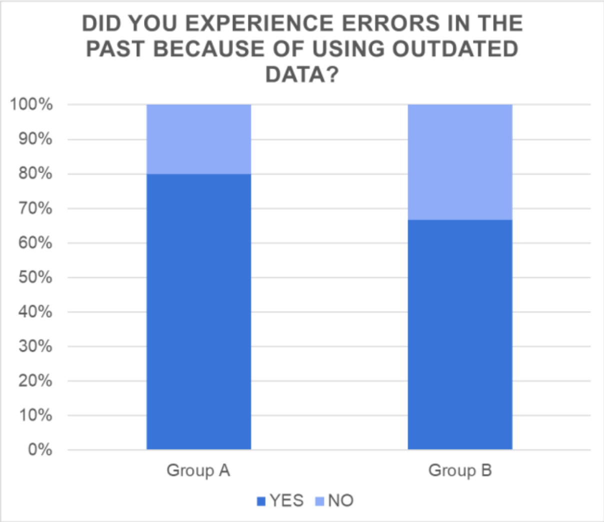


Figure 11: Question 5: Did you experience errors in the past because of using outdated data?

4.2.4 Key Performance Indicators

Respondents were asked to rate the importance of tracking Key Performance Indicators (KPIs), as power production, measured voltage, capacity factor etc, during the realisation of the processes (hybrid experimental campaign) (see Figure 12).

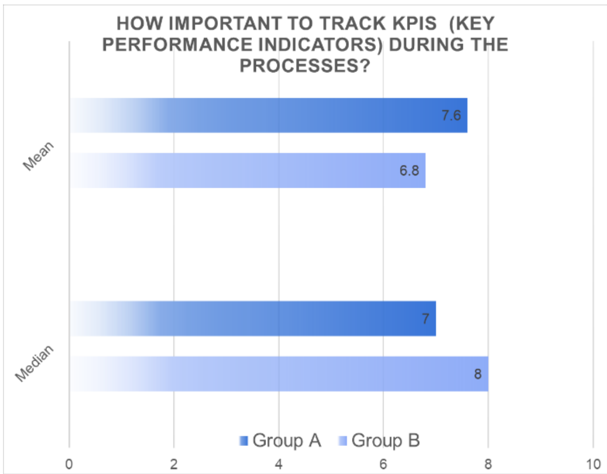


Figure 12: Question 7: How important is to track KPIs (Key Performance Indicators) during the processes?

The average rating given by Group A is greater than that assigned by Group B, thus suggesting a higher importance attributed to the possibility of tracking KPI. However, this difference does not reflect the view of the majority of respondents in Group B (the median of the ratings if greater for Group B than for Group A), and it is essentially due to the very low ratings given by some respondents in this group.

All the respondents in Group A reported to normally monitor KPIs during the processes (Figure 13). Continuous visualization of KPIs was considered important by all the respondents (Figure 14), as suggested by the median of 8 for both groups. Current practice in monitoring KPIs differs between the two groups: while all respondents in Group A keep track of KPIs during processes, only 16% of respondents in Group B do it (Figure 15).

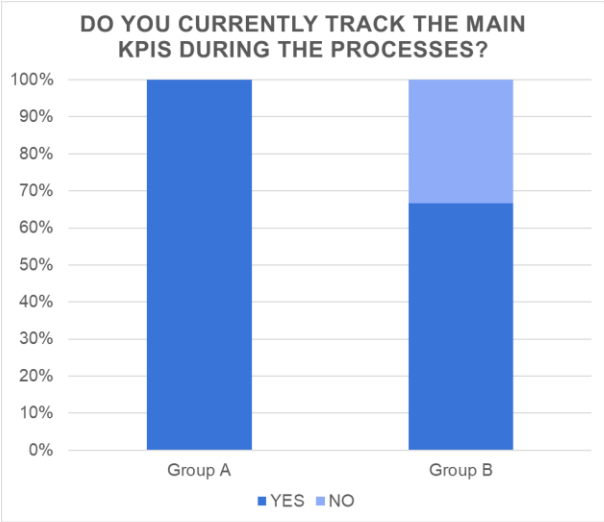


Figure 13: Question 8: Do you currently track the main KPIs during the processes?

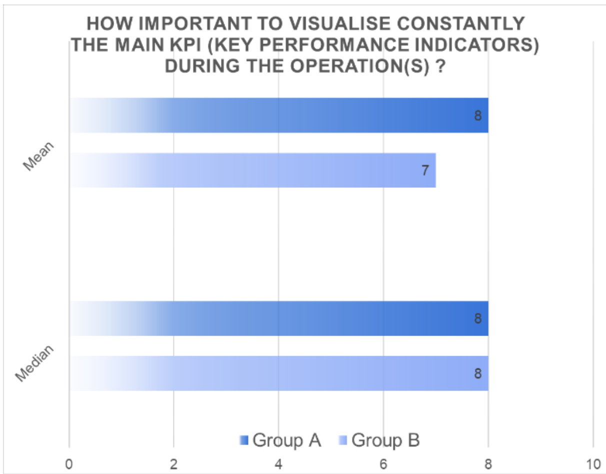


Figure 14: Question 9: How important to visualise continuously the main KPIs (key performance indicators) during the operation(s)?

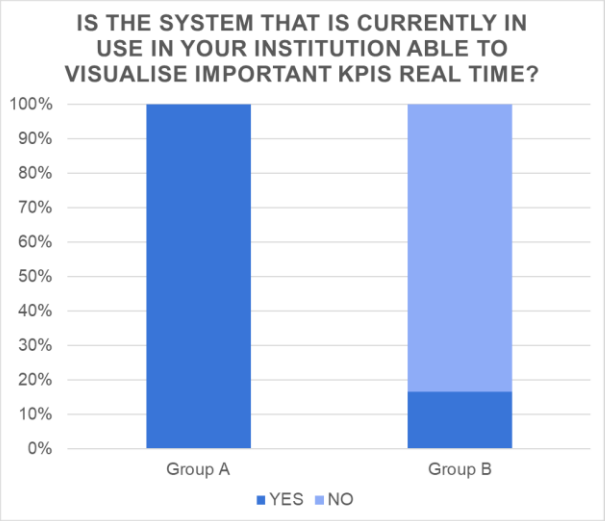


Figure 15: Question 10: Is the system that is currently in use in your institution able to visualise important KPIs real time?

4.2.5 Traceability of Data

Traceability of the data throughout the whole lifecycle of the projects is seen more important for Group B rather than for Group A (Figure 16).

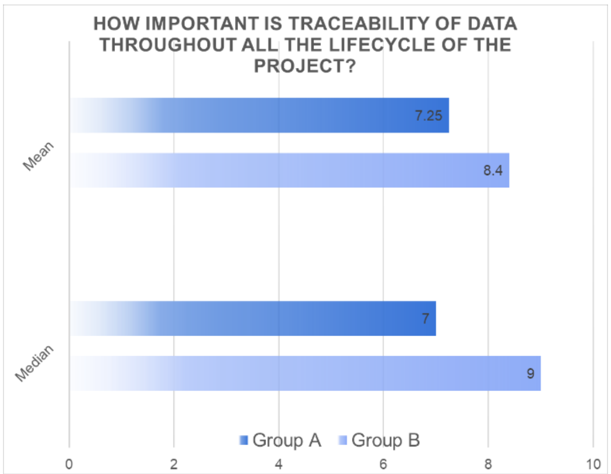


Figure 16: Question 11: How important is traceability of data throughout the whole lifecycle of the project?

In Question 12, the respondents were asked to describe how data could be tracked during the processes in the systems currently used at their organizations. The answers were very dispersed. Group A mentioned that they generally refer to systems to track changes in databases, naming conventions, version controls for numerical models, and Vault for CAD. For Group B, data traceability was ensured via commercial solutions for version tracking such as Microsoft SharePoint, or dedicated software such as AVL CRETA for calibration and AVL FUSE for functional safety.

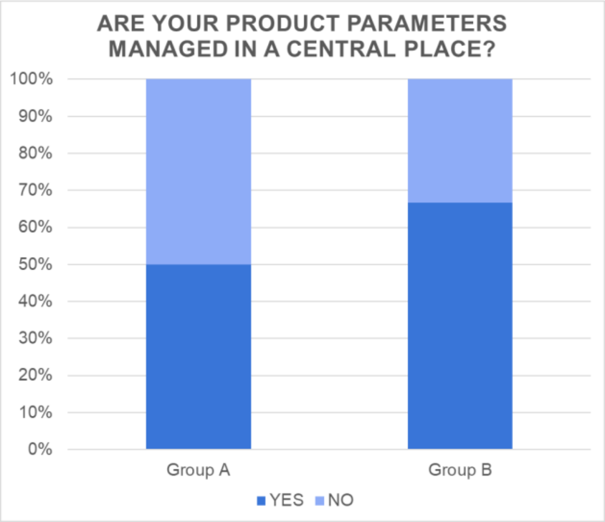


Figure 17: Question 13: Are your product parameters managed in a central place?

4.2.6 Centralised Management

Only 50% of the respondents in Group A use centralised systems for managing product parameters, while the percentage increases to 66% for users in Group B (see Figure 17). The parameters are saved in central databases, via VPN, using SAP systems, Microsoft SharePoint or AVL Creta. Furthermore, only 40% of users in Group A generally uses a centralised place for managing the models, while still the 66% of users in Group B does it, as shown in Figure 18. This is done using GitLab repositories, servers, Microsoft SharePoint or web based in-house tools.

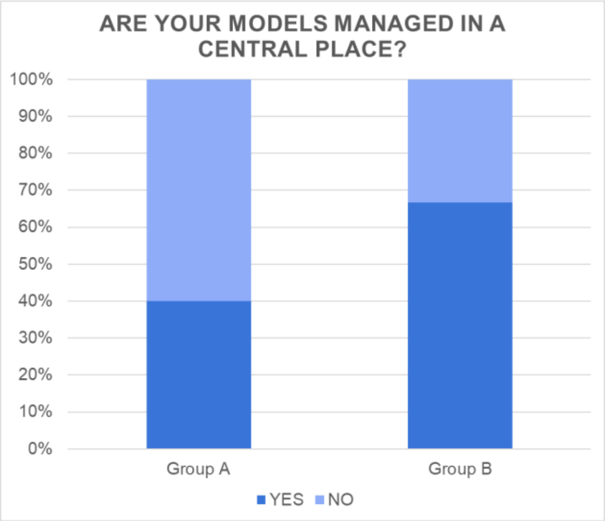


Figure 18: Question 15: Are your models managed in a central place?

A similar trend is noticed for the management of measurement data (Figure 19). Only 40% of the users in Group A manage measurements in a central place, with respect to more than 80% in Group B. Databases, SCADA applications, or in-house developments as AVL Santorin/Santorin MX are some of the options that the respondents indicated for the central managements of measurements.

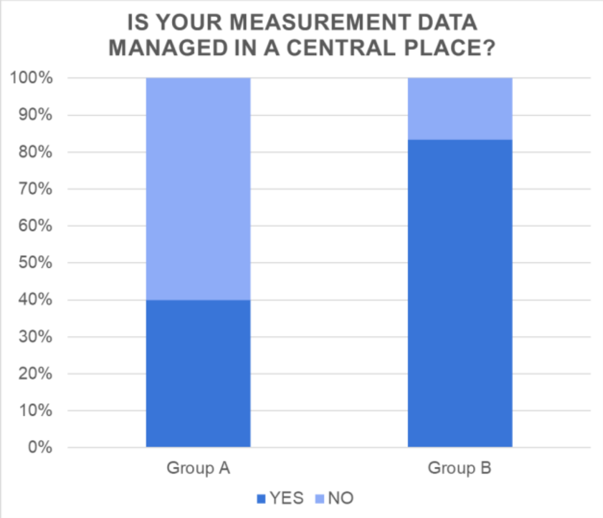


Figure 19: Question 17: Is your measurement data managed in a central place?

4.2.7 Commercial Data

Respondents were asked if their commercial data were linked to project data (Figure 20). The response to this question was affirmative for 40% of the respondents in Group A, and for more than 80% of the respondents in Group B. When the link exists, it is implemented via SAP or file/data structures.

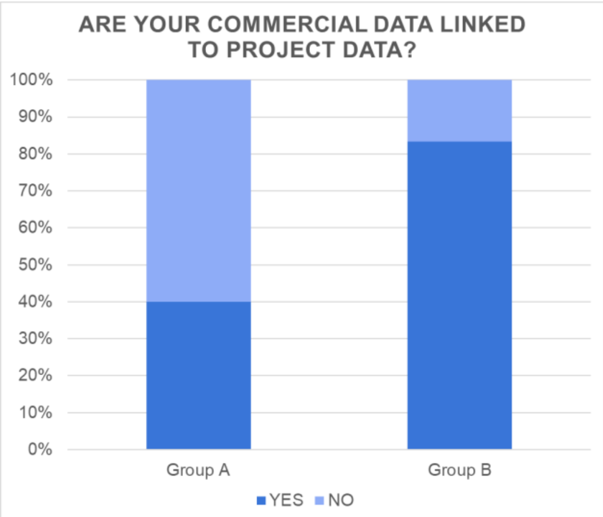


Figure 20: Question 19: Are your commercial data linked to project data?

4.2.8 Metadata

Most of the respondents of Group A (80%) reported to apply data science techniques during the post processing stage, for example to automatize the detection of failures and other types of critical events. This percentage of data science practitioners drops down to 33% among the respondents in Group B. The purpose of data science methods was to signal the occurrence of critical deviations from operational conditions.

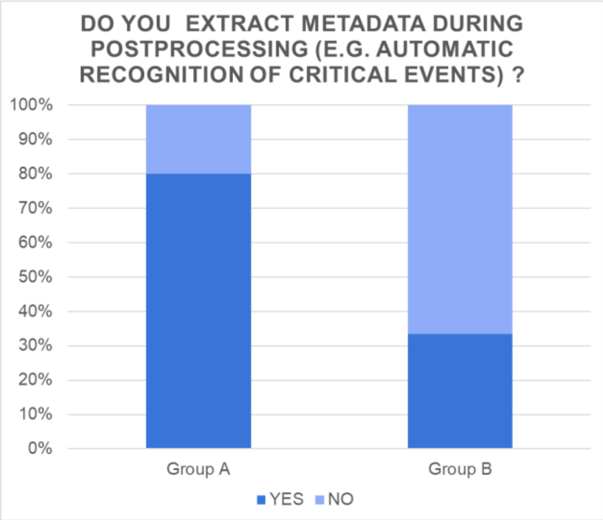


Figure 21: Question 21: Do you extract metadata during postprocessing (e.g., for automatic detection of critical events)?

4.3 Data Maintenance Cycle

4.3.1 Data Pre- and Post-processing

The first two questions in the Data Maintenance Cycle section of the survey were intended to investigate the importance of planning ahead the time for pre-processing (see Figure 22) and post-processing of (see Figure 23) the data, which is a preventive action to minimize the risk of delays in the preparation of the experimental campaign or in the analysis of the outcomes.

Group B gave the time requirements for these operations a much higher relevance than Group A, as suggested by the difference in the average ratings between the two groups (9.5 compared to 7.8). Both groups were therefore asked to indicate their preferred methods for processing the signals (see Figure 24). Although the answers were quite dispersed, methods based on filtering seemed to be predominant with respect to advanced statistics (e.g., regression) (less than 30% of the answers for both groups).

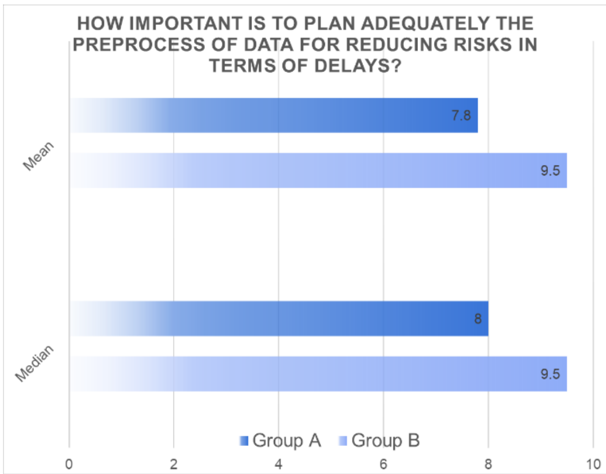


Figure 22: Question 1: How important is to plan adequately pre-processing of the data to reduce the risk of delays?

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

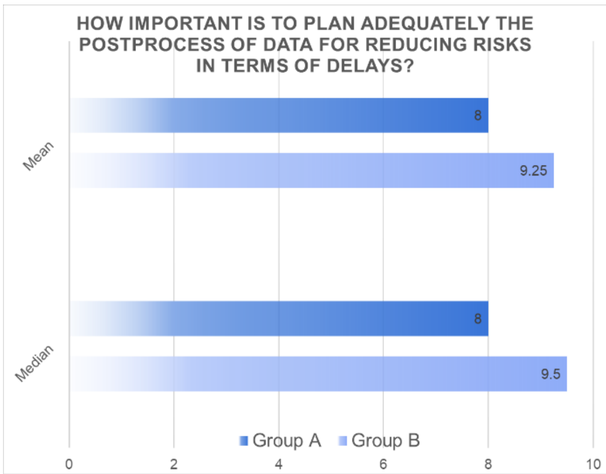


Figure 23: Question 2: How important is to plan adequately post-processing of the data to reduce the risk of delays?

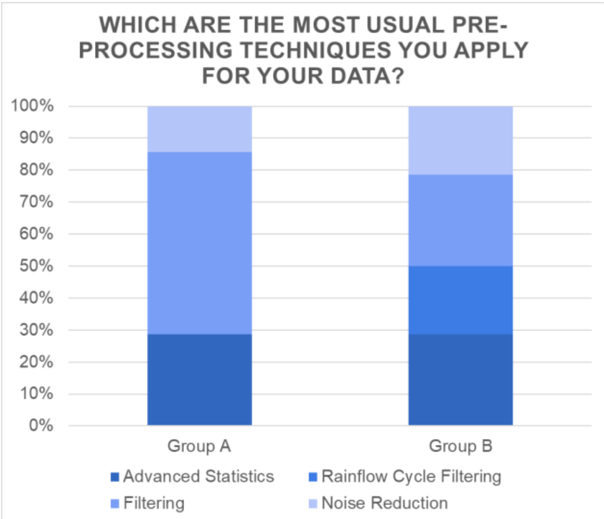


Figure 24: Question 3: Which are the most usual pre-processing techniques you apply to your data?

4.3.2 Data Accessibility and synchronisation

In Question 4, the respondents were asked how they organized the accessibility to the data for different types of users. More than 80% of the respondents in Group B reported that access to data is normally restricted only to people working on a specific project, whereas for the remaining (less than) 20%, the access is usually granted to everybody within their organization or the whole consortium (see Figure 25). These ratings changed to more than 40% for Group A, with a fraction of respondents who even opted to share their data openly within their community of potential users.

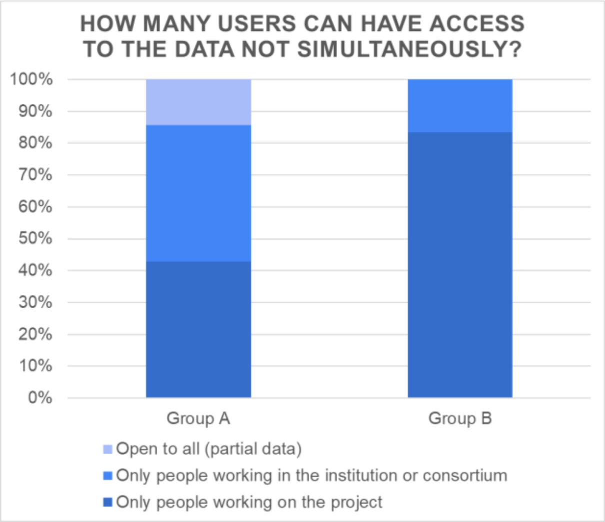


Figure 25: Question 4: How many users can have access to the data not simultaneously?

A different trend was observed in the answers to Question 5, where the respondents were asked the number of users that could access to the data simultaneously. For Group A, 80% of the respondents answered that the data should be available for as many people as it needs. The remaining 20% answered that the data should be available for one person at the time. In Group B, again the most frequent answer was that data should be available for as many users as required (50% of the answers), followed by the possibility to restrict further the number of users (around 30% of the answers), and for the remaining share (less than 20%) only one user at the time is granted access to the datasets.

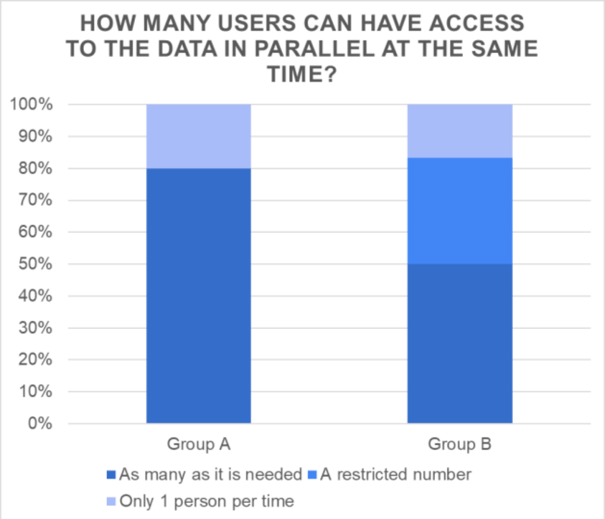


Figure 26: Question 5: How many users can have access to the data at the same time?

4.3.3 Metadata Management

None of the respondents in Group B reported to use any dedicated tool to manage metadata, in contrast to 20% of the respondents in Group A who did it (see Figure 27). Essentially, one respondent in Group A treats timeseries in a specific database for storing and accessing the signals.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

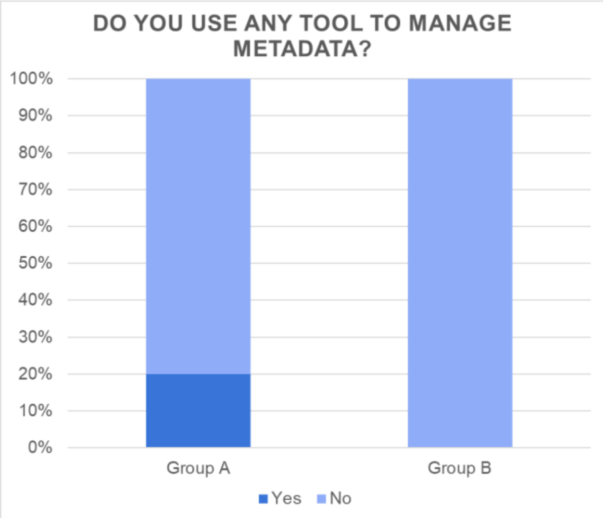


Figure 27: Question 6: Do you use any tool to manage metadata?

4.3.4 Use of Standards

A sizeable discrepancy was found between groups of respondents who answered Question 8 (see Figure 28). The majority of respondents in Group A (80%) reported that no standards are available in their domain, while the remaining (20%) expressed a lack of knowledge about the existence of pertinent standards. In contrast, 66% of respondents in Group B declared that standards are available in their domain, while the remaining 33% were equally distributed between the options "No" and "Don't know".

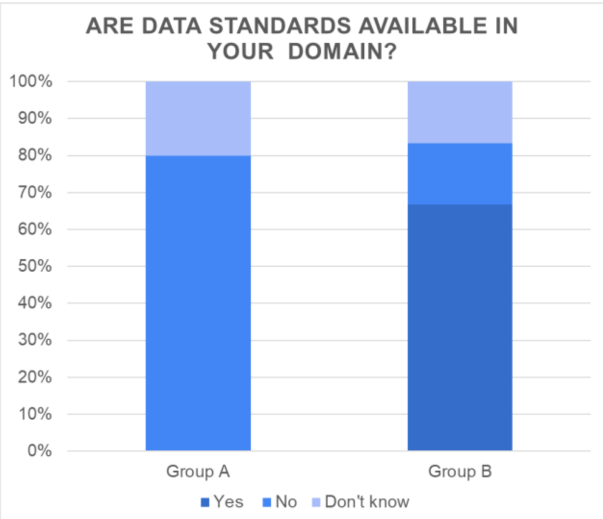


Figure 28: Question 8: Are data standards available in your domain?

4.3.5 Data Sharing

None of the respondents in Group A reported to use standardized data formats, while Group B was equally split between users and not-users of standardized formats (see Figure 29). The adopted data formats are generally software-driven, i.e. they are selected on the basis of compatibility with software tools (for example, open FOAM or MATLAB) or as in the case of bathymetries which use the netCDF format.

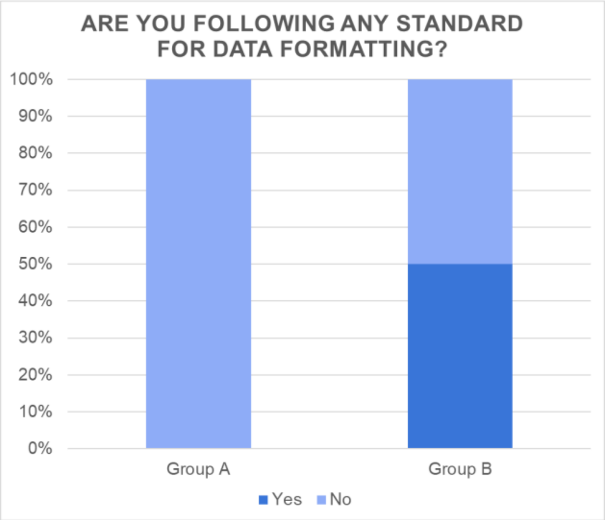


Figure 29: Question 9: Are you following any standard for data formatting?

In Group B, data internal to the project resulted to be shared mainly via SharePoint (70%) or Dropbox (30%). In Group A other alternative were proposed, such as servers, databases, undefined APIs, although the most common platform for data exchange was SharePoint also in this group (43%, see Figure 30).

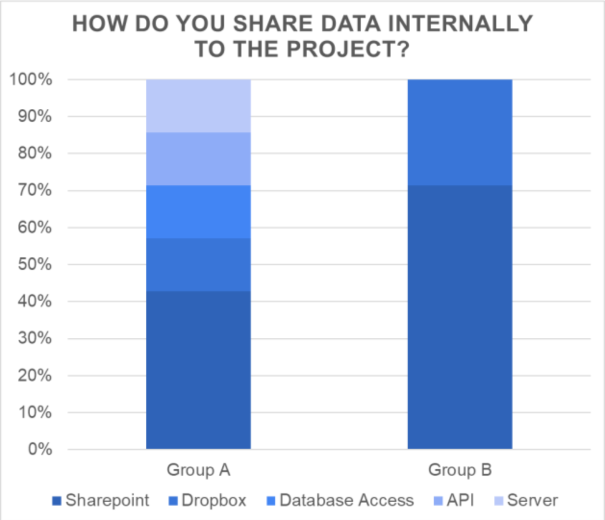


Figure 30: Question 11: How do you share data internally to the project?

Regarding data sharing with partners outside the project consortium, SharePoint was still the most frequently used solution (42% in Group A and 70% in Group B). Other reported options were Dropbox, or local solutions such as websites, APIs, etc (see Figure 31).

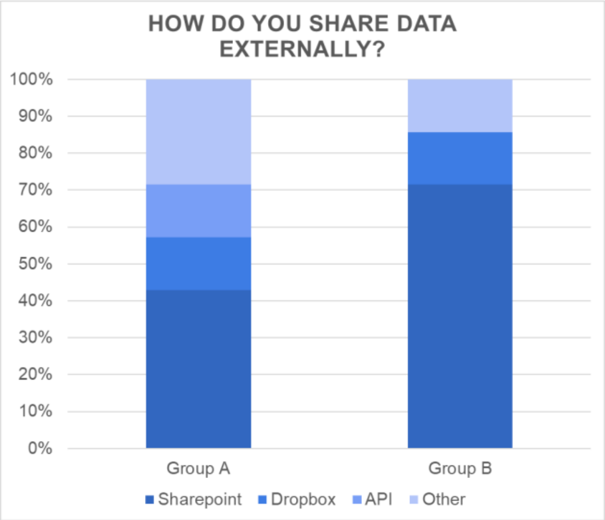


Figure 31: Question 12: How do you share data outside the project consortium?

4.3.6 Data Quality

Respondents in both Group A and B attributed high importance to the quality of the data, as indicated by the median rating equal to 10 in the answers to Question 13 from both groups (see Figure 32).

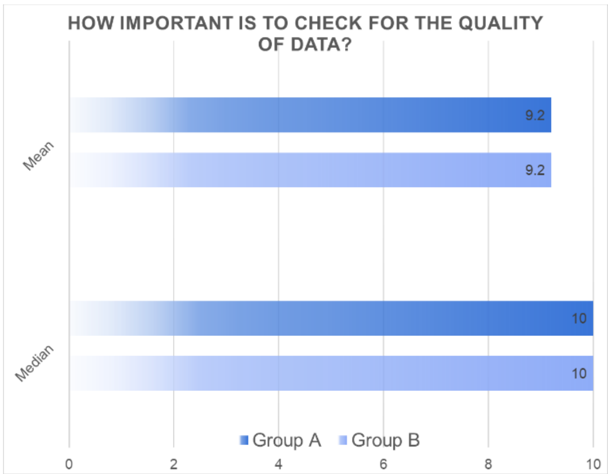


Figure 32: Question 13: How important is to check for the quality of data?

4.3.7 Data processing time and accuracy

The results presented in Figure 33 suggest a substantial consensus between Group A and B that a good trade-off between accuracy and speed of the simulations is either to accept slow pre-process when higher accuracy is needed, or to accelerate pre-process when low accuracy could be admissible.

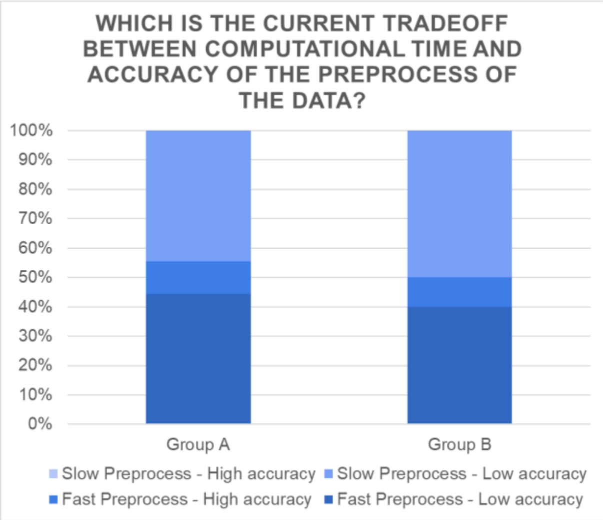


Figure 33: Question 14: Which is the current trade-off between computational time and accuracy of the pre-process of the data?

4.3.8 Data Sources

All the respondents in Group A reported to be familiar with using data from different sources. The percentage drops down to 67% among the respondents in Group B (see Figure 34). The analysis of the sources of data indicated that both groups collect data from a mix of sources which includes public repositories, real time simulations, internal and external databases. In Group A, the respondents included also offline simulations in the list of data sources considered in some cases (see Figure 35).

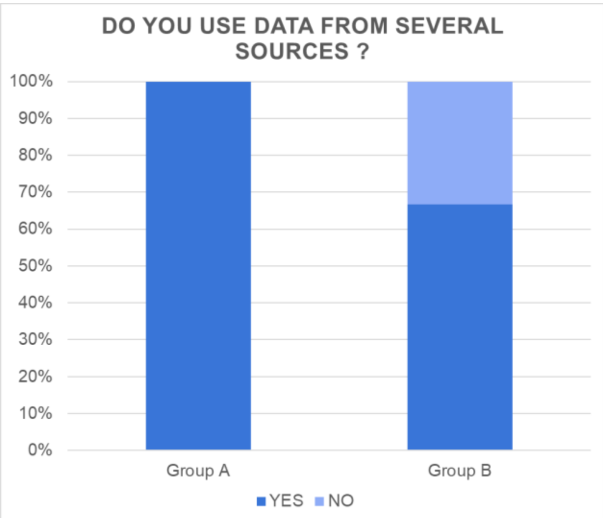


Figure 34: Question 15: Do you use data from several sources?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

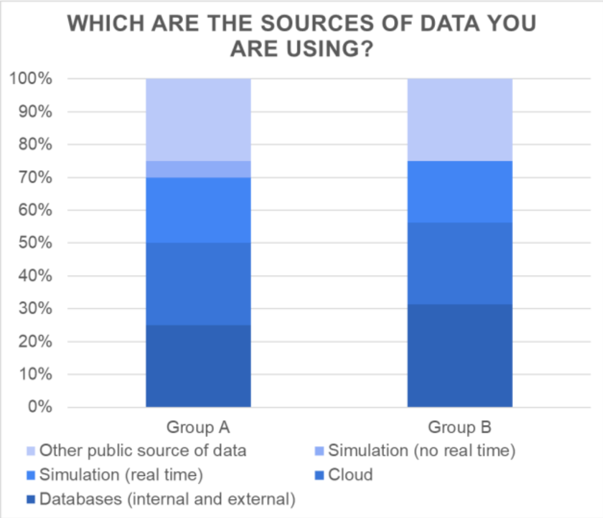


Figure 35: Question 16: What sources of data you are using?

4.3.9 Data Consistency/Robustness Checks

Only 40% of the respondents in Group A and 50% in Group B reported to carry out a real time check of data integrity, that is their consistency and robustness (see Figure 36). When such a check is performed, commercial solutions or in-house developed tools seemed to be chosen with the same frequency (50% for both groups).

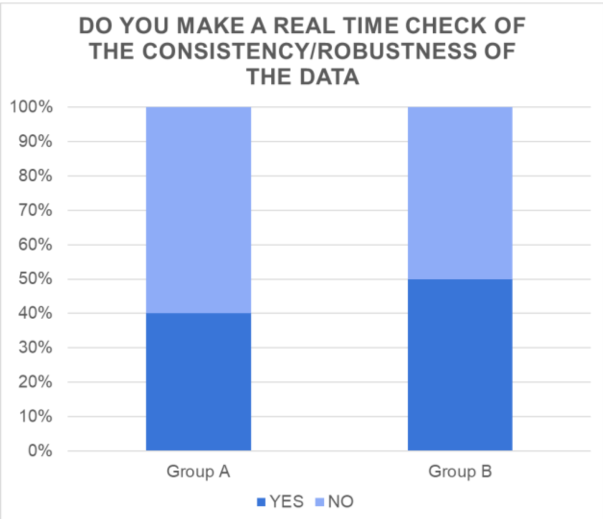


Figure 36: Question 17: Do you make a real time check of the consistency/robustness of the data?

4.3.10 Data Redundancy

Only 40% of the respondents in Group A reported to apply redundancy in data acquisition, while this share increased to 60% in Group B (see Figure 37). Redundancy was said to be implemented typically by the usage of virtual sensors and/or numerical models, but in some cases additional physical sensors are also deployed.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

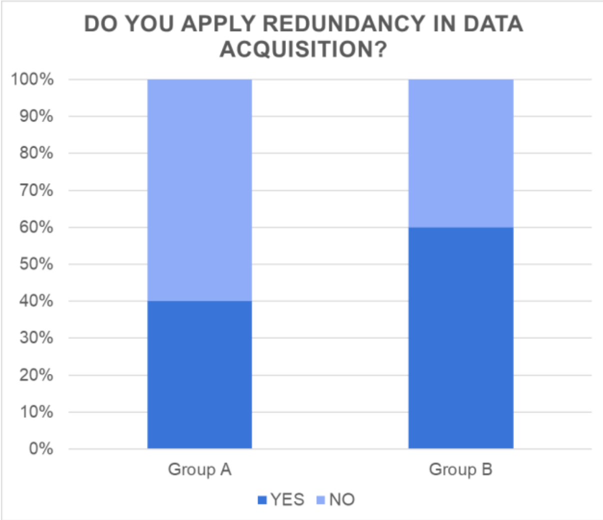


Figure 37: Question 19: Do you apply redundancy in data acquisition?

4.3.11 Data Versioning

Only 20% of to the respondents in Group A reported to use some tool for versioning the data (typically, Gitlab). This share increased to 60% for Group B, and Gitlab as well as alternative tools such as AVL Creta were indicated as technical solutions for data versioning (see Figure 38).

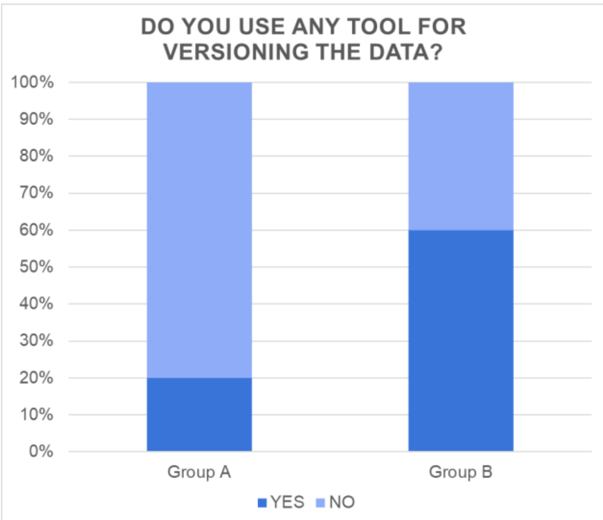


Figure 38: Question 21: Do you use any tool for versioning the data?

4.3.12 Data Organisation

Only 40% of the organizations that participated in the survey (both in Group A and Group B) declared to organise their data in a way that facilitates the access to the most updated data, which are generally stored in clouds, intranets, or even locally (see Figure 39).

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

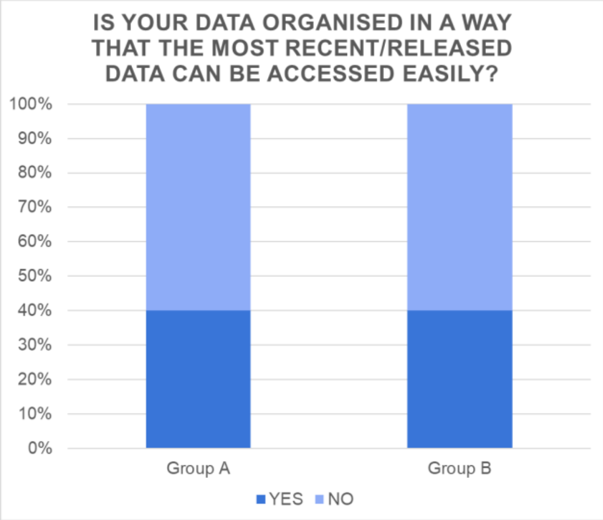


Figure 39: Question 23: Is your data organised in a way that the most recent/released data can be accessed easily?

4.3.13 Data Formats

Figure 40 and Figure 41 display the responses from Group A and B, respectively about how typical data found in the design and testing applications considered in the VALID project are distributed over the four most common data formats: scalar, array, multidimensional array, and metadata The usage of metadata is indicated to be restricted only to describe specifications (Group B) or constraints (Group A), while in most of the other cases, scalar quantities, arrays, or multi-dimensional arrays constitute the most common formats.

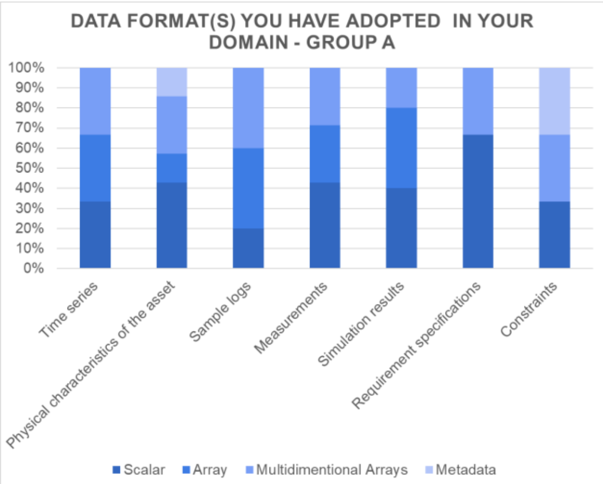


Figure 40: Question 25: used data formats – Group A

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

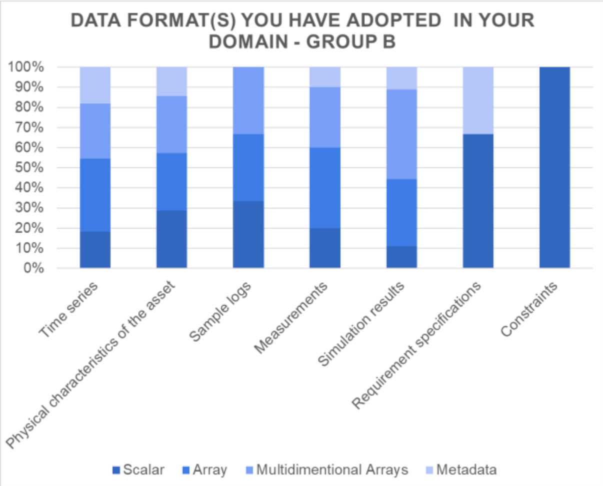


Figure 41: Question 25: used data formats – Group B

The results presented in Figure 42 indicate that saving one file per test represents the most common strategy for Group A. Alternative approaches such as using workspaces, saving data junks, or even saving a file for each variable (Group B) were also reported to be relatively common.

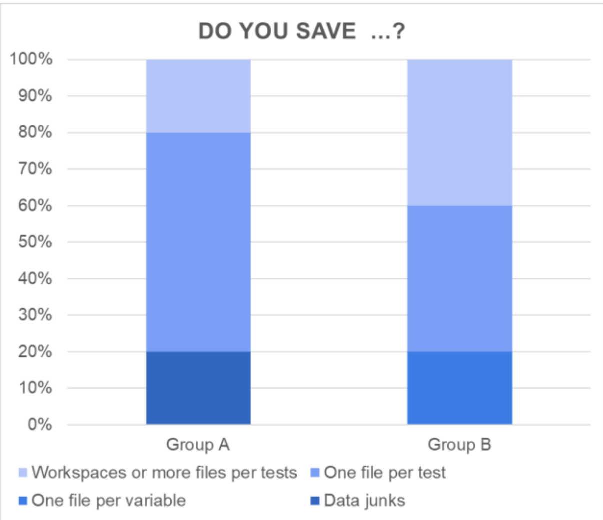


Figure 42: Question 26: Do you save…?

4.3.14 Management of Large Amounts of Data

The aim of Question 27 in the survey was to map the size of datasets that are typically managed in hybrid testing. Half of the test rig developers, managers, and owners represented in Group A reported that their operations require relatively small input datasets (less than 1 MB). In contrast, about 66% of the respondents in Group B indicated a larger size range for typical input datasets (between 1 and 100 MB), as shown in Figure 43. Regarding the amount of data produced during a hybrid test, the majority of respondents in both groups indicated that most of the output datasets are definitely larger than 1 MB and, according to Group B, even larger than 100 MB (75% of the answers from that group).

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

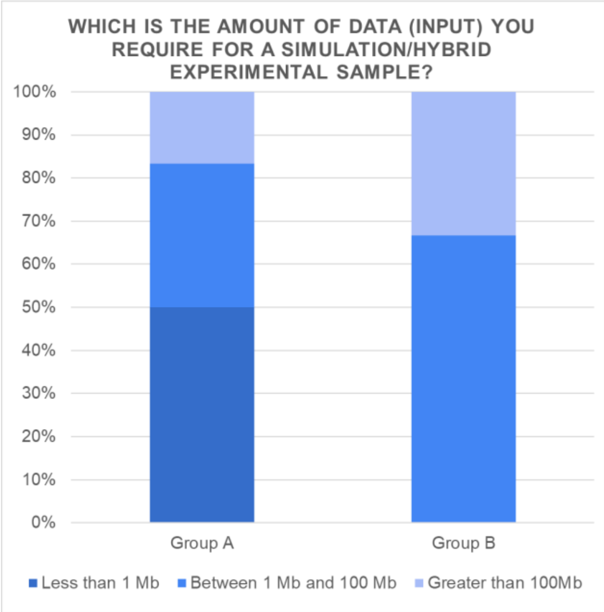


Figure 43: Question 27: Which is the amount of (input) data you require for a simulation/ hybrid experimental sample?

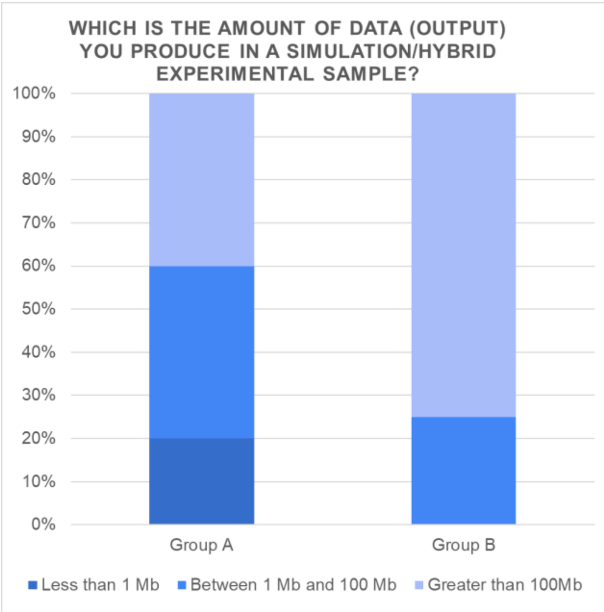


Figure 44: Question 28: Which is the amount of (output) data you produce in a simulation/ hybrid experimental sample?

The majority of the respondents in Group A (75%) reported to require or have already available at least 32 channels in their logging systems, whereas the answers from Group B were fragmented in several intervals (see Figure 45).

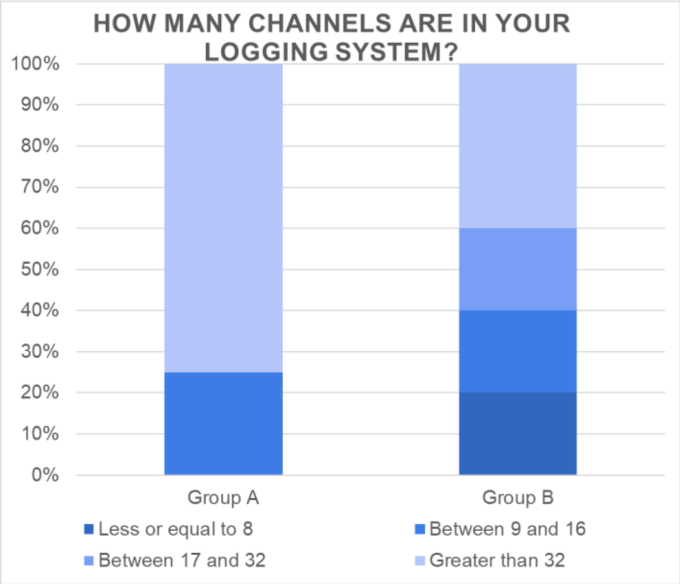


Figure 45: Question 29: How many channels are in your logging system?

All the respondents in Group A and B reported to record time series for at least 5 minutes and for 60% of them, the minimum duration of recorded time series raised up to 30 minutes (see Figure 46). Regarding sampling rates (Figure 47), 60% of the test rig developers and managers who participated to the survey indicated that sampling frequencies could be higher than 10 Hz, whereas for the remaining 40% the range of typical sampling frequencies was comprised between 5 and 10 Hz. Half of respondents in Group B considered also saving the data with a frequency less than or equal to 1 Hz.

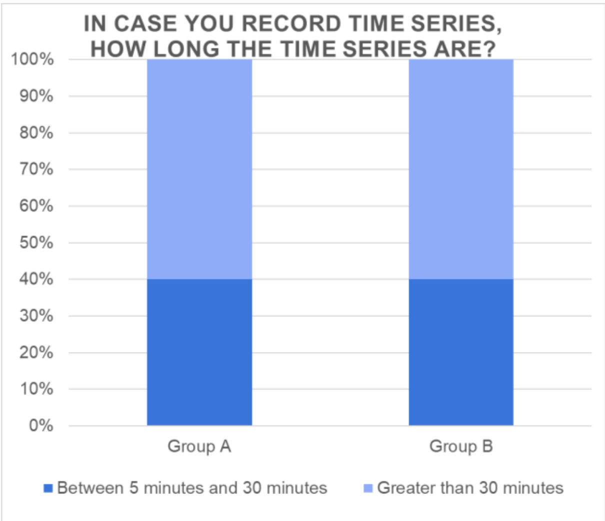


Figure 46: Question 30: In case you record time series, how long the time series are?

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

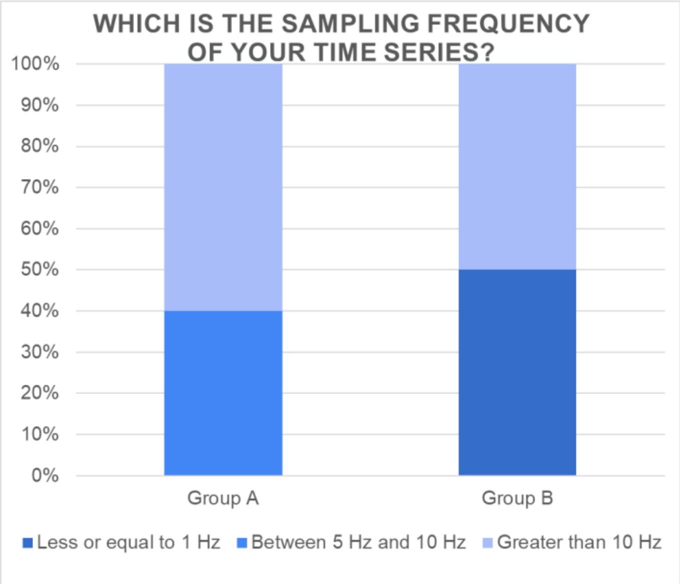


Figure 47: Question 31: Which is the sampling frequency of your time series?

5 Data Management Recommendations for the hybrid Testing Methodology

5.1 Overall Data Management Requirements

As pointed out in Section 3, from a methodological perspective the results of the survey cannot be used in a statistical sense for drawing general conclusions about requirements for data handling in hybrid testing of wave energy converters. However, the survey has provided a sufficiently broad list of needs of wave energy developers and test rigs managers in the framework of the VALID project, as well as an insight of the capabilities and expertise matured within the consortium. The detailed critical review of the requirements per section will be done in Section 5.2 and Section 5.3. In general, the answers provided from Group A show a clear trend and they will be marked inside a box. These are recommendations that should be followed while planning, carrying out, and analysing the results of the hybrid test campaign in VALID. However, in some cases, a specific requirement cannot be identified from the answers of Group A, hence the need for the adoption of tailored solutions.

In general, respondents in Group A have shown substantially convergent views about their needs:

- **Automatic workflows** for data management are considered instrumental to increase the robustness of the dataflow, the integrity of data, automatic and continuous checks on the KPIs and extraction and treatment of metadata.
- Current systems for **data storage** at the organizations represented in Group A are characterized by limited robustness, as highlighted by several errors detected in the past as a result of the unintended use of outdated data. In most cases, no centralised repositories for data and models are used or implemented at the organizations which participated to the survey.
- The absence of any **data standards** for the management of the information leads to a wide range of adopted data formats.
- **Quality of data** is indicated as a priority for Group A, although techniques for increasing measurement redundancy or versioning of data are reportedly no part of current practice.
- **The management of a big amount of output data** presents specific challenges, due to the combination of length of the time series, the sampling frequency and the number of channels in the logging systems.

More diversity was found in the answers of Group A regarding the tools and data formats that are currently used in processes. This can be partly justified to the specificness of the study. However, it seems reasonable that within the framework of the VALID project, for improving the quality of the outcome, some measures could be generalised, as for example using Gitlab repositories for management of the models (if public), at storage in clouds, data sharing via common platforms.

The survey highlighted **some** strong **similarities between the wave energy and the automotive sectors**, regarding the expected requirements for efficient data management in product development. The features identified from the survey and listed in Section 2.3 are found to be equally relevant for both industrial sectors: traceability of simulation results within an integrated environment that allows for continuous development and storage of data about physical and virtual artifacts, as well as the assessment of KPIs. However, some **differences** were also noticed, and interpreted as an effect of the different level of product maturity. Indeed, whereas the use of a **standardized process** to build up trust in the results of simulation/test is seen as a requirement in the automotive sector, the lack of standardisation in the wave energy sector is identified as a gap. This means that even if the wave energy sector considers the use of standardised procedures as a priority, still the implementation of standardisation procedures is impossible due to the lack of their own standards.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

5.2 Data Intelligence Requirements

Automated approaches for data management workflows appeared to be important to test rig managers and wave energy developers in the VALID project. Their implementation requires adequate measures, such as:

- **To increase the robustness of the data storage systems, using for example storage in the cloud, centralised file systems, local and remote access databases**.

- **The systems should guarantee versioning of the data and easy access to the most updated data**. Respondents in Group A reported the occurrence of errors in the past due to the usage of wrong/outdated data. Although the traceability of data was rated "medium", versioning of models and data is not part of current practice according to respondents in Group A. Systems that would improve the traceability of data, a would avoid (or, at least, limit) the occurrence of errors that result from the use of outdated data.

- User case developers and test rig managers **have already access to systems that enable tracking and visualization of important KPIs**.

- **Centralisation of information is not seen as critical** by respondents in Group A and indeed it is not implemented in the current systems. However, the answers of Group B revealed that some members of the consortium have developed significant expertise on this approach to information management, which could be deployed in the VALID project to increase the robustness of the storage systems.

- Besides measuring and visualising relevant KPIs, the hybrid test platform developed in the VALID project, should be capable to **build metadata and apply basic and advanced techniques of data analytics** and data science the occurrence of critical events and degradation. These techniques are already included in the operations of some of the organizations represented in the survey, and their application may be extended to all the user cases.

The list above is not exhaustive and additional measures could be introduced for each User Case. The best approach for managing the data in each case depends also on the skills present in the consortium, as confirmed by the answers of Group B. Therefore, an effort in the implementation of such measures should be carried out during the hybrid testing campaigns in order to reach the goals of the project. Supporting partners should help in identifying appropriate alternatives for each User Case.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

5.3 Data Maintenance Cycle Requirements

During the different phases of the lifecycle of the hybrid testing, the following requirements have been identified:

- Users in Group A generally assigned medium importance to the time allocated to pre-processing and post-processing of the signals. Furthermore, a variety of techniques were considered to analyse the signals. In contrast, users in Group B assigned a higher importance to these tasks. A possible explanation for the different attitudes shown by the two groups is that Group A might have already overcome some technical challenges in previous physical test campaigns. However, given the concerns raised by Group B, **adequate time** should be reserved **for the pre-process and postprocess of the data when planning the test campaigns**, in order to minimize the risk for unwanted delays. It is generally **accepted longer pre/post-process of data when increasing the accuracy**.

- The platforms (e.g., databases, file systems, clouds) chosen for storage and processing of the data should be **accessible to as many users as needed by the hybrid testing procedure**. Accessibility specifications are a strong driver in the selection of the data platform, which should be able to support different levels of restrictions and responsibilities for the users (e.g., read-only privileges, managers or owners of the data) in order to avoid data corruption, thus preserving the integrity of the datasets. Accessibility should account for the requirements described in Section 5.2, such as the adoption of appropriate versioning systems.

- **Metadata, when available, are not managed by any dedicated tool.** This does not seem to be required, therefore, during the hybrid testing campaign in the VALID project.

- According to Group A, **no data standards are available.** This is evidently a limit of the wave energy sector, and it cannot be fully addressed within the extent of the VALID project. The lack of standardization in data format hinders the sharing of data with external parts, and it practically requires using several data formats for the same physical quantities. This lack of a common language or accessible translation tools affect the quality of data (for example, in the sampling rates and positioning of the sensors), which was also indicated as a highly relevant requirement by the respondents in Group A.

- Despite data quality is seen as a priority, no real time consistency or robustness checks are currently performed during physical testing, according to the respondents in Group A. Furthermore, no techniques for increasing redundancy in the physical testing are currently implemented. The **installation of extra physical sensors or the implementation of some basic virtual sensors** based on indirect measurement might be included in the VALID hybrid testing protocols, in order to ensure data robustness and reliability.

- **Big amount of data must be managed during VALID hybrid testing**. The number of logging channels required (greater than 32), the sampling frequencies (higher than 10 Hz), the length of the time signals (at least 5 minutes in real scale) as well as the number of tests for load cases, environmental conditions, etc… make the amount of data to be analysed particularly cumbersome to handle. Another concern associated

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

with the number of tests and the length of each signal is about the duration of the experimental campaign, as the task of acceleration of tests via hybrid testing could turn out to be challenging.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

6 Nomenclature

Abbreviations

API	Application Programming Interface
BoM	Bill of Material
CAD	Computer-Aided Design
EC	European Commission
EU	European Union
H2020	Horizon 2020
IODP	Integrated and Open Development Platform
KPI	Key Performance Indicator
OEM	Original Equipment Manufacturer
PDM	Product Data Managmement
PDP	Product Development Process
PLC	Product Lifecycle Collaboration
PLM	Product Lifecycle Management
PTO	Power Take-Off
SCADA	Supervisory Control and Data Acquisition
TRL	Technology Readiness Level
UC	User Case
VDP	Vehicle Development Process
VPN	Virtual Private Network
WEC	Wave Energy Converter
WP	Work Package

7 References

[1] Hirz, Mario. (2009). Basics of Product Data Management in Automotive Engineering.

[2] http://www.npd-solutions.com/pdm.html

[3] Puntigam, W., Zehetner, J., Lappano, E., Krems, D. (2020). Integrated and Open Development Platform for the Automotive Industry. In: Hick, H., Küpper, K., Sorger, H. (eds) Systems Engineering for Automotive Powertrain Development. Powertrain. Springer, Cham. https://doi.org/10.1007/978-3-319-68847-3_28-1

[4] proStep, 2003; White Paper for the Use of Standards based Data Communication Methods in Automotive Industry

[5] https://www.ni.com/de-at/solutions/transportation/automotive-data-management.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

Annex 1. The Questionnaire

A questionnaire to be filled by the partners in the VALID consortium has been the main tool for collecting information about the current approaches and the needs in terms of data management during the hybrid testing campaigns.

The questionnaire is available at the link:

https://docs.google.com/forms/d/e/1FAIpQLSdDgNgG2hXESb1OOUCrKHjZEFg26JF9W0iPKkQriB4iLf916w/viewform?usp=sf_link

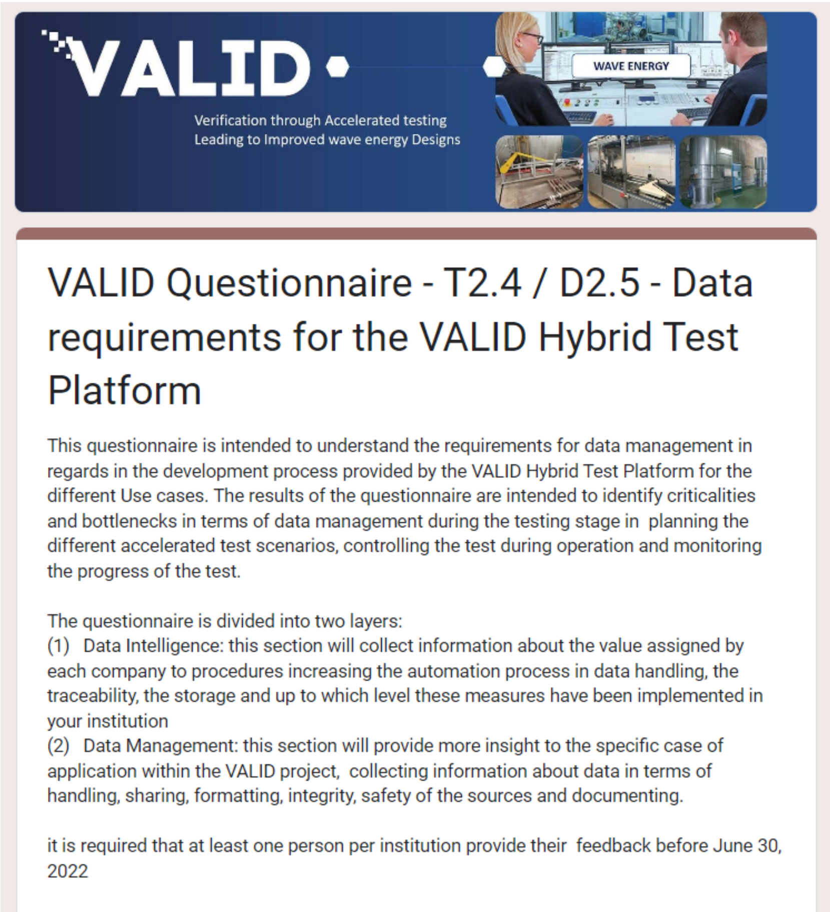
A part of an introductory section, the questionnaire included three sections:

- Registry
- Data Intelligence
- Data Management cycle

A total of 67 questions were included.

Introduction

The landing page is an introductory section (see Figure 48), describing the aim of the survey, the sections it is structured and some formal requirements (initial deadline and people required to answer).



VALID Questionnaire - T2.4 / D2.5 - Data requirements for the VALID Hybrid Test Platform

This questionnaire is intended to understand the requirements for data management in regards in the development process provided by the VALID Hybrid Test Platform for the different Use cases. The results of the questionnaire are intended to identify criticalities and bottlenecks in terms of data management during the testing stage in planning the different accelerated test scenarios, controlling the test during operation and monitoring the progress of the test.

The questionnaire is divided into two layers:
(1) Data Intelligence: this section will collect information about the value assigned by each company to procedures increasing the automation process in data handling, the traceability, the storage and up to which level these measures have been implemented in your institution
(2) Data Management: this section will provide more insight to the specific case of application within the VALID project, collecting information about data in terms of handling, sharing, formatting, integrity, safety of the sources and documenting.

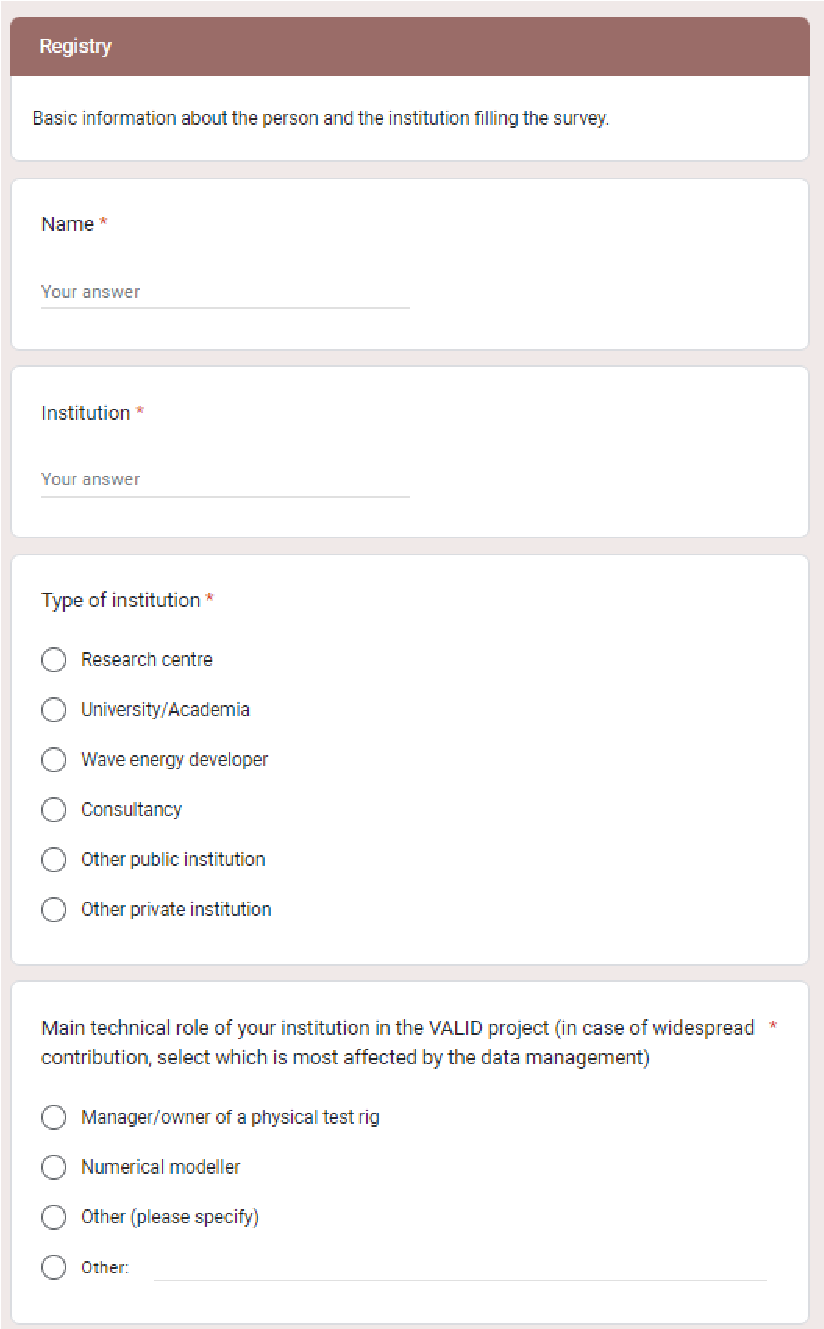
it is required that at least one person per institution provide their feedback before June 30, 2022

Figure 48: Introduction to the questionnaire

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

Registry

In the registry section (see Figure 49), the user is asked to fill basic information in order to identify them, the institution represented and the role of the institution inside the VALID project. It was asked, indeed, to answer to the questions relatively to the scope of the university/company/research centre within the project.


Registry

Basic information about the person and the institution filling the survey.

Name *

Your answer

Institution *

Your answer

Type of institution *

- Research centre
- University/Academia
- Wave energy developer
- Consultancy
- Other public institution
- Other private institution

Main technical role of your institution in the VALID project (in case of widespread contribution, select which is most affected by the data management) *

- Manager/owner of a physical test rig
- Numerical modeller
- Other (please specify)
- Other:

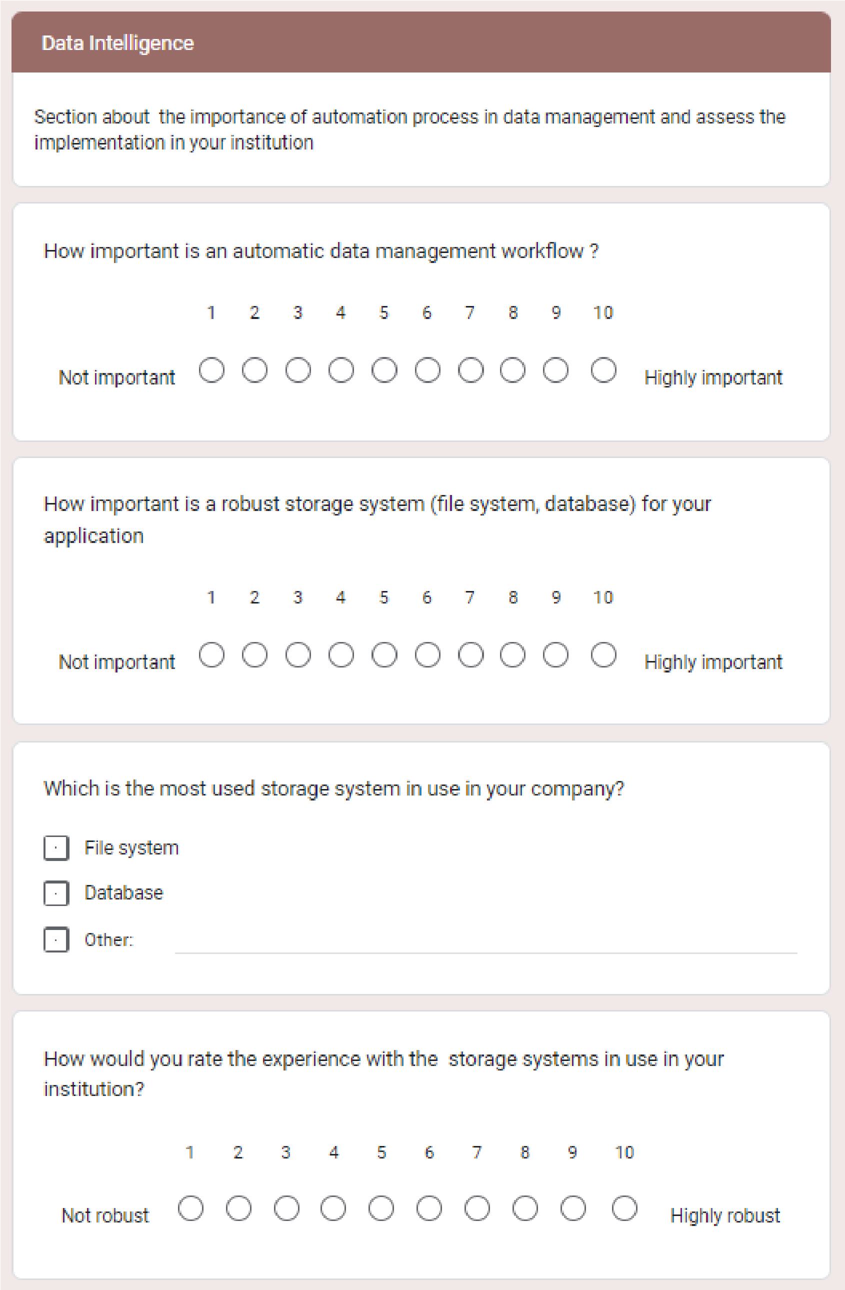
Figure 49: Registry Section

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.

Data Intelligence

In the Data Intelligence sections (see Figure 50, Figure 51, Figure 52, Figure 53 and Figure 54) information pertinent to the automation process in data managements in the different institution is collected. A total of 22 questions are asked. The user is asked about the relevance of automated data workflows in their tasks, the robustness of the data storage approaches used in their institution and the occurrence of failures in data treatment. Moreover, in order to understand the needs of the company, the survey asks about the importance of tracking KPIs real time, i.e., during the realisation of the experiment, and visualisation of metadata. Information about the centralisation of data and models as well as the usage and treatment of public data is finally gathered to inform the needs in terms of shared resources and confidentiality.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101006927.



Data Intelligence

Section about the importance of automation process in data management and assess the implementation in your institution

How important is an automatic data management workflow ?

	1	2	3	4	5	6	7	8	9	10	
Not important	○	○	○	○	○	○	○	○	○	○	Highly important

How important is a robust storage system (file system, database) for your application

	1	2	3	4	5	6	7	8	9	10	
Not important	○	○	○	○	○	○	○	○	○	○	Highly important

Which is the most used storage system in use in your company?

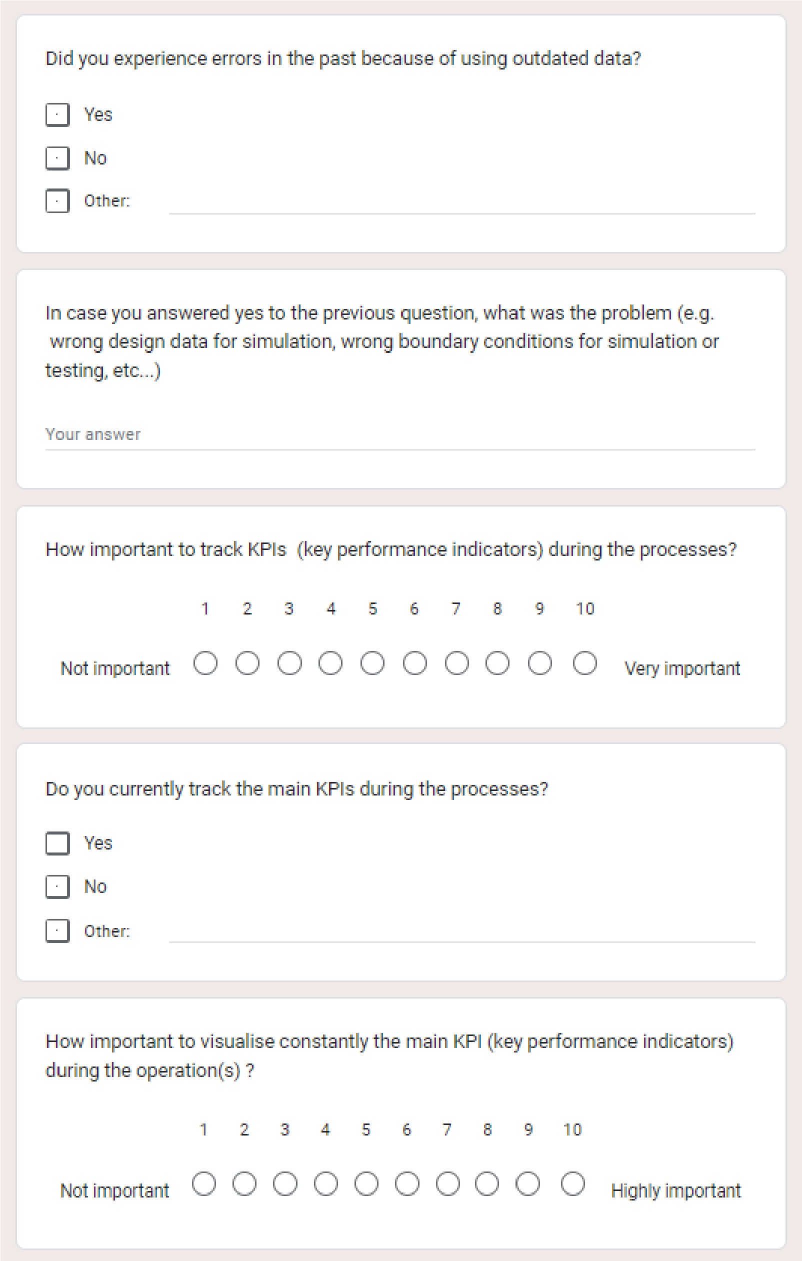
- ☐ File system
- ☐ Database
- ☐ Other: ____________

How would you rate the experience with the storage systems in use in your institution?

	1	2	3	4	5	6	7	8	9	10	
Not robust	○	○	○	○	○	○	○	○	○	○	Highly robust

Figure 50: Data Intelligence (I)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.



Did you experience errors in the past because of using outdated data?

- [ ] Yes
- [ ] No
- [ ] Other:

In case you answered yes to the previous question, what was the problem (e.g. wrong design data for simulation, wrong boundary conditions for simulation or testing, etc...)

Your answer

How important to track KPIs (key performance indicators) during the processes?

	1	2	3	4	5	6	7	8	9	10	
Not important	○	○	○	○	○	○	○	○	○	○	Very important

Do you currently track the main KPIs during the processes?

- [ ] Yes
- [ ] No
- [ ] Other:

How important to visualise constantly the main KPI (key performance indicators) during the operation(s) ?

	1	2	3	4	5	6	7	8	9	10	
Not important	○	○	○	○	○	○	○	○	○	○	Highly important

Figure 51: Data Intelligence (II)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

VALID

Is the system that is currently in use in your institution able to visualise important KPIs real time?

- [ ] Yes
- [ ] No
- [ ] Other: ____________

How important is traceability of data throughout all the lifecycle of the project?

	1	2	3	4	5	6	7	8	9	10	
Not important	○	○	○	○	○	○	○	○	○	○	Highly important

Which are the traceability systems adopted in your institution?

Your answer ____________

Are your product parameters managed in a central place?

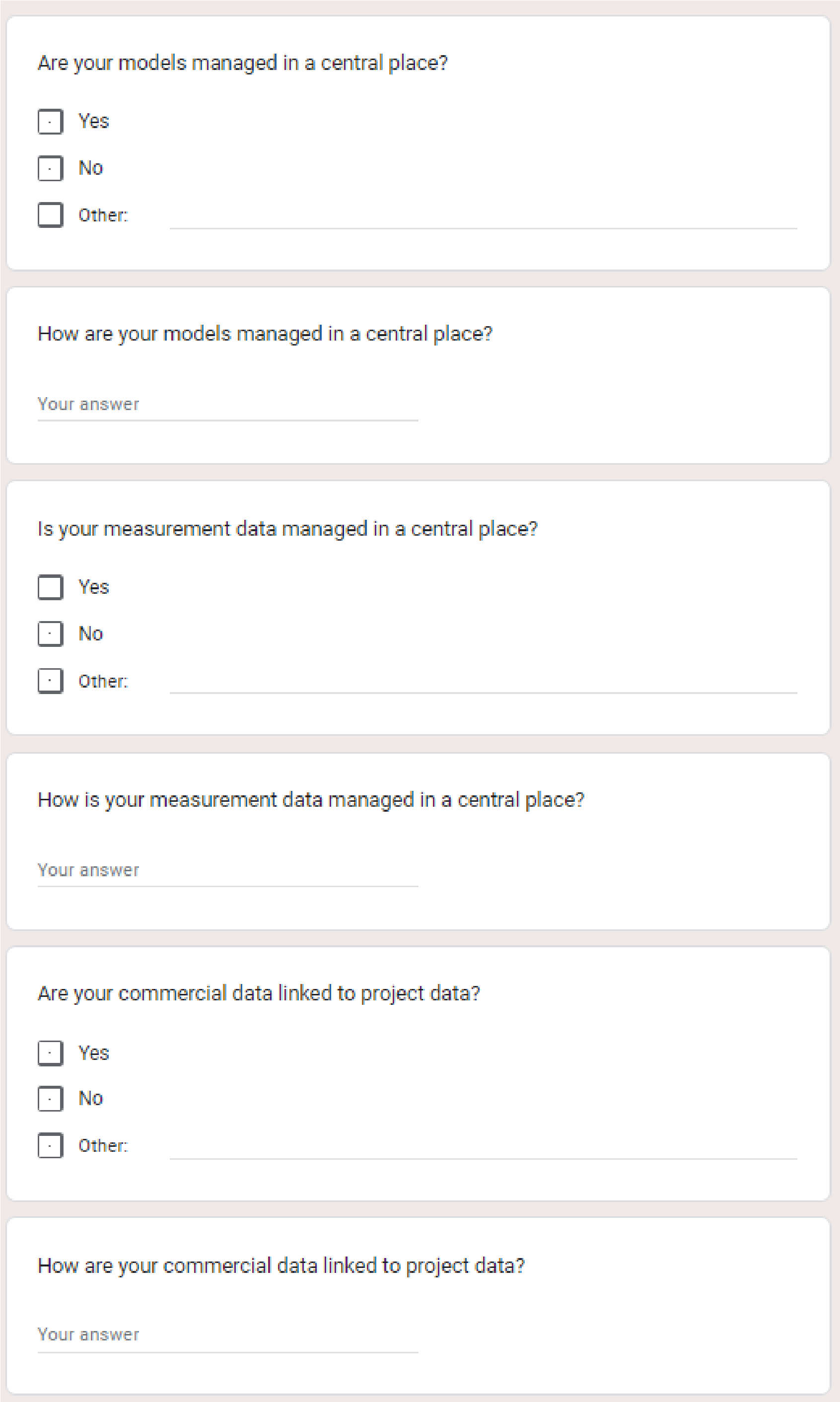
- [ ] Yes
- [ ] No
- [ ] Other: ____________

How are your product parameters managed in a central place?

Your answer ____________

Figure 52: Data Intelligence (III)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.



Are your models managed in a central place?

- [ ] Yes
- [ ] No
- [ ] Other:

How are your models managed in a central place?

Your answer

Is your measurement data managed in a central place?

- [ ] Yes
- [ ] No
- [ ] Other:

How is your measurement data managed in a central place?

Your answer

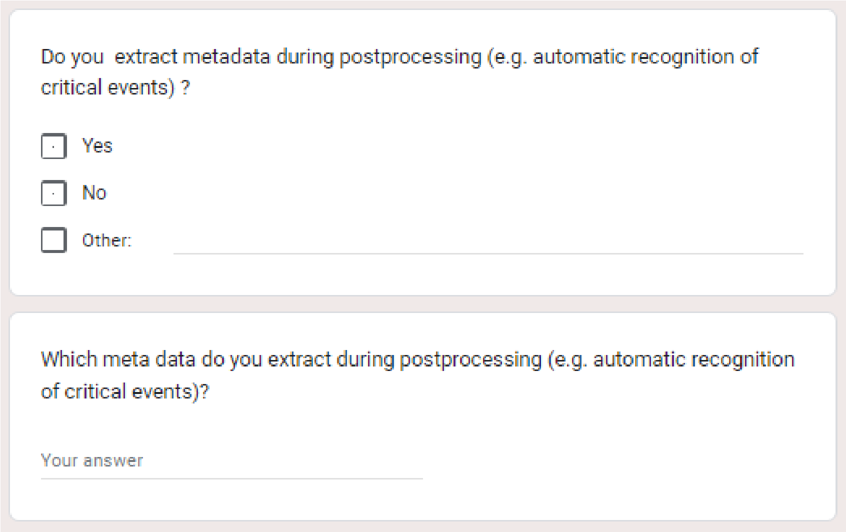
Are your commercial data linked to project data?

- [ ] Yes
- [ ] No
- [ ] Other:

How are your commercial data linked to project data?

Your answer

Figure 53: Data Intelligence (IV)



Do you extract metadata during postprocessing (e.g. automatic recognition of critical events) ?

- [ ] Yes
- [ ] No
- [ ] Other:

Which meta data do you extract during postprocessing (e.g. automatic recognition of critical events)?

Your answer

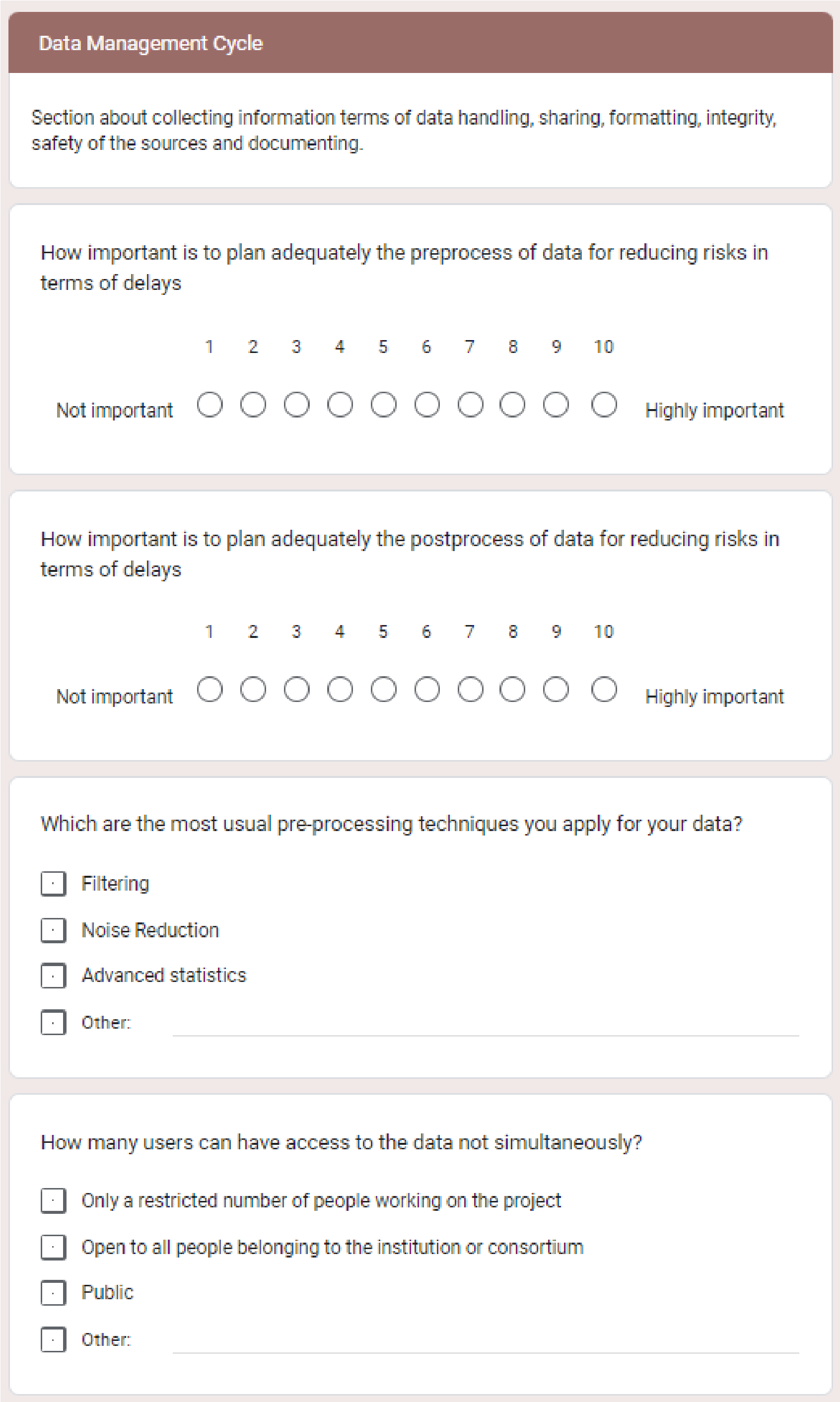
Figure 54: Data Intelligence (V)

Data Management Cycle

A total of 31 questions were asked, finally, in order to understand the mechanisms and protocols adopted by each institution during all the lifecycle of data management:

- ***Pre-process of the data***: planning of the pre-process, techniques, acceptable duration of this phase.
- ***Operations with the data***: number of users with simultaneous access to the data, tools, management of metadata and presence of standards in the domain, tools for sharing the data among user working with the same datasets in the same project.
- ***Data Quality***: accuracy versus computational time, sources, checks about accuracy and robustness, redundancy of measurements and track of the versioning.
- **Storage and data amount**: data formats, size of the datasets, number of sampling channels during simulations/physical activities, frequency of sampling and length of signals.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.



Data Management Cycle

Section about collecting information terms of data handling, sharing, formatting, integrity, safety of the sources and documenting.

How important is to plan adequately the preprocess of data for reducing risks in terms of delays

	1	2	3	4	5	6	7	8	9	10	
Not important	○	○	○	○	○	○	○	○	○	○	Highly important

How important is to plan adequately the postprocess of data for reducing risks in terms of delays

	1	2	3	4	5	6	7	8	9	10	
Not important	○	○	○	○	○	○	○	○	○	○	Highly important

Which are the most usual pre-processing techniques you apply for your data?

- [ ] Filtering
- [ ] Noise Reduction
- [ ] Advanced statistics
- [ ] Other: ______

How many users can have access to the data not simultaneously?

- [ ] Only a restricted number of people working on the project
- [ ] Open to all people belonging to the institution or consortium
- [ ] Public
- [ ] Other: ______

Figure 55: Data Management Cycle questions (I)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

How many users can have access to the data in parallel at the same time?

- [ ] Only 1 person per time
- [ ] A restricted number
- [ ] As many as it is needed
- [ ] Other: ______

Do you use any tool to manage metadata

- [ ] Yes
- [ ] No
- [ ] Other: ______

Which tools do you use to manage metadata?

Your answer ______

Are data standards available in your domain?

- [ ] Yes
- [ ] No
- [ ] Don't know

Are you following any standard for data formatting?

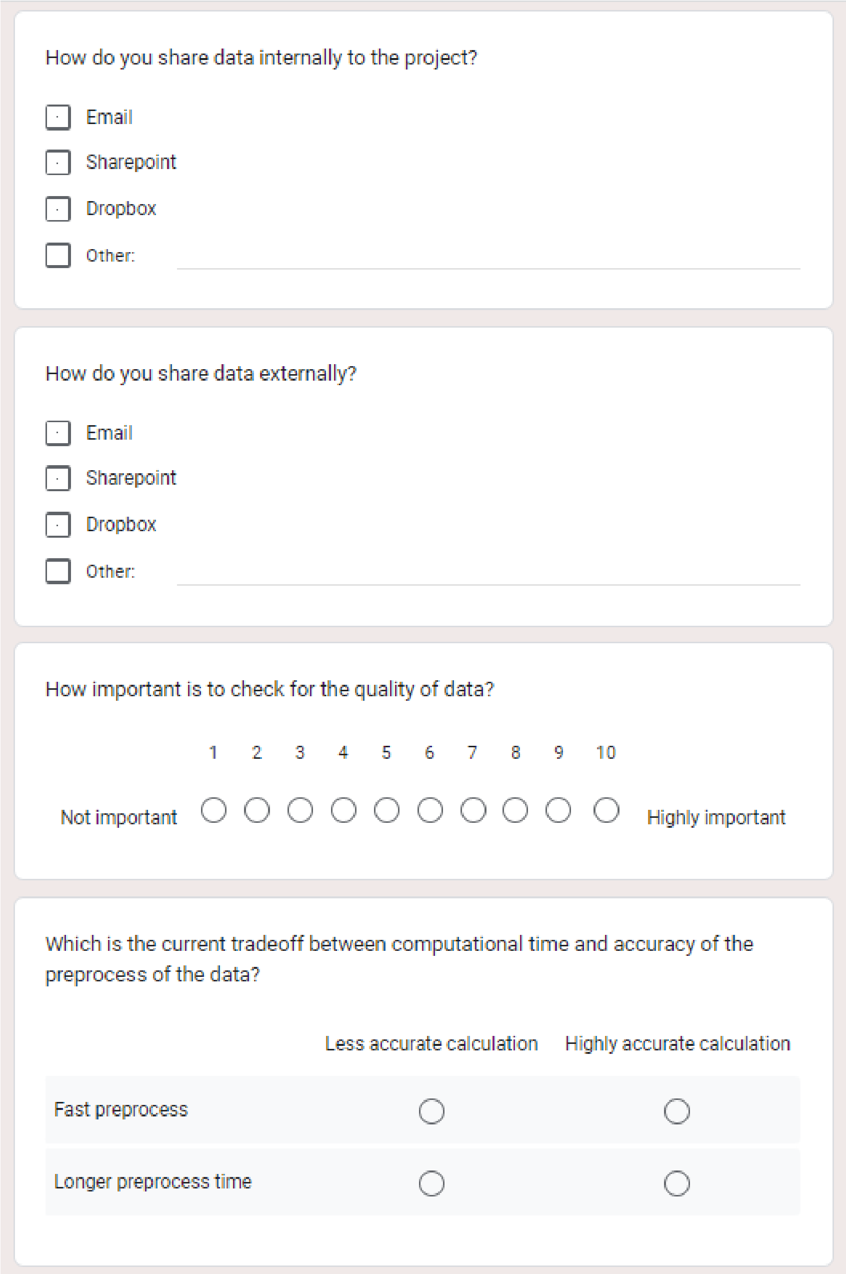
- ○ Yes
- ○ No
- ○ Other: ______

Which data standard are you using for formatting your data?

Your answer ______

Figure 56: Data Management Cycle questions (II)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.



How do you share data internally to the project?

- [ ] Email
- [ ] Sharepoint
- [ ] Dropbox
- [ ] Other: ____

How do you share data externally?

- [ ] Email
- [ ] Sharepoint
- [ ] Dropbox
- [ ] Other: ____

How important is to check for the quality of data?

	1	2	3	4	5	6	7	8	9	10	
Not important	○	○	○	○	○	○	○	○	○	○	Highly important

Which is the current tradeoff between computational time and accuracy of the preprocess of the data?

	Less accurate calculation	Highly accurate calculation
Fast preprocess	○	○
Longer preprocess time	○	○

Figure 57: Data Management Cycle questions (III)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

Do you use data from several sources

- [ ] Yes
- [ ] No
- [ ] Other: ______

Which are the sources of data you are using?

- [ ] Real time simulation
- [ ] Data stored locally in databases
- [ ] Data stored in a cloud, external database, public source of data, etc...
- [ ] Real time measure
- [ ] Other: ______

Do you make a real time check of the consistency/robustness of the data

- [ ] Yes
- [ ] No
- [ ] Other: ______

Which tools do you use to check consistency and robusteness of the data?

- [ ] In-house software
- [ ] A commercial software
- [ ] No check on the consistency of input data?
- [ ] Other: ______

Do you apply redundancy in data acquisition?

- [ ] Yes
- [ ] No
- [ ] Other: ______

Figure 58: Data Management Cycle questions (IV)

How do you create data redundancy?

- [ ] Additional physical sensing
- [ ] Virtual sensing and numerical models
- [ ] No redundancy applied
- [ ] Other:

Do you use any tool for versioning the data?

- [ ] Yes
- [ ] No
- [ ] Other:

Which tools do you use for versioning the data?

- [ ] None
- [ ] Gitlab or similar for not binary files
- [ ] Jisapp or similar for binary files
- [ ] Other:

Is your data organised in a way that the most recent/released data can be accessed easily?

- [ ] Yes
- [ ] No
- [ ] Other:

Where is the data stored?

- [ ] Network access
- [ ] Locally
- [ ] Other:

Figure 59: Data Management Cycle questions (V)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

For each of the following examples of data, select the format(s) you have adopted in your domain

	Scalar	Array	Multi-dimensional array	Metadata structure	Not used in my domain
Time series	☐	☐	☐	☐	☐
Physical characteristics of the asset	☐	☐	☐	☐	☐
Sample logs	☐	☐	☐	☐	☐
Measurements	☐	☐	☐	☐	☐
Simulation results	☐	☐	☐	☐	☐
Requirement specifications	☐	☐	☐	☐	☐
Constraints	☐	☐	☐	☐	☐

Do you save

- ☐ Data junks
- ☐ One file per test
- ☐ Other: ________

Which is the amount of data (input) you require for a simulation/hybrid experimental sample?

- ☐ Less than 1 Mb
- ☐ Between 1 Mb and 100 Mb
- ☐ Greater than 100Mb
- ☐ Other: ________

Figure 60: Data Management Cycle questions (VI)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

Which is the amount of data (output) you produce in a simulation/hybrid experimental sample?

- [ ] Less than 1 Mb
- [ ] Between 1 Mb and 100 Mb
- [ ] Greater than 100Mb
- [ ] Other: ____

How many channels are in your logging system?

- [ ] Less or equal to 8
- [ ] Between 9 and 16
- [ ] Between 17 and 32
- [ ] Greater than 32
- [ ] Other: ____

In case you record time series, how long the time series are?

- [ ] Less or equal to 5 minutes
- [ ] Between 5 minutes and 30 minutes
- [ ] Greater than 30 minutes
- [ ] Not measuring time series
- [ ] Other: ____

Which is the sampling frequency of your time series?

- [ ] Less or equal to 1 Hz
- [ ] Between 1 Hz and 5 Hz
- [ ] Between 5 Hz and 10 Hz
- [ ] Greater than 10 Hz
- [ ] Not measuring time series
- [ ] Other: ____

Figure 61: Data Management Cycle questions (VII)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006927.

Annex 2: List of participants

A total of 11 answers were collected. Table 2 collects the main information of the respondents to the survey, derived from the registry section of the questionnaire.

Table 2: List of participants

Name	Institution	Type of institution	Main technical role of your institution in the VALID project (in case of widespread contribution, select which is most affected by the data management)
Patxi Etxaniz	IDOM	Wave energy developer	Technology developer
Günter Lang	AVL List GmbH	Other private institution	Development / customizing of VHTP
João Cruz	Yavin Four Consultants	Consultancy	Other (please specify)
George Lavidas	Delft University of Technology	University/Academia	Numerical modeller
Natalia Rojas	Aquatera Atlántico	Consultancy	Lead of Task 2.2 Model requirements for the VALID Hybrid Test Platform
Jon Lekube	BiMEP	Other public institution	Manager/owner of a physical test rig
Timur Delahaye	CorPower Ocean	Wave energy developer	Manager/owner of a physical test rig
Eider Robles	TECNALIA	Research centre	Manager/owner of a physical test rig
Claes Eskilsson	Aalborg University	University/Academia	Numerical modeller
Troels Lukassen	Wavepiston	Wave energy developer	Numerical modeller
Pär Johannesson	RISE	Research centre	Researcher in mechanical reliability