H2020-EINFRA-2017
EINFRA-21-2017 - Platform-driven e-infrastructure innovation
DARE [777413] "Delivering Agile Research Excellence on European e-Infrastructures"

D6.3 Pilot Tools and Services, Execution and Evaluation Report I

Project Reference No	777413 — DARE — H2020-EINFRA-2017 / EINFRA-21-2017
Deliverable	D6.3 Pilot Tools and Services, Execution and Evaluation Report I
Work package	WP6: EPOS Use Case
Tasks involved	T6.3 EPOS Pilot Development
Type	R: Document, report
Dissemination Level	PU = Public
Due Date	30/06/2019 Deadline extended to 31/07/2019 in agreement with the PO
Submission Date	31/07/2019
Status	Final Draft
Editor(s)	Federica Magnoni (INGV), Emanuele Casarotti (INGV)
Contributor(s)	Athanasios Davvetas, Iraklis Klampanos (NCSR-D)
Reviewer(s)	Antonis Lempesis (ATHENA RC)
Document description	This deliverable will report on the implementation and execution of the Rapid Assessment (RA) test case of the EPOS Use Case within the DARE platform, the set of activated DARE components and their

	configuration parameters. It will also report on the pilot development evaluation phase.

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
v1	14/06/2019	Initial Structure	F. Magnoni (INGV)
v2	28/06/2019	Sections on the test case structure and on the WP6 training	F. Magnoni (INGV)
v3	01/07/2019	Section on DARE API and suggestions	A. Davvetas (NCSRD)
		Update of the sections	F. Magnoni (INGV)
v4	03/07/2019	Completion of the WP6 training section and update of sections 2	F. Magnoni (INGV)
v5	04/07/2019	Comments on the document	E. Casarotti (INGV)
		Completion of sections 2	F. Magnoni (INGV)
v6	05/07/2019	Completion of sections 4 and 5	F. Magnoni (INGV)
		Changes in section 2.3 and comments on the document	E. Casarotti (INGV)
v7	08/07/2019	Comments and updates on the sections	A. Davvetas (NCSRD)
v8	12/07/2019	Document ready for internal review	F. Magnoni (INGV)
v9	22/07/2019	Comments from Reviewer received	A. Lempesis (ATHENA RC)
v10	29/07/2019	Final document	F. Magnoni (INGV)

Executive Summary

This deliverable reports on the present structure of the seismic ground motion Rapid Assessment (RA) test case within the EPOS seismological Use Case and its implementation in the first prototype of the DARE platform. The main DARE components, with corresponding configuration parameters, required for the implementation and execution, are also described. Finally, the deliverable reports on the training evaluation event organised by WP6 to present to a selected group of users the pilot development and execution using the DARE platform, and focuses on the gathered feedback and suggestions in order to address future works.

Table of Contents

List of Figures

List of Terms and Abbreviations

Abbreviation	Definition
RA	Rapid Assessment
SS	Seismic Sources
ES	Ensemble Simulations
EIDA	European Integrated Data Archive
FDSN	Federation of Digital Seismographic Network
PE	Processing Element
CWL	Common Workflow Language
RI	Research Infrastructure
MPI	Message Passing Interface

1 Introduction

1.1 Purpose and Scope

The objective of this deliverable is to present the structure of the **ground motion Rapid Assessment (RA) test case**, prioritised in the framework of the EPOS seismological Use Case, and the stages of its implementation and execution in the first prototype of the DARE platform. The RA workflow is used as a first basic test case to address the platform structure and functionalities. A description of the required DARE components is also presented in order to highlight the main advantages of using the DARE platform to ease development of scientific applications and exploitation of required RI services.

The deliverable has also the scope of reporting on the pilot evaluation training organised by WP6 in order to obtain feedback from the target users, mainly research developers, on the usefulness of the platform for their present work and the future developments they are planning, and on possible updates required in the second phase of the project.

1.2 Approach and Relation to other Work Packages and Deliverables

The approach used will be to firstly describe the high level steps of the RA workflow and their implementation and execution in the DARE platform. Then the WP6 training evaluation is presented focusing on the gathered user feedback that will address future works planned in the general context of the EPOS Use Case implementation. This deliverable is closely linked to WP8, responsible for training and dissemination activities, and in particular deliverable D8.4 (Training and Consulting report I) that describes in detail the organisation of the training event, its purposes and the chosen participants. It is also linked to WP2 that works at the architectural design, WP3 for lineage and process management, and WP4 responsible for the platform development.

1.3 Methodology and Structure of the Deliverable

The structure of this deliverable is as follows. First, the steps of the underlying workflow of the RA test case are described specifying how they would have to be performed without DARE compared to the advantages of exploiting the platform. Then the specific DARE components used to implement the workflow in the DARE platform are presented followed by the description of the workflow execution through a Jupyter Notebook. Finally, the organisation of the pilot evaluation event is summarised and the feedback from the target users are analysed in detail to qualitatively and quantitatively assess the present achievements and the required future updates.

2 EPOS Use Case First Prototype Implementation

This section will summarise the high level structure of the RA workflow and its implementation and execution through the first prototype of the DARE platform. An overview of the main required DARE components is also given.

2.1 Test Case Description

In the framework of the EPOS seismological Use Case, in this first 18 months of the DARE project we prioritised the strong ground motion RA test case, as reported in deliverable D6.1. The scope of this test case is to analyse seismic wavefields right after the occurrence of large earthquakes and produce quick, on-demand estimates of ground motion parameters as peak values of ground velocity or acceleration. This allows us to construct maps of these parameters, that can be useful to communicate about the earthquake in the emergency context and above all can be compared with maps based on recorded ground motion data to understand how our models work and, in future applications, to combine observed and synthetic information improving our knowledge on ground behaviour. The RA workflow is used as an initial basic test case to address the platform development. Many steps of this workflow can be useful for other more complex seismological applications, as those represented by the other test cases foreseen in the EPOS Use Case, the Seismic Source (SS) analyses and the Ensemble Simulation (ES) analyses (see Figure 1 and deliverable D6.1). This will facilitate their future implementation exploiting the common steps already constructed for RA test case and benefiting from the modularity and flexibility of the built platform.

In order to allow and guide the implementation of the RA test case in the first DARE prototype, we identified the main high level steps that constitute the underlying workflow and that will be translated into a series of executable steps as described in section 2.3. Figure 1 summarises these steps, extensively described in deliverable D6.1:

- Users have to choose a model to describe the physical properties and geometry of the wave propagation medium, i.e. a wavespeed model and a mesh.
- A model of the seismic source that represents the earthquake should be also defined as well as a list of seismic stations to be used in the analyses.
- With the previous inputs a waveform simulation can be run to obtain the synthetic seismograms; in the implemented example the chosen numerical code is SPECFEM3D_Cartesian [*Peter et al.*, 2011], specific for local seismic waveform simulations.
- Corresponding raw observed seismograms for the chosen earthquake should be downloaded e.g. from public European archives, such as EIDA[1].
- Pre-processing is applied on both data and synthetics, that includes e.g. filtering the traces or removing the instrument response from the recorded data; this prepares the seismograms for the following analyses assuring the consistency between data and synthetics.
- Chosen ground motion parameters are calculated from both simulated and observed wavefields and comparisons between them are performed. In the simple example so far considered for RA, the determined ground motion parameters are peak values of displacement, velocity and acceleration of the ground and the damped spectral acceleration, then we just calculate the difference between their synthetic and recorded values.

[1] http://eida.rm.ingv.it

- The output products are finally stored along with their metadata and complete provenance.

Considering the listed steps, in the following we analyse how a workflow like RA is usually developed and executed without the DARE platform and the typical issues to be faced. This will allow us to highlight in section 2.3 the advantages of using the DARE platform based on the description of its main components and execution approach (see sections 2.2 and 2.3).

- Manual configuration, compilation and run of the numerical codes should be done by testing the suitable configuration and compilation parameters and handling all the specific requirements of the computing environment; despite the usage of well established community codes (as e.g. SPECFEM3D), this step usually consumes a large part of the working time often requiring trial-and-error procedures.
- In our specific RA test case, as in typical seismological applications, external services providing input data should be accessed (in this case public archives to gather recorded seismograms). This requires to deal with the main issue of interfacing with specific web service protocols and data format. Ad-hoc query procedures are every time manually written by the developers.
- Another issue is to manage transfer, analysis and storage of input and output data that can be potentially large and/or numerous. I/O optimization is then manually included in the used software and tools often clashing with the limits of available resources.
- Especially for more complex scientific applications, keeping trace of the performed runs and analyses is fundamental since it could be necessary to search and reuse them several times, combine multiple analyses and improve the procedures by tracking errors and their source. Typical seismological tools do not include a programmatic management of metadata and provenance thus users often resort to in-house strategies or more elegant but ad-hoc developers' solutions.
- Finally, the steps of the workflows have to be manually assembled e.g building a sequence of calling scripts exposing the procedures to the usual errors that can occur because of each manual intervention.

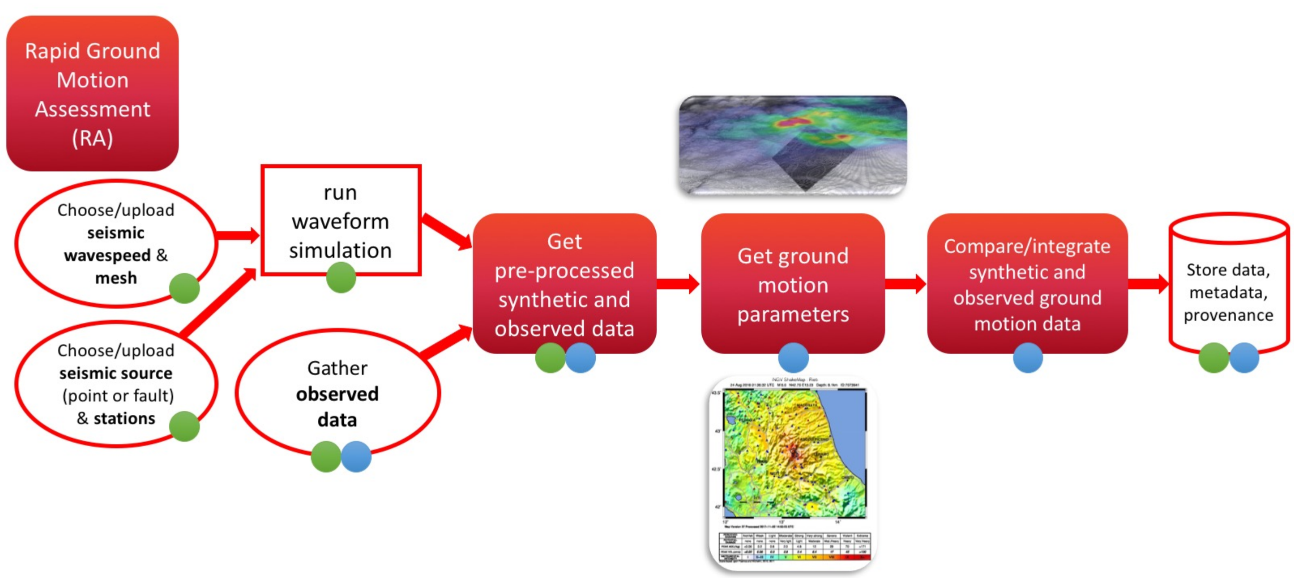


Figure 1: Steps of the workflow for the RA test case. The coloured dots at the corners of the boxes indicate the steps that are in common with the other proposed test cases of the general EPOS Use Case (compare with figures in the Annex of deliverable D6.1): green is for SS analysis test cases, blue for ES test case.

2.2 Overview of DARE API

The implementation and execution of the described RA test case within the DARE platform is based on the exploitation of the DARE API and its main components.

The DARE API is a grouping of exposed web APIs of individual components that are part of the DARE platform (see Figure 2). Web APIs allow the external use of DARE components by enabling machine-to-machine communication through the Hypertext Transfer Protocol, broadly known as HTTP. Therefore, the DARE API acts as a gateway between the cloud deployed DARE platform components and the User Interfaces that the user interacts with as part of his use case.

The DARE platform can be described as composition of containerised services. Containerised services are software components that are wrapped inside Docker containers. A docker container is a runtime instantiation of container image, with container image being a template of executable steps that ensure the availability of code and dependencies required for the operation of a single application/software component.

The advantages of such a composition can be summarised as:

- Effortless cloud infrastructure deployment
- Software isolation (enabled by containerisation)
- Scalable and flexible platform through the use of container orchestration systems such as Kubernetes[2]
- Across component communication using exposed REST APIs

The components used during the first iteration of executing the RA workflow on DARE platform are:

- Execution API
- Dispel4py Registry (also referenced as Processing Elements Library)
- sProv & sProv-viewer (both composing the Provenance component)

The **Execution API** enables the distributed and scalable execution of SPECFEM3D simulations and dispel4py workflows (e.g. used to describe all the other steps of RA except for the simulation, see section 2.3), which can be extended to other execution contexts. Through the Role-based access, control Execution API can itself access the Kubernetes API and spawn container clusters on-demand, while at the same time enabling shared file system access with itself and the execution contexts and monitoring the status of executed jobs. All execution contexts are built to use Common Workflow Language (CWL) which allows for Dynamically parameterised executions. In addition, the execution API offers services such as uploading/downloading and referencing of data and process monitoring. The data referencing as well as the uploading and downloading are not only available within the services of the DARE platform but also by using the EUDAT B2DROP service[3], already exploited for Climate Use Case and planned to be used in the next phase also for EPOS Use Case.

The **Dispel4py Registry** provides functionality for registering workflow entities, such as processing elements (PEs), functions and literals. Dispel4py Registry provides a workspace structure for registering workflow entities that enables user individuality and encourages sharing and collaboration via groups and workspaces. Workflow registration through the Dispel4py Registry allows transparency and provides a manageable way of allowing users to run arbitrary code through the execution API.

[2] https://kubernetes.io/

[3] https://eudat.eu/services/b2drop

The **sProv** records metadata regarding execution aspects such as: distribution of computation units, generated data products and runtime changes. sProv is natively supported by dispel4py and offers visualization of captured provenance through the sProv-viewer component.

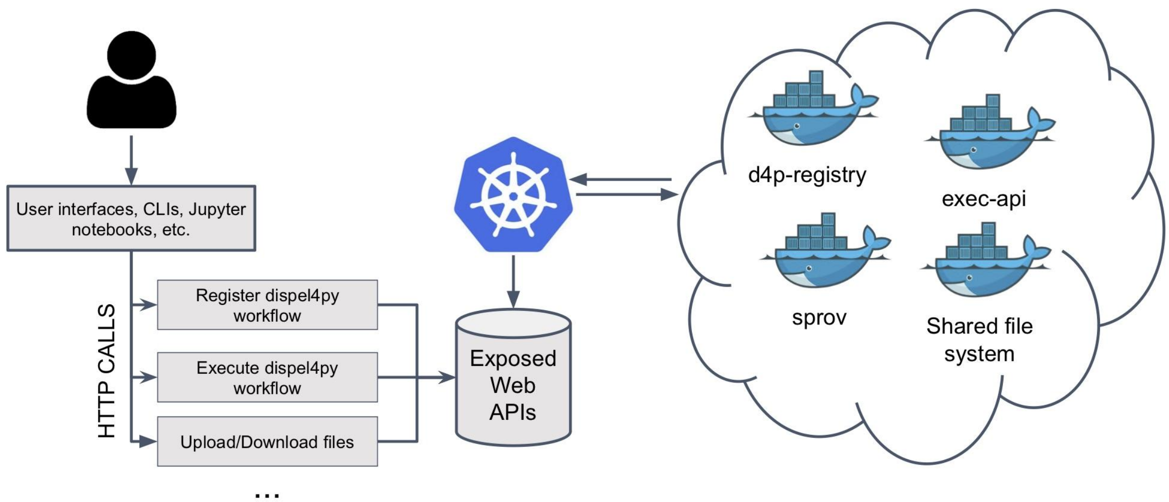


Figure 2: Overview of the DARE API

2.3 Implementation and Execution of RA Test Case

In the following we describe how the RA workflow has been implemented and executed within the first prototype of the DARE platform exploiting its components introduced in section 2.2.
Looking at Figure 1, the initial steps of selecting structure and source models are so far handled by asking users to prepare these input files and upload them in the computing environment. This will change in the future and the advantage of DARE is that the platform is in principle ready to be linked to whatever external graphical interface that allows to make these selections (e.g. the VERCE portal).
Then, for the simulation step, the dockerized version of the numerical code (here SPECFEM3D), containing all the dependencies required for execution, is used. Thus, through a specific API call, an MPI cluster is created in Kubernetes, it runs the code and is destroyed at the end.
Finally, the other steps of the RA test case are described by dispel4py PEs and each of them is considered as a separate executable workflow launched through another specific API call. Each call now starts a dispel4py MPI cluster in Kubernetes, the corresponding PE is executed there producing the output files, and at the end the cluster is destroyed.
The interface used to execute the RA workflow within DARE platform is so far a Jupyter Notebook[4] using the DARE API under the hood, also shown at the training demo. The steps are the following:

1. Register to the dispel4py registry and get a token for following authorization:

```
# Constant hostnames of exec-api and d4p-registry api
EXEC_API_HOSTNAME = 'https://testbed.project-dare.eu/exec-api'
D4P_REGISTRY_HOSTNAME = 'https://testbed.project-dare.eu/d4p-registry'
# token
auth_token = F.login(USERNAME, PASSWORD, D4P_REGISTRY_HOSTNAME)
```

[4] https://gitlab.com/project-dare/dare-api/blob/master/examples/wp6/WP6%20Rapid%20assessment.ipynb

2. Create a DARE workspace:

```
workspace_url, workspace_id = F.create_workspace("", [wp_name], "", creds)
```

here and in the following brackets '[]' contain names of variables

3. API call to launch a SPECFEM3D simulation on the DARE cluster (using `n_nodes` nodes):

```
F.submit_specfem(n_nodes,data_url=[zip_input_file],
                 token=F.auth(), creds=creds)
```

`zip_input_file` contains the input parameter files of the code and is uploaded using its url.

4. Upload required input json files:

```
F.upload(token=F.auth(),
         path=[file_path], local_path=[file_name],
         creds=creds)
```

Then, each of the other steps of RA workflow, represented by a dispel4py PE, is launched by similar API calls and the structure is always the same:

- Create a signature to identify the specific PE/workflow:

```
pe_url = F.create_pe(desc="", name=[pe_name], conn=[],
      pckg=[wp_name],workspace=workspace_url, clone="", peimpls=[],
      creds=creds)
```

- Register the PE to the dispel4py PE registry:

```
impl_id = F.create_peimpl(desc="",
      code=open([pe_file]).read(),
      parent_sig=pe_url, pckg=[wp_name],name=[pe_name],
      workspace=workspace_url, clone="", creds=creds)
```

- dispel4py API call to execute the PE:

```
F.submit_d4p(impl_id=impl_id,pckg=[wp_name],
             workspace_id=workspace_id,pe_name=[pe_name],
             token=F.auth(), creds=creds,
             n_nodes,n_processes, iterations,
             reqs=[requirements], target=[d4p_model],
             inputdata=[input_files])
```

Using this structure, we can execute the following steps of the RA workflow:

5. Creation of the data download parameter file starting from the step #3 input files

6. Download recorded seismograms querying FDSN web services

7. Pre-process recorded and synthetic seismograms

8. Calculate ground motion parameters and compare their values from recorded and synthetic data

9. Plot a final map of the results

After every call, we can monitor the status of the job and see and download the output, as introduced in section 2.2:

- Monitor the call:

```
F.monitor(creds=creds)
```

- See output directories/files:

```
resp = F.myfiles(token=F.auth(), creds=creds)
resp = F._list(path=[API_LOCAL_PATH], creds=creds)
```

- Download files:

```
F.download(path=[file_path], creds=creds,
           local_path=[local_download_path])
```

All the dispel4py scripts include the instructions required to track and store metadata and complete provenance, and users can fully customise them. Moreover, an identification number (RUN_ID) is assigned at run time (generated from the execution API) at each job and can be passed to the following job in order to allow for future exploration of the lineage.
The presented structure highlights the flexibility of the DARE platform since it is easy to customise a workflow: a new PE is written, tested and then implemented in the general workflow using a new API call with always the same form.
A single step can be also run by itself (e.g. the pre-processing, if raw data and synthetics are already available), an evidence of the modularity of the structure that we constructed, which allows to easily reassemble the workflow and adapt it to a new application.
This basic example is particularly suitable to test workflow implementation and execution in the DARE platform. Of course, in the future more complex examples of RA workflow can be implemented as e.g. directly gathering from public archives maps of ground motion parameters based on observed data (like Shakemaps[5]) and comparing or integrating them with synthetic information, as initially proposed in D6.1. However, thanks to the modularity and flexibility of the platform this wouldn't change the implementation approach because in principle it is just a matter of making some changes in the implemented steps or adding new ones following the procedure described above. The same for the more complex test cases SS and ES whose implementation, planned for the next phase of the project, will also benefit from some common steps already implemented for RA.

So now the main advantages that we see in exploiting the DARE platform, for this test case but also for a more general scientific application, are:

- Possibility of exploiting the Cloud for execution, this means elasticity in acquiring and using resources and also on demand availability of computing and storage resources.

[5] http://shakemap.rm.ingv.it/shake/archive/

- Transparent set up and execution of runs without need to deal with environment specificity and details of code/scripts execution. Here a single call is used to do all the required steps to prepare the environment and run a SPECFEM3D simulation (see section 2.1).
- Possibility of exploiting RI services by including them in the whole workflow procedure, so taking care of the required input, query parameters and gathered output. Here a simple call allows to query FDSN web services of European archives to download recorded waveforms.
- Rapid and transparent data analyses and transfer between co-working environments.
- Automatic description and storage of complete lineage and multiple metadata that allow us to track runs and data through the whole workflow, to easily search and reuse them and to also combine numerous outputs from multiple workflows that is fundamental in many scientific applications.
- High level description of workflow steps that are as abstract as possible to increase the flexibility in reusing them to assemble different workflows.
- Existence of managed knowledge-bases (e.g. the PE registry) that allows users to easily exchange information about what they deployed and executed.
- Workflow structure and provenance information that can be customised.

Despite the highlighted advantages, there are still issues to be faced and updates to be made to the platform. These aspects emerged also from the training event described in section 3 and will be discussed in detail in section 4.

3 Training Event and Evaluation Results

This section will summarise the pilot training event organised by WP6 and focus on the gathered results in order to address the platform development and update in the second part of the DARE project. More details on the training organisation are reported in deliverable D8.4.

3.1 Training Purposes and Development

The WP6 training event was organized by INGV on June 27th, 2019. The agenda of the training[6] was structured in order to:

- give the participants a general overview of the project and its main goals;
- present the scientific problem behind the RA test case and the main advantages in using the DARE platform to reduce the difficulties that researchers usually face developing workflows;
- describe the main components that drive the DARE platform functioning (API, dispel4py, provenance);
- show the RA workflow execution using the DARE platform through a live demo;
- gather participants' feedback through anonymous questionnaires.

The trainers included two members of WP6 from INGV to introduce the test case and perform the demo and they were physically present at INGV, while the other trainers were connected remotely: two of

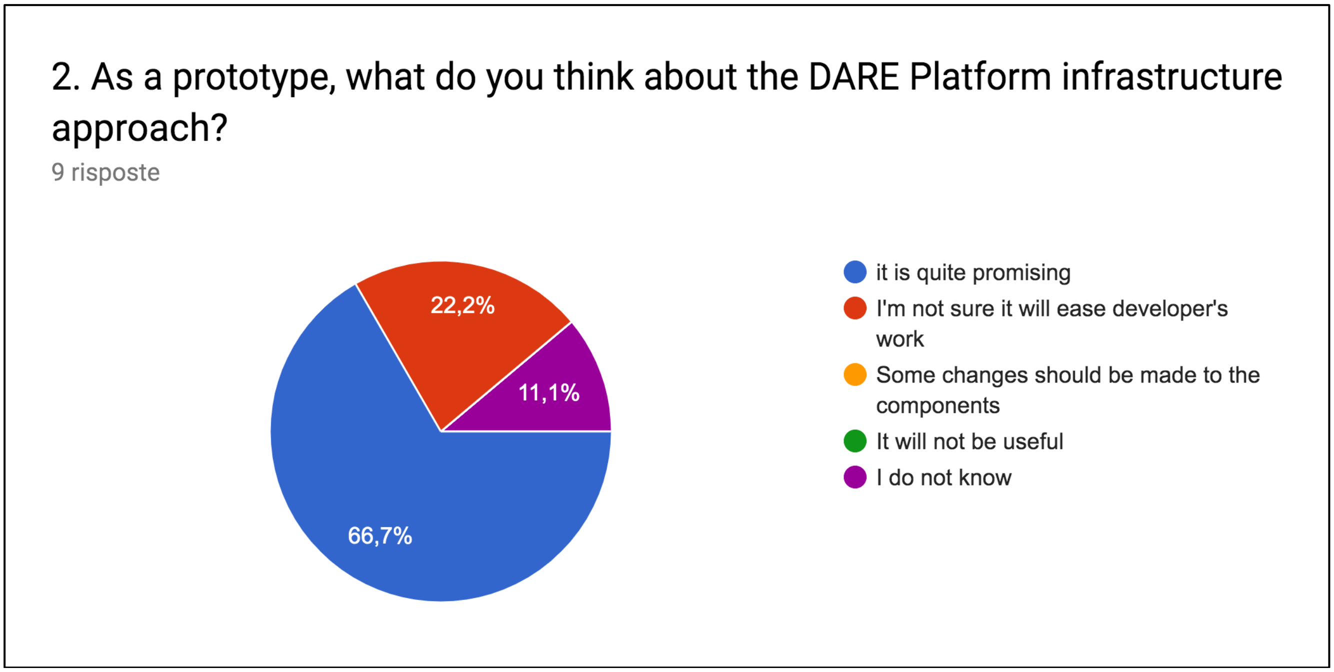
[6] https://docs.google.com/document/d/1UocmWmNWOZi8vNl5SppFVjn6KN9wF9ReqYCe_V_QWYo/edit

WP4 from NCSRD for the introduction of the DARE project and components, one member of WP2 from UEDIN to present dispel4py and one of WP3 from KNMI to present the main concepts of provenance. This multiplicity of figures was thought to offer the participants different approaches and points of view in order to appreciate the multidisciplinarity of the project.
The participants were carefully selected in order to gather useful feedback on the work that has been done in this first part of the project and support the next development. Their expertise is heterogeneous (as emerged from question #1 of the evaluation questionnaire, section 3.2) but basically all of them have experience in developing and implementing workflows. The total number of attendees was 9 and they all completed the evaluation questionnaire; the list and details of the participants can be found in deliverable D8.4.

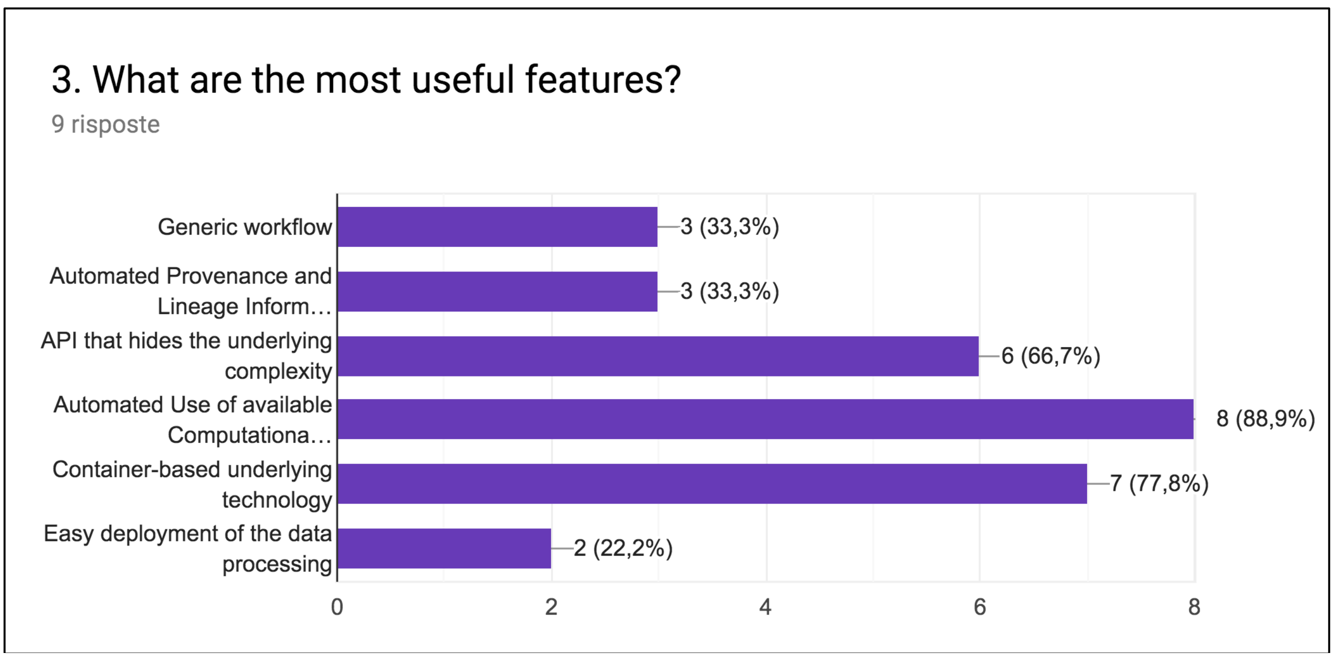
3.2 Evaluation Results

The evaluation questionnaire, anonymously filled in by the attendees, is available at this link:
https://docs.google.com/forms/d/e/1FAIpQLSeyBL6vVeyHlv2yBghodiUqi-ESOaSPlOdBtslnHLrkCC8QaA/viewform
In the following we present the results of the evaluation discussing the questions one by one. Then general comments are inferred.

In this question we gave the users the possibility to define themselves choosing the level of expertise for each role. This was done on purpose to highlight the heterogeneous expertise typical of the seismological research framework of INGV. Indeed, our research developers have inevitably a mix of technical and domain specific competencies. This is clearly a benefit in order to appreciate the operational advantages of DARE for the proposed applications and to suggest useful updates. More details in deliverable D8.4.



The majority of attendees thinks that the DARE platform approach is quite promising. Two of the attendees (22,2%), moreover, suggest that the functionality of the platform that will ease developers' work could be improved in the next releases.



All the proposed features were recognised of some usefulness with a particular focus on:

- Automated use of the available Computational Resources with Kubernetes
- Container-based underlying technology
- API that hides the underlying complexity

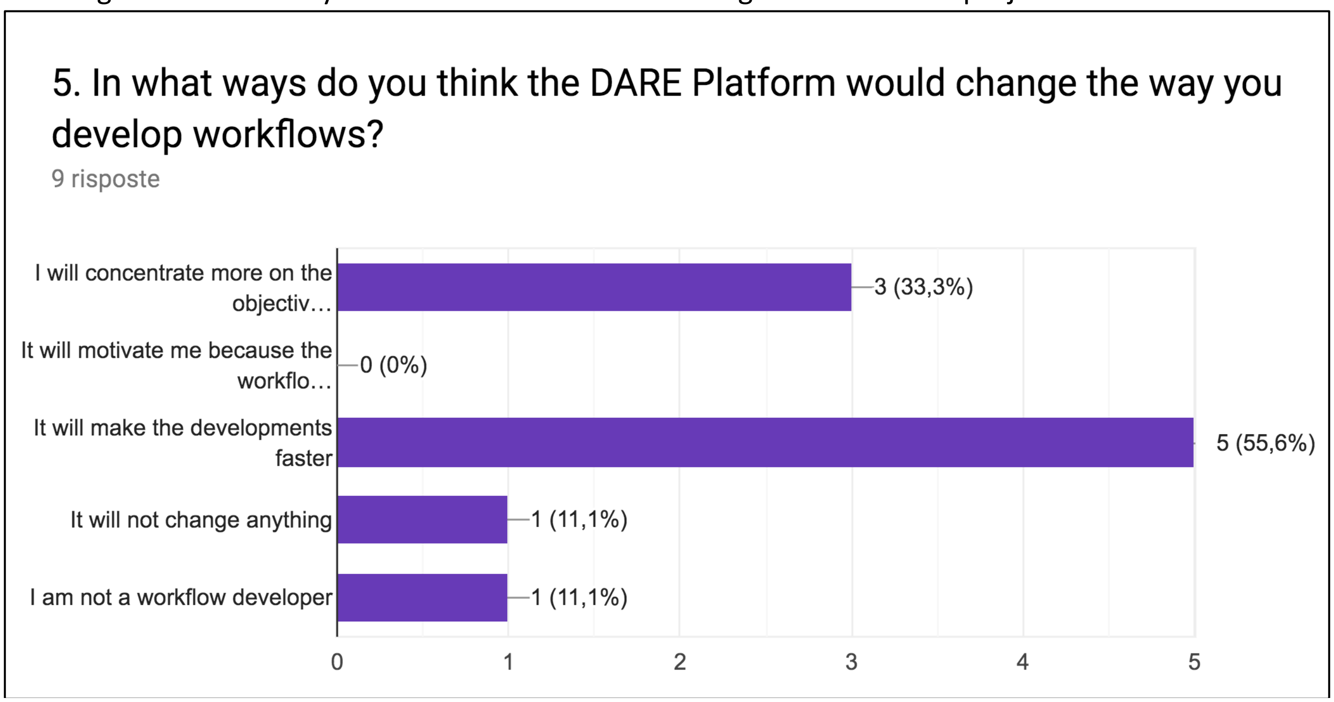
This highlights how the container-base technology and management of computational tasks of the workflow are extremely important in Earth Science field as well as the need of avoiding all the technical complexities, in order to concentrate on the main research objectives.

4. Which other features would you like to be available?

5 risposte

- I do not know
- I don't know all the features available at the moment, so it is hard for me to answer to this question.
- potential use tests ;
- more work on workflow management
- Event queue, event driven workflow, user defined containers

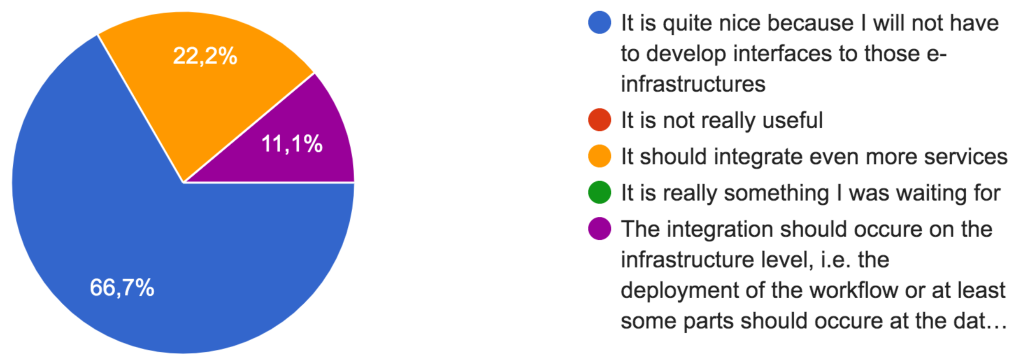
The last three answers emphasise that another important requirement of the Earth science community is the possibility to customise the platform. Research developers aim to easily access and modify the available components and, in view of their complementary domain specific expertise, would like to have a final product effectively usable for their practical and operational scientific applications. Allowing flexible and easy customisation is a crucial final goal of the DARE project.



Here attendees could give multiple answers and particularly they recognised that the platform will ease research developers' work making the developments faster and leaving them focused on their workflows' objectives, as already emerged from answers to question #3. These are some of the main goals of DARE and attendees' feedback demonstrates that the first platform prototype already gained promising achievements.

6. What do you think about the integration with available e-infrastructure Services (EUDAT, EIDA, ...)?

9 risposte

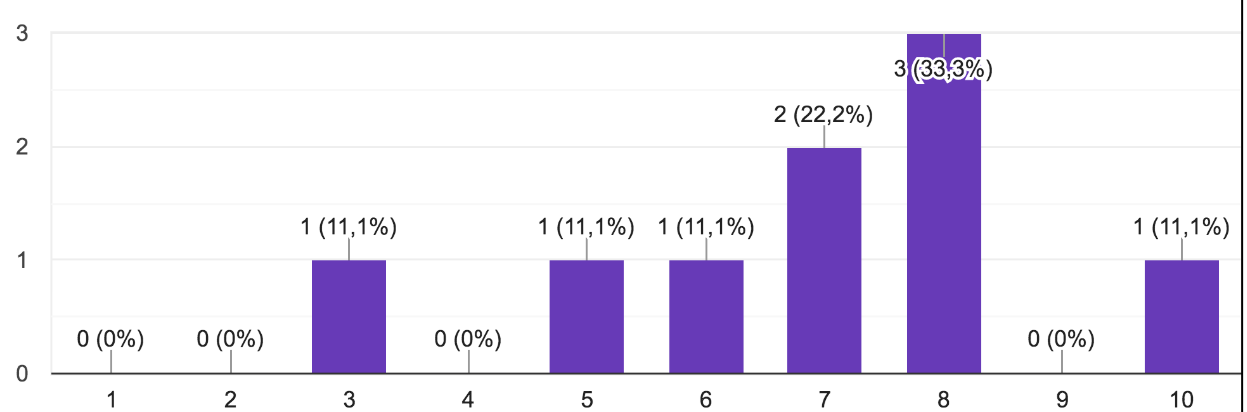


The majority of attendees clearly recognised how much DARE eases the interaction and integration with available e-infrastructure services and two of them are very interested in having even other services integrated. This is one of the main goals of DARE and gives a positive prompt for next developments.

Regarding the only open answer (*"The integration should occur on the infrastructure level, i.e. the deployment of the workflow or at least some parts should occur at the data centres"*), integration can be done at every level and we could lose flexibility if DARE constrains it only at *infrastructure level*. In any case DARE aims at high level integration.

7. For your knowledge, is the DARE Platform duplicating similar efforts or it is unique and interesting?

9 risposte



For the answers, '1'='It duplicates completely existing solutions', '10'='It is unique and interesting'.
The majority of the attendees recognised that the platform is quite unique and innovative (66.6% gave 7/10 or more). This demonstrates that the potentialities of the first prototype are evident. Then, the

work in the second phase of the project should be oriented to exploit them in order to create a unique and essential tool for research developers in Earth science.

8. If you assigned 5 or less at the previous question, can you list the projects with similar efforts?

2 risposte

I could be wrong, but Azure, IBM, AWS, etc, do not have similar services?

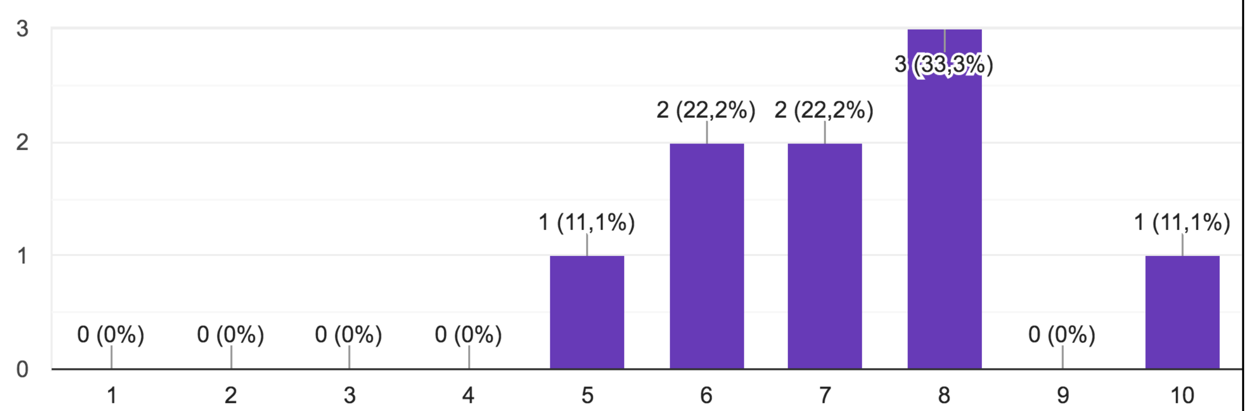
Storm, Flink, Spark, Singularity

The aforementioned cloud infrastructure providers offer automated resource management and expose some functionality through workflow services. However, these simple workflow services provide basic functionality regarding their respective components. DARE allows the creation, management and execution of complex workflows and additionally offers high level workflow description, as well as, recording provenance and data lineage.

The aforementioned software components (Storm, Flink, etc.) are completely orthogonal to the DARE platform. These software components can be integrated as execution contexts within the DARE platform.

9. What is the usefulness of the DARE Platform based on this first Prototype?

9 risposte

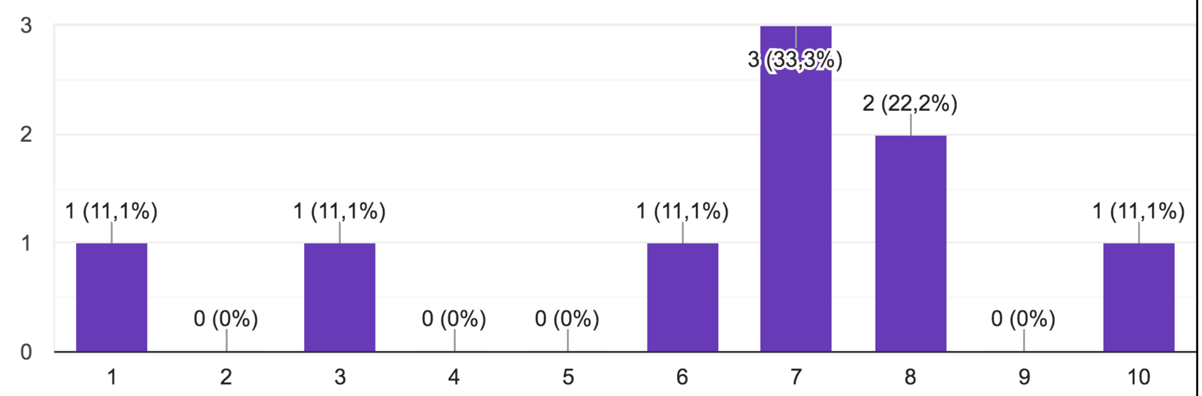


For the answers, ‘1’=’Not useful’, ‘10’=’It will be very useful’.

Almost all the attendees recognised the usefulness of the first DARE prototype both from the technical and domain specific point of view. Given the multidisciplinary role of the participants, this is a very positive feedback.

10. How likely you consider using the DARE Platform when it will be operational?

9 risposte



For the answers, ‘1’=’Not likely’, ‘10’=’Very likely’.

The answers reflect those of the previous questions, highlighting that the good impression on the first prototype generated good expectations on the next developments. Earth science community is really looking for a tool that can change and improve the way of working and this is exactly one of the main goals of DARE. A more advanced release of the platform, showing the applicability on more complex test cases, will then help to gain consensus also from researcher groups not yet interested in the platform usage.

11. Please add your comments and suggestions. This is important to have as much information as possible to report!

5 risposte

I suggest to show a comparison between the time spent to face one (or more) typical problem from scratch and by using the DARE platform

Regarding question #10, the use I'll do of the DARE platform will depend on how difficult will be to add new seismological analysis. Regards question #12, I'm a seismologist so some talks were hard to follow for a not authorised personnel.

The users need to know how to use/introduce their problems/computations in this platform.
Concerning the evaluation demo I believe that it should be interned for more dedicated people.

In order to achieve a full and real integration of DARE products with other RIs like EPOS, Web APIs should be improved and also some work on WF engine would be required.
The driving idea should be make it a sort of black box that enable easy communication through APIs, and guarantees deployment and (in case) elastic allocation of resources

I believe an overarching workflow concept, which would allow to define and execute complete workflow would be necessary (d4p further up??). Deriving this workflows from already existing execution graphs (provenance record) would be great.
Probably, also some messaging infrastructure for events, to trigger next steps and monitor progress, manage exceptions, etc. would be a helpful feature.
The dyspel data workflows might be quite expressive, though their definitions seems to be extremely verbose especially for simpler cases. Probably simplified definition, e.g. implicit graph definition trough calls, concept of sets, etc. should be possible.
The API and the deployment management look very interesting, it would gain further potential once it allows to deployment of user defined containers and some handles of storage/compute resource integration and locality.

Answer #1:

This comment highlights the need of a practical evidence of the platform usefulness in easing and making developers' work faster. This is definitely something that we have to show and plan for the next phase including for example the production of some material, also useful in the dissemination context, showing how to use and take advantage of the platform.

Answer #2:

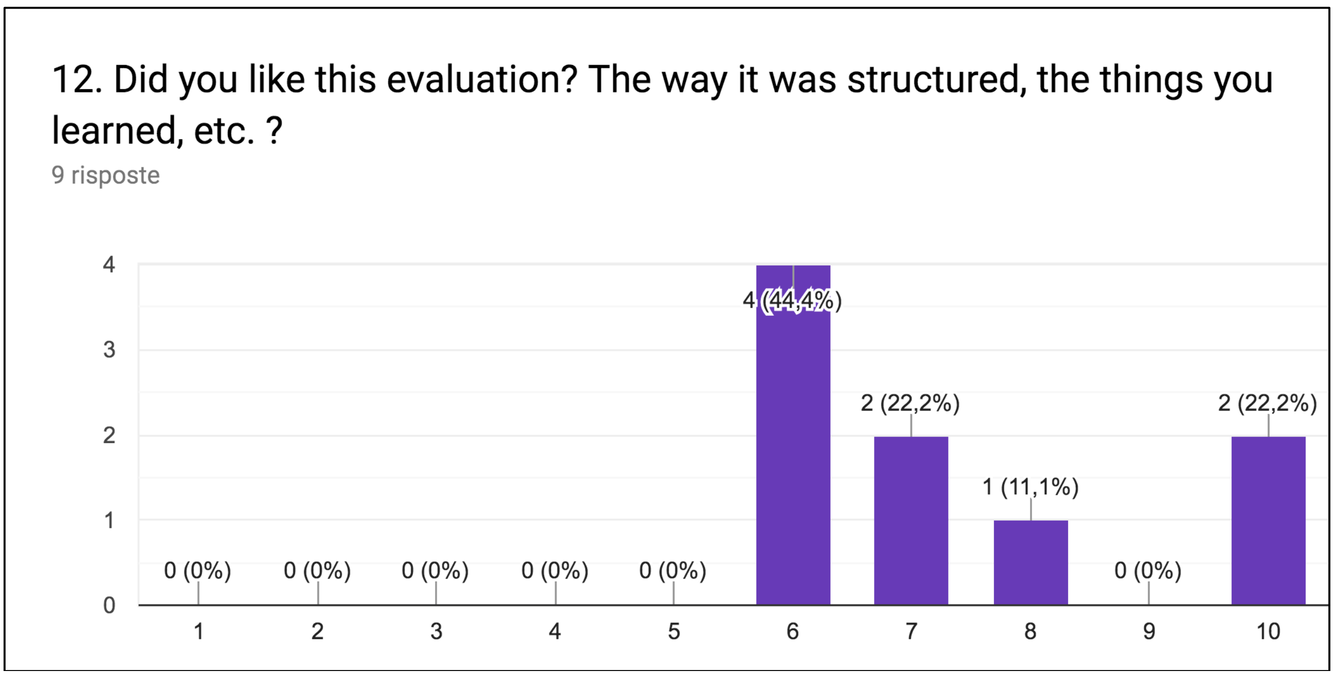
The answer evidences once again the importance of creating a flexible and customisable product that can be in principle applied to many different scientific test cases and research fields. Although this feature already emerged from the first prototype, we are encouraged to potentiate this aspect and in case show in a future training a practical example of introducing changes on an initial workflow.

Answer #3:

This answer perfectly reflects one of the main goals for the next phase of the project and reassures on the fact that the attendee has very clearly in mind the objectives and potentialities of the project.

Answer #4:

This comment contains useful suggestions for the next developments. Indeed, it supports one of the main objectives of the next phase of the project that is working on the optimisation in order to obtain a higher level of abstraction. Moreover, managing messages to trigger next steps, monitor progress and control errors is something we have to improve as well as an easier handling of computing and storage resources. The attendee recognises the potentialities of the platform and stresses again on the customisation e.g. looking for user defined containers.



For the answers, '1'='No', '10'='Absolutely yes'.

All participants appreciated the training evaluation so we can infer that its structure and the chosen audience (detailed in D8.4) were quite appropriate. It is evident that attendees with a more pronounced domain expertise had difficulties following the more technical parts especially because they didn't previously know or use the presented components, but in the end they were able to understand their main advantages.

As a general comment, participants showed a large interest on the first platform prototype and promising expectations on its future usage. The answers and suggestions offer useful feedback to drive the project work of the following months.

4 Next Development Phases

Based on the evaluation results, additional work is needed to stress on some features still missing or less developed in the prototype in order to reach also less interested researcher groups.
Considering this and the work already planned for the second half of the project, the next development phases include the following.

- Improve the facility of customisation of workflows e.g. easing the registry access and modification, allowing for more dynamic import of external modules and definition of environmental variables, or considering the possibility of directly writing, testing, registering and linking new PEs directly in the Jupyter Notebook used for execution.
- Automation of computation, analysis, transfer and storage should be even more stressed based on the strong requirement from the end-user audience of hiding technical complexities to concentrate on main research objectives.
- Authentication and authorization are technical components still missing in the first platform prototype and focusing on them is fundamental in order to release an operational testbed for practical applications.
- Based on the strong requirement emerged from the evaluation of evidencing the operational applicability of the platform, it could be useful to develop a real application of the RA workflow for a test area (e.g. central Italy) of scientific interest. In this sense, also the implementation of the other more complex test cases (SS, ES) will provide a practical demonstration of the platform usage and usefulness both from the technical and scientific point of view.
- Integrate even more existing e-infrastructure services transparently to the user e.g. making use of B2DROP also for the seismological use case that will provide a more user-friendly management of input and output data.
- A well managed and easily accessible file system is required.
- Improve control messages for workflow management like those to monitor the processes.
- Produce some dissemination material showing practical examples of platform deployment and usage, workflow execution and customisation, implementation of new resources.
- Intelligent optimisation of the workflows avoiding user manual intervention.
- Allow for easy deployment of the DARE platform into local computing resources or existing Cloud resources (e.g. EOSC, Amazon, etc.) that will increase the usage and sustainability of the project also after its end.
- A future training, probably including a hands-on, could be useful to test the platform updates with respect to the first prototype and also for dissemination and outreach purposes to involve an increasing number of users.

5 General Conclusions

The RA test case was implemented and executed through the first prototype of the DARE platform following the work plan delineated in deliverable D6.1.

The workflow was successfully presented by showing a short demo during the pilot training organised by WP6 at INGV on June 27th, 2019.
The overall feedback from the training evaluation is very positive. The participants, thanks to the heterogeneity of their skills and expertise, were able to catch up the main advantages of using the platform, perfectly understanding the goals of DARE project. In general, they demonstrated large interest on the first prototype of the platform recognising how much this can improve their work. Thus they were very propositional about possible updates in future releases, looking forward to having a complete product directly usable for their operational work.
Based on the comments and suggestions gathered from the training, we refined and updated the work plan for the next development phase of the project aiming at creating a final comprehensive platform widely used among the Earth science community.

6 References

[Atkinson *et al.* 2015] M.P. Atkinson, M. Carpené, E. Casarotti, S. Claus, R. Filgueira, A. Frank, M. Galea, T. Garth, A. Gemünd, H. Igel, I. Klampanos, A. Krause, L. Krischer, S. H. Leong, F. Magnoni, J. Matser, A. Michelini, A. Rietbrock, H. Schwichtenberg, A. Spinuso, and J.-P. Vilotte, *VERCE delivers a productive e-Science environment for seismology research*, in Proc. IEEE eScience 2015.

[Peter *et al.* 2011] D. Peter, D. Komatitsch, Y. Luo, R. Martin, N. Le Goff, E. Casarotti, P. Le Loher, F. Magnoni, Q. Liu, C. Blitz, T. Nissen-Meyer, P. Basini & J. Tromp, 2011. *Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes*, Geophys. J. Int., 186, 721–739