Topic	H2020 – INFRAIA-2018-2020
Short Title	EurofleetsPlus
Title	An alliance of European marine research infrastructures to meet the evolving requirements of the research and industrial communities
Project Number	824077
Delivery Date	28/11/2022
Deliverable No	D3.26
Lead Beneficiary	OGS
Dissemination Level	Public

Report on validated virtual playground tools on the EVIOR portal

Document information	
Document Name	Report on validated virtual playground tools on the EVIOR portal
Document ID	EUROFLEETS_WP3_D3-26_09112022_V1.2_OGS
Revision	V1.2
Revision Date	10 November 2022
Author	Paolo Diviacco - OGS
Security	Public

Approvals			
	Name	**Organisation**	**Date**
Coordinator	Aodhán Fitzgerald	Marine Institute	25 November 2022
Activity Coordinator	Paolo Diviacco	OGS	25 November 2022
WP Leader	Arturo Castellon	CSIC	25 November 2022

History			
Revision	**Date**	**Modification**	**Author**
V1.0	07-11-2022	First draft	Paolo Diviacco
V1.1	09-11-2022	Review	Dick M.A. Schaap
V1.2	09-11-2022	Finalised and submitted	Paolo Diviacco

This document contains information, which is proprietary to the EUROFLEETS+ consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to any third party, in whole or in parts, except with prior written consent of the EUROFLEETS+ Coordinator.

The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

TABLE OF CONTENTS

Contents

1 Executive summary

The purpose of WP3 T3.1.4 is to extend the notion of VRE towards that of a Virtual Playground (VP) aiming to pave the road to a new vision of collaborative science that will facilitate to bridge the gaps between different scientific domains and ways of thinking while at the same time preserving peculiarities and backgrounds.

The Eurofleets+ Virtual Playground that has been made available within the JRA 3 task T3.1.4 aims to create a space of data, information and knowledge in which collaborative research can take place in the easiest way.

This deliverable describes the developed software and how to use it. The system can be found at: https://colla.ogs.it

It also can be found as one of the components of the EVIOR portal: see https://evior.eurofleets.eu/

The software has been extensively tested and has been presented at the Eurofleets+ Seismic workshop that took place 10-12 October 2022.

Chapter 2: Introduction. This gives an overview of the system architecture and an introduction to the software.

Chapter 3: System description and usage. This chapter describes the Functionalities of the EF+ VP system architecture. This section describes the different perspective used within the system in terms of the ontologies used and the Virtual Research Environment and how to access and use it.

Chapter 4: COLLA system architecture main components. This chapter describes the technical components of the system in terms of functionalities

Chapter 5: Seismic Workshop. This chapter describes the activities that took place during the Eurofleets+ Seismic workshop that took place 10-12 October 2022.

Appendix A provides references to better understand the system structure

Appendix B provides an acronym list and glossary

Appendix C Presentation of the Virtual playground at the EF+ Seismic workshop.

2 Introduction

The availability of new technologies, the limitations in resources and from a certain point of view also the COVID19 Pandemic, suggests that collaboration among scientists can be moved, for at least some aspects, to the web. There, researchers should confront their knowledge with other researchers from their or other disciplines.

OGS was already working on a collaborative tool aiming at supporting this new perspective through a project called COLLA (see reference section). COLLA aimed at supporting mostly the communicative aspects of collaboration, essentially defining a common map of knowledge and exchanging file, messages and information.

In this perspective the Virtual Playground (VP) developed within EF+ is a first trial to add to what already exists by providing a Virtual Research Environment where researchers have the possibility to process data and share not only results but also methods. This is achieved by sharing scripts themselves. This way it is possible to enlarge the possible methods to pack, transmit, decode and use knowledge. This can be done by means of formalized and non-formalized knowledge.

Deliverable D3.13 extensively described the implications of this perspective and highlighted the difficulties in handling the two approaches from a theoretical and practical point of view. D3.13 proposed a mix between those two approaches that should offer the possibility to retrieve information and data as needed. This is based on the instantiation of artifacts called boundary objects that allows to integrate at the same time different backgrounds and cultures while guiding the actual collaborative work.

Boundary objects are graphic maps that represent nodes and their relations. Nodes can be considered repositories for data and information. In this perspective the system with its map and data, information and knowledge that populate each node becomes the current status of knowledge of the designated community.

Using the COLLA framework, a knowledge space can be built on top of an ontology that can guide researchers' activities. The ontology can be built using software such as for example Protégé and uploaded in COLLA to be visualized using Ontodia, a web based ontology visualizer.

Specifications of the Virtual Playground are proposed in D3.17. The overall system description planned that three main modules should be present, and namely:

- Map (ontology) Viewer that allows to show the boundary object ontology to users and that redirects them to the information contained in each ontology node.
- Messaging within which data preview and processing can be accessed and shared.
- Virtual Research Environment (computing+data): where actual computation and processing on data can be done

This project has received funding from the EU
H2020 research and innovation programme
under Grant Agreement No 824077

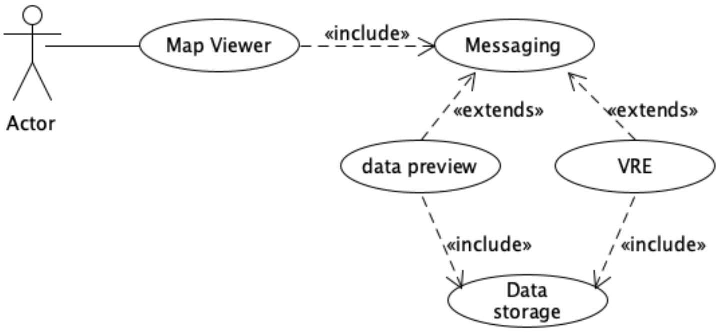


Figure 1 Virtual Playground system structure

3 System description and usage

This chapter describes the developed EF+ VP web based system in terms of its structure, functionalities and use

3.1 Functionalities

Colla functionalities can be gathered in two main areas:

- The project knowledge graph and space
- The project management related space

The former is explicitly visualized and used by the members of the team in order to drive through the topics of the discussion and reach the information they need, be it data, files, messages processing scripts or external nodes.

All this acquires meaning being mapped in the project graph and ontology, while the rest of knowledge can remain un-formalized being embedded in discussions, files or documents.
In addition to this a further layer of systematization is overlaid implicitly structuring the discussion in threads within each concept node.
The project management space is related to the hidden mechanisms that are used by COLLA to organize the collaboration.

This relates to user and access management which uses an ontology albeit this is not made public but used only internally by the software. For example, information on the users is stored using FOAF ontology although team members' basic information only is exposed.

This project has received funding from the EU H2020 research and innovation programme under Grant Agreement No 824077

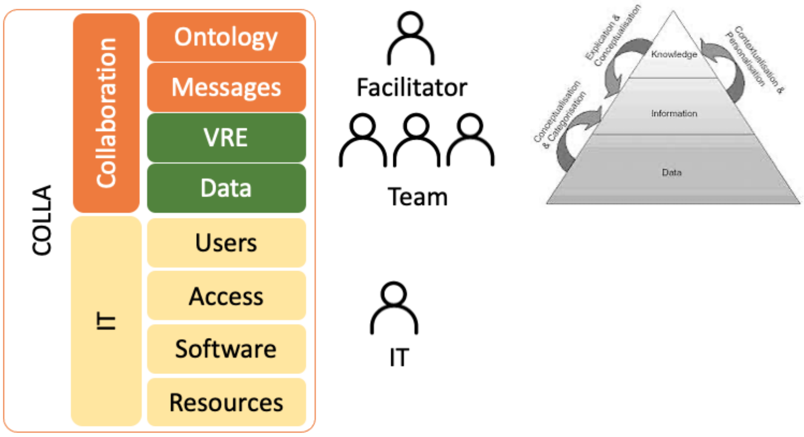


Figure 2 VP stack of modules and functionalities

3.2 Ontology Perspectives

Colla handles collaborative ontologies in two perspectives:

- **Domains**
- **Projects**

Domain can be used to represent a knowledge domain such as for example the relations between concepts in a scientific theory or hypothesis. This can be used to isolate observations, data processing and analysis or data products within homogeneous nodes so that users can easily retrieve information browsing the map.

The use of these types of ontologies is not guided and team leaders can upload any kind of ontology depending on what he/she considers useful for the description of the designated scientific domain. These ontologies must be created outside COLLA, using for example Protege, and uploaded in COLLA. Of course, there is no control on what is uploaded, that can be completely not useful or even wrong. The Project perspective is based on a set of predefined ontology classes and data and object properties.

These can be combined to produce the specific project ontology. This latter is built directly within COLLA using the *Project ontology builder*. This is a portal wizard that allows to create individuals and associate properties to them. Once the ontology is created it can directly be visualized and used as the project graph.

This project has received funding from the EU
H2020 research and innovation programme
under Grant Agreement No 824077

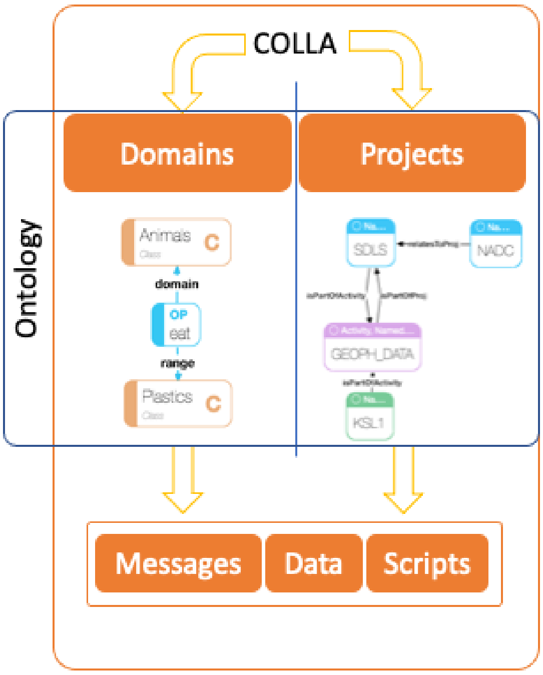
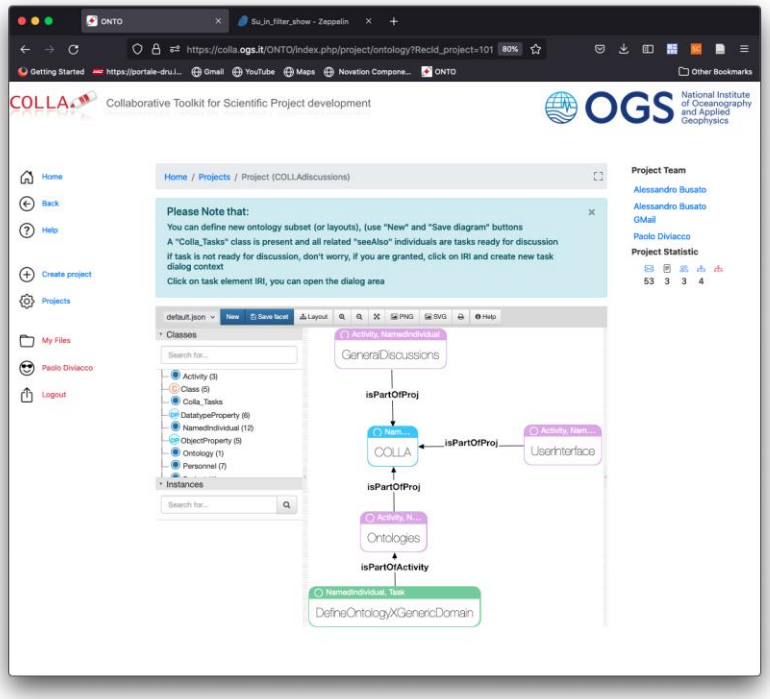


Figure 1 Two perspectives in using of COLLA, as a Domain or as Project representation. Each node of the represented ontology can accessCOLLA messages, data and scripting tools.

Both domain and project perspectives can access messages and data stored within each ontology node with the same mechanism. At the same time, it is possible to access and use the Virtual Research Environment (VRE) from both domains and projects, using the scripting tools and file manager provided by COLLA and at the same time embedding the scripts within COLLA messages.

The Map view tool shows an interactive map where nodes are listed. These nodes can be gathered in a graph. Multiple graphs can coexist since they refer to the same set of nodes.

This project has received funding from the EU
H2020 research and innovation programme
under Grant Agreement No 824077



Please Note that:

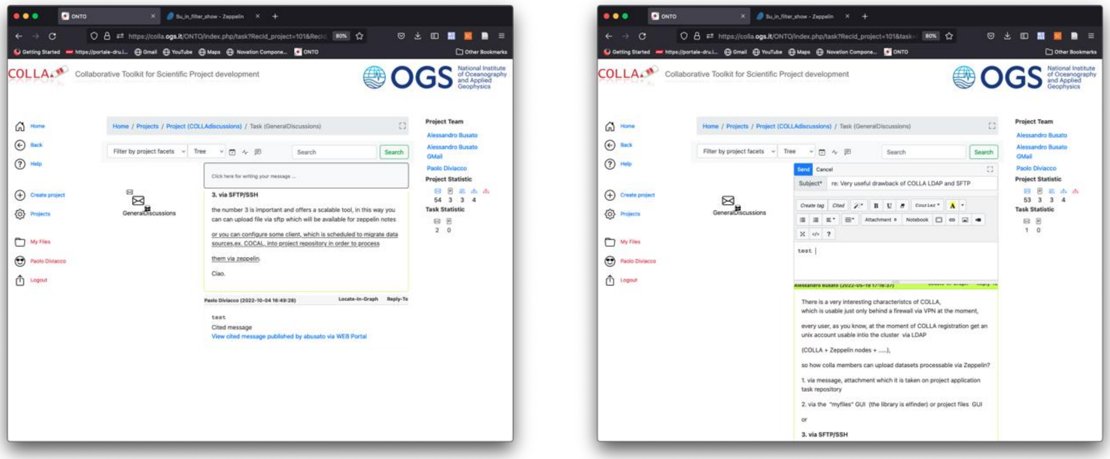
You can define new ontology subset (or layouts), (use "New" and "Save diagram" buttons

A "Colla_Tasks" class is present and all related "seeAlso" individuals are tasks ready for discussion

if task is not ready for discussion, don't worry, if you are granted, click on IRI and create new task dialog context

Click on task element IRI, you can open the dialog area

Figure 2 Ontology visualization in COLLA


ONTO
Su_in_filter_show - Zeppelin
https://colla.ogs.it/ONTO/index.php/task?RecId_project=101&RecId
80%
COLLA
Collaborative Toolkit for Scientific Project development
OGS
National Institute of Oceanography and Applied Geophysics
Home
Back
Help
Create project
Projects
My Files
Paolo Diviacco
Logout
Home / Projects / Project (COLLAdiscussions) / Task (GeneralDiscussions)
Filter by project facets
Tree
Search
Search
GeneralDiscussions
Click here for writing your message ...
Project Team
Alessandro Busato
Alessandro Busato GMail
Paolo Diviacco
Project Statistic
54 3 3 4
Task Statistic
2 0
Locate-In-Graph
Reply-To
https://colla.ogs.it/ONTO/index.php/task?RecId_project=101&task
Send
Cancel
Subject*
Create tag
Cited
Courier
Attachment
Notebook
Project Statistic
53 3 3 4
Task Statistic
1 0

3. via SFTP/SSH

the number 3 is important and offers a scalable tool, in this way you can can upload file via sftp which will be available for zeppelin notes

or you can configure some client, which is scheduled to migrate data sources.ex. COCAL. into project repository in order to process

them via zeppelin.

Ciao.

Paolo Diviacco (2022-10-04 16:49:28)

test
Cited message
View cited message published by abusato via WEB Portal

Subject* re: Very useful drawback of COLLA LDAP and SFTP

test

There is a very interesting characteristcs of COLLA,
which is usable just only behind a firewall via VPN at the moment,

every user, as you know, at the moment of COLLA registration get an unix account usable intio the cluster via LDAP

(COLLA + Zeppelin nodes + ...),

so how colla members can upload datasets processable via Zeppelin?

1. via message, attachment which it is taken on project application task repository

2. via the "myfiles" GUI (the library is elfinder) or project files GUI

or

3. via SFTP/SSH

Figure 3 Writing and replying to COLLA messages

Clicking one of the nodes it is possible to enter the messaging tool. There, messages are related graphically gathering them by multiple threads of subject. In figure 5 within a test node a single message was written, later a second message is sent as reply and is automatically mapped in the graph that represents the thread of discussion.

This project has received funding from the EU
H2020 research and innovation programme
under Grant Agreement No 824077

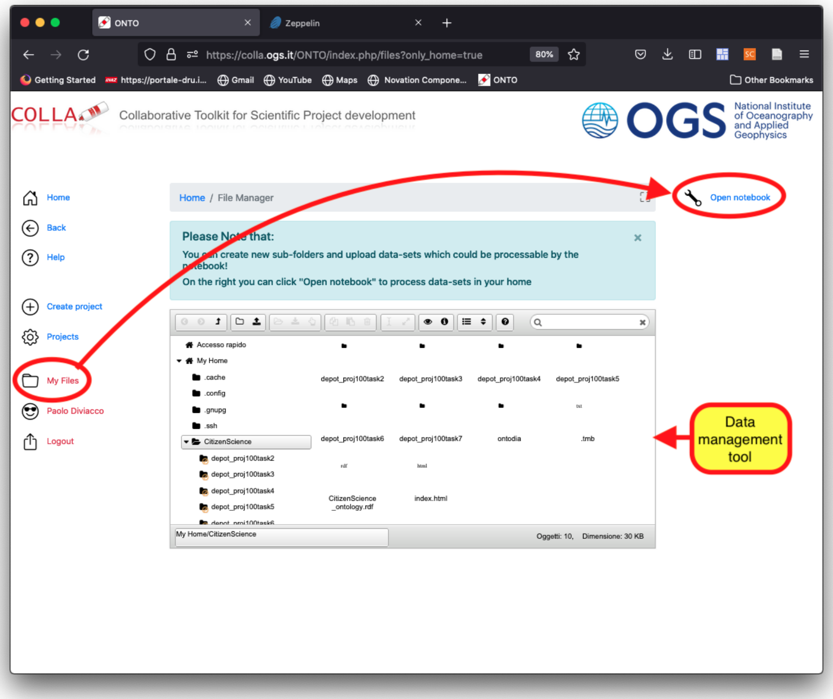
3.3 Virtual Research Environment (VRE)

Within the COLLA stack of functionalities the user is enabled to process data through scripts that are run within a notebook paradigm. Notebooks are web-based interactive computing platforms that are based on a server-client application that allows editing and running notebook documents via a web browser. The System is document centric.

Notebooks have two components: a front-end web page and a back-end kernel. The front-end web page allows data scientists to enter programming code or text. The browser then passes the code to the back-end kernel which runs the code and returns the results. Results are shown on the web browser.

Several notebook frameworks exist. COLLA VRE is based on Apache Zeppelin. Although Jupyter is a more known notebook framework, Zeppelin has several advantages. Among them the possibility to combine multiple paragraphs into one line, multi-user configuration via LDAP/Active Directory connectivity and a good integration with Hadoop.

Zeppelin provides interpreters for a lot of specific language/data-processing-backend. This allows users to use Python, C or almost any other scripting language.
Colla team developed a customized interpreter for the specific target of seismic data. This introduces an environment that allows for example to embed all functionalities available while scripting CWP Seismic Unix on remote shells.



Please Note that:
You can create new sub-folders and upload data-sets which could be processable by the notebook!
On the right you can click "Open notebook" to process data-sets in your home

Figure 4 Sections of the Virtual Playground on the portal

This project has received funding from the EU H2020 research and innovation programme under Grant Agreement No 824077

It is possible to access the VRE tools from within COLLA clicking on My Files in the menu. This triggers the visualization of the data management tool and of the notebook menu. The data management tool is an interactive file manager window where all data, images and files in general can be accessed.

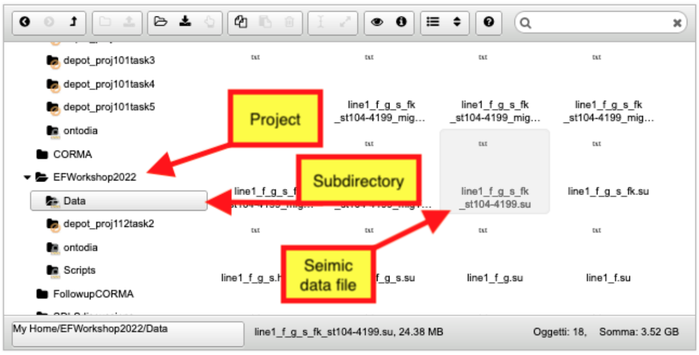


Figure 5 COLLA file manager

Files are organized automatically depending on the project they reside in (if they were uploaded within COLLA), while the tool allows users to upload others independently.

Each project has a specific folder. Each node of the ontology has a specific subfolder, where data can be located, in addition it is possible to create, link or mount directories that will be visualized in the file manager.

From the data management tool, each file can be quickly viewed and managed.

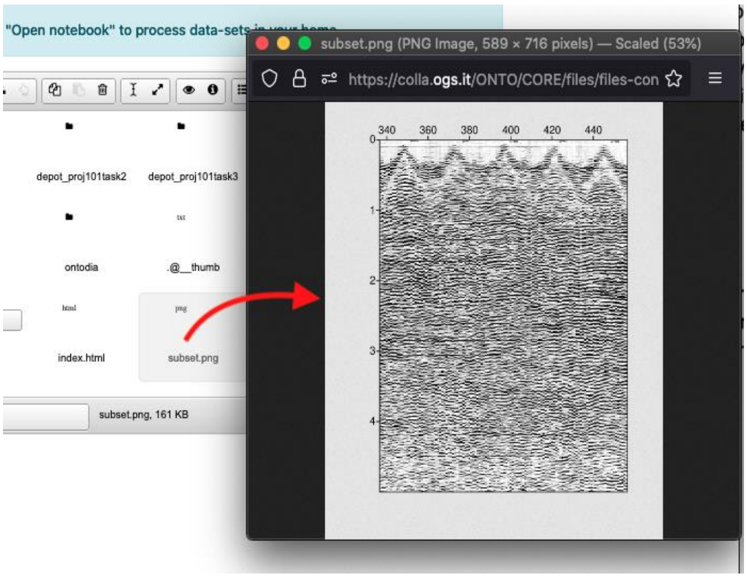
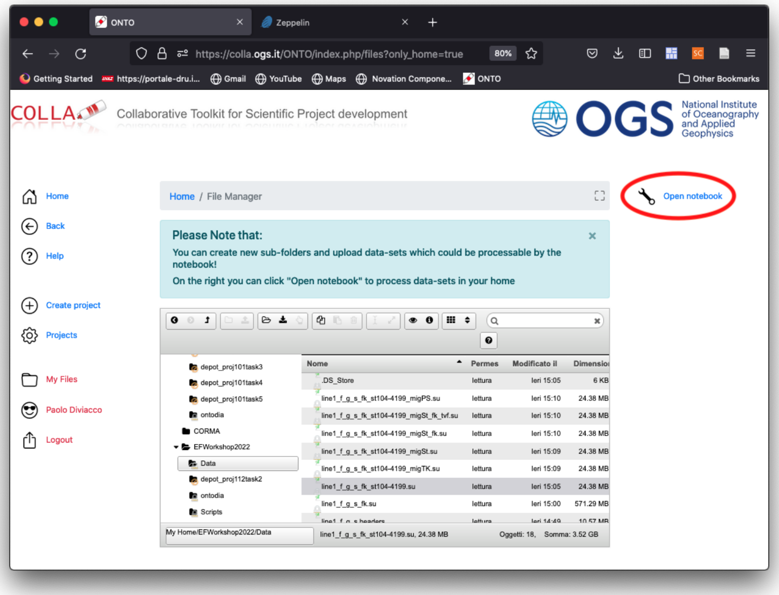

"Open notebook" to process data-sets in your home
subset.png (PNG Image, 589 × 716 pixels) — Scaled (53%)
https://colla.ogs.it/ONTO/CORE/files/files-con
depot_proj101task2
depot_proj101task3
ontodia
.@__thumb
index.html
subset.png
subset.png, 161 KB
340
360
380
400
420
440
0
1
2
3
4

Figure 6 quick view of files in COLLA File Manager

This project has received funding from the EU H2020 research and innovation programme under Grant Agreement No 824077

From the file management tool, it is possible to open the Processing tool, which launches the Notebook server (Zeppelin) where scripts can be executed, where notebook documents can be visualized and launched



Please Note that:

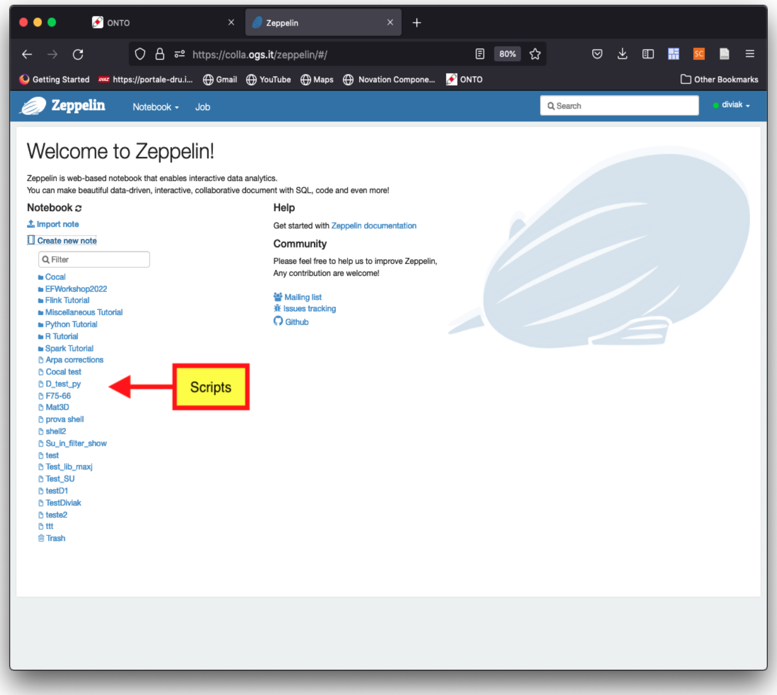
You can create new sub-folders and upload data-sets which could be processable by the notebook!

On the right you can click "Open notebook" to process data-sets in your home

Nome	Permes	Modificato il	Dimensio
.DS_Store	lettura	Ieri 15:05	6 KB
line1_f_g_s_fk_st104-4199_migPS.su	lettura	Ieri 15:10	24.38 MB
line1_f_g_s_fk_st104-4199_migSt_fk_tvf.su	lettura	Ieri 15:10	24.38 MB
line1_f_g_s_fk_st104-4199_migSt_fk.su	lettura	Ieri 15:10	24.38 MB
line1_f_g_s_fk_st104-4199_migSt.su	lettura	Ieri 15:09	24.38 MB
line1_f_g_s_fk_st104-4199_migTK.su	lettura	Ieri 15:09	24.38 MB
line1_f_g_s_fk_st104-4199.su	lettura	Ieri 15:05	24.38 MB
line1_f_g_s_fk.su	lettura	Ieri 15:00	571.29 MB
line1_f_g_s_headers	lettura	Ieri 14:49	10.57 MB

My Home/EFWorkshop2022/Data   line1_f_g_s_fk_st104-4199.su, 24.38 MB   Oggetti: 18,   Somma: 3.52 GB

Figure 7 How to access the VRE



Welcome to Zeppelin!

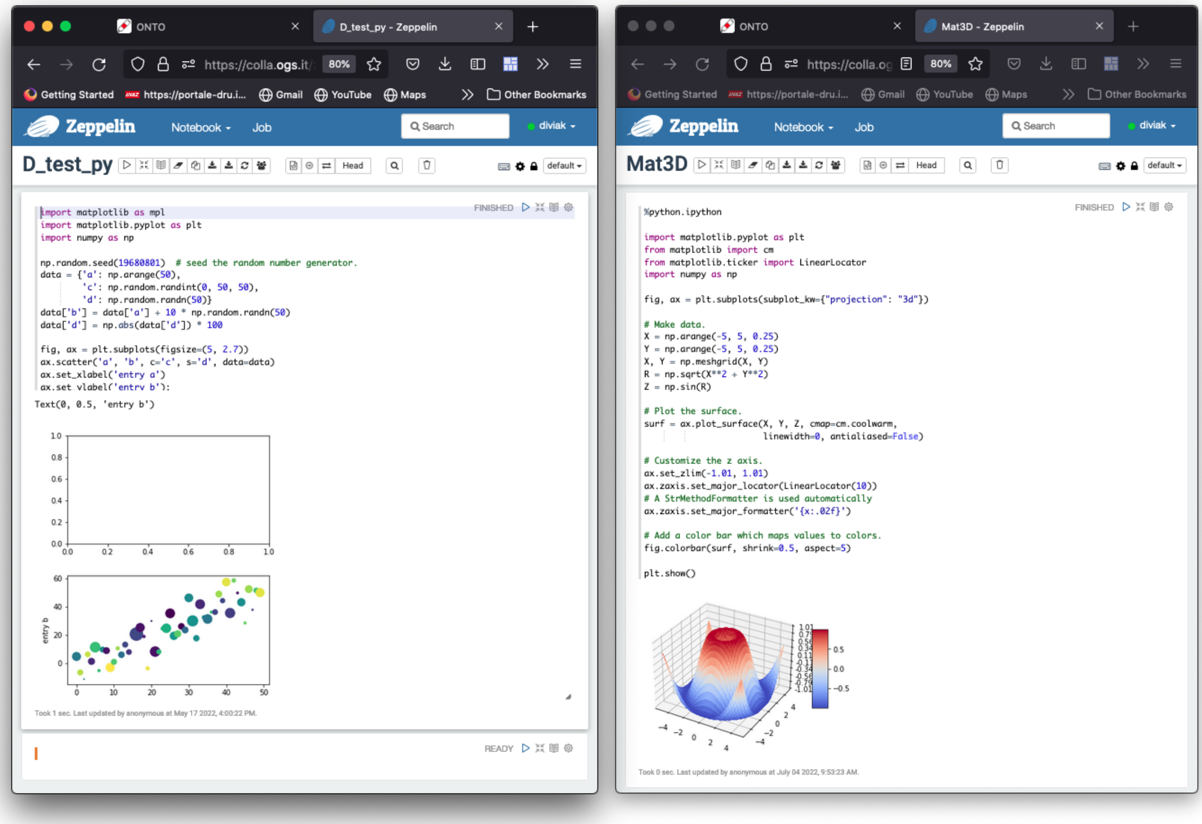
Zeppelin is web-based notebook that enables interactive data analytics.
You can make beautiful data-driven, interactive, collaborative document with SQL, code and even more!

Get started with Zeppelin documentation

Please feel free to help us to improve Zeppelin,
Any contribution are welcome!

Figure 8 VRE sections

This project has received funding from the EU
H2020 research and innovation programme
under Grant Agreement No 824077



```
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

np.random.seed(19680801)  # seed the random number generator.
data = {'a': np.arange(50),
        'c': np.random.randint(0, 50, 50),
        'd': np.random.randn(50)}
data['b'] = data['a'] + 10 * np.random.randn(50)
data['d'] = np.abs(data['d']) * 100

fig, ax = plt.subplots(figsize=(5, 2.7))
ax.scatter('a', 'b', c='c', s='d', data=data)
ax.set_xlabel('entry a')
ax.set_ylabel('entry b');
Text(0, 0.5, 'entry b')
```

Took 1 sec. Last updated by anonymous at May 17 2022, 4:00:22 PM.

```
%python.ipython

import matplotlib.pyplot as plt
from matplotlib import cm
from matplotlib.ticker import LinearLocator
import numpy as np

fig, ax = plt.subplots(subplot_kw={"projection": "3d"})

# Make data.
X = np.arange(-5, 5, 0.25)
Y = np.arange(-5, 5, 0.25)
X, Y = np.meshgrid(X, Y)
R = np.sqrt(X**2 + Y**2)
Z = np.sin(R)

# Plot the surface.
surf = ax.plot_surface(X, Y, Z, cmap=cm.coolwarm,
                       linewidth=0, antialiased=False)

# Customize the z axis.
ax.set_zlim(-1.01, 1.01)
ax.zaxis.set_major_locator(LinearLocator(10))
# A StrMethodFormatter is used automatically
ax.zaxis.set_major_formatter('{x:.02f}')

# Add a color bar which maps values to colors.
fig.colorbar(surf, shrink=0.5, aspect=5)

plt.show()
```

Took 0 sec. Last updated by anonymous at July 04 2022, 9:53:23 AM.

Figure 9 Virtual Playground executing complex python scripts and visualizing result

It is important to say that, currently, the approach is based on workflows that consist in a sequence of executions of programs that provide a result that can be visualized.

This approach has a limitation in interactivity. So far users provide parameters entering them manually in the document rather than graphically.

This in most of the cases is not a real problem. In some cases, where picking values graphically in needed, such as for example in the case of velocity analysis in seismics, this could be problematic. The limitation is under consideration for future development and we believe that the next versions of COLLA will overcome the problem.

Following the described workflow, we can run any kind of script (e.g. in figure python scripts based on matplotlib) to produce a graphic representation of the computation.

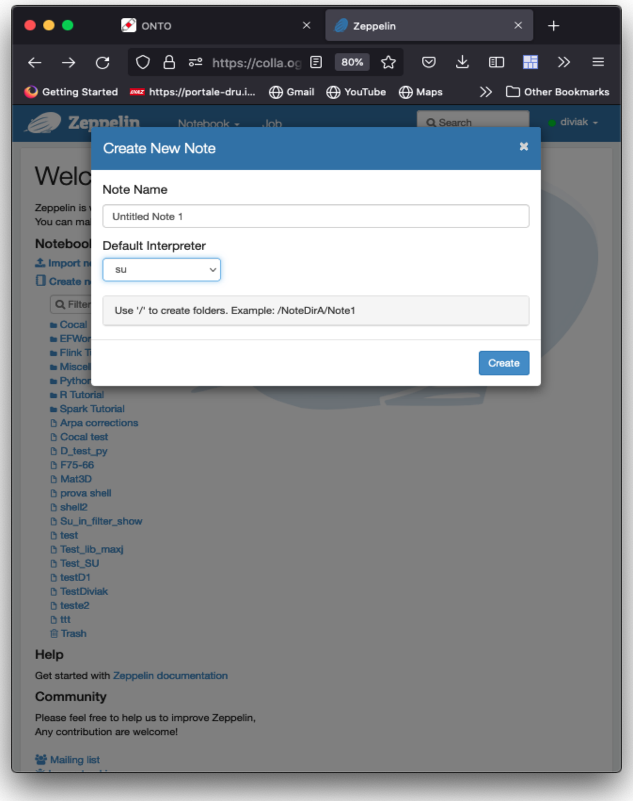
In order to use seismic data, we developed a specific interpreter in order to accommodate some problems with visualization and setting of environmental variables.

This project has received funding from the EU H2020 research and innovation programme under Grant Agreement No 824077

This interpreter was called ‘su’.

When new scripts aiming at seismic data processing need to be written, this profile should be used. For data processing of seismic data we rely on the CWP Seismic Unix software (https://cwp.mines.edu/software/).

Some parts of the code of the processing software were tuned and recompiled.
We will here show a practical example



Create New Note

Note Name: Untitled Note 1

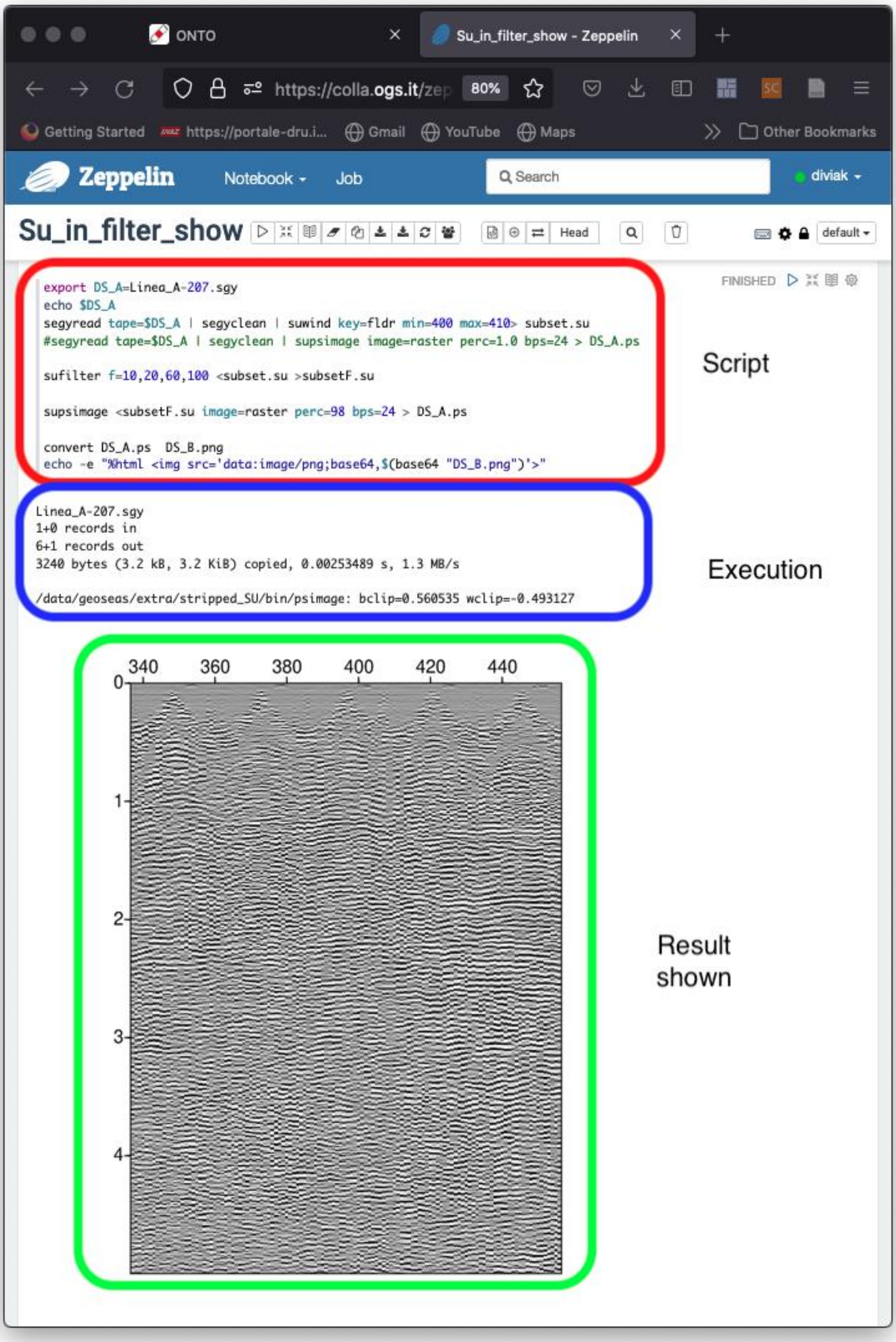
Default Interpreter: su

Use '/' to create folders. Example: /NoteDirA/Note1

Figure 10 Selecting the Seismic Unix interpreter

<table>
<tr><th>script command</th><th>comment</th></tr>
<tr><td>export DS_A=Linea_A-207.sgy</td><td>defines a variable
e.g.The data should reside in the directory where the the scripts runs or the absolute/relative path should be added</td></tr>
<tr><td>echo $DS_A</td><td>shows the variable value</td></tr>
<tr><td>segyread tape=$DS_A | segyclean | suwind key=fldr min=400 max=410> subset.su</td><td>Multiple commands can be appended through the '|' char. here segyread takes the filename from the DS_A variable. Data read are sent to segy clean that clean up the headers. Later a subset of the data is selected using the fldr header in the range form 400 to 410 to produce a file subset.su</td></tr>
<tr><td>sufilter f=10,20,60,100 <subset.su >subsetF.su</td><td>the subset is sent to a filter with the specified frequencies and a second file is produced</td></tr>
<tr><td>supsimage <subsetF.su image=raster perc=98 bps=24 > DS_A.ps</td><td>Very important: in order to visualize data a file has to be produced. SU usually show xwin unix windows, but since here we process data remotely using scripts this eventually has to produce a postscript file.</td></tr>
<tr><td>convert DS_A.ps DS_B.png</td><td>The postscript file need to be convertet to a file that can be visualized on a web browser (.png)</td></tr>
<tr><td>echo -e "%html <img src='data:image/png;base64,$(base64 "DS_B.png")'>"</td><td>This final line is very important because takes the file and shows it in the browser. It is a mix between scripting and HTML. The file that will be shown is DS_B.png</td></tr>
</table>

Figure 11 Example of Seismic Unix scripts to be executed within the Virtual Playground with comments on the right



```
export DS_A=Linea_A-207.sgy
echo $DS_A
segyread tape=$DS_A | segyclean | suwind key=fldr min=400 max=410> subset.su
#segyread tape=$DS_A | segyclean | supsimage image=raster perc=1.0 bps=24 > DS_A.ps

sufilter f=10,20,60,100 <subset.su >subsetF.su

supsimage <subsetF.su image=raster perc=98 bps=24 > DS_A.ps

convert DS_A.ps  DS_B.png
echo -e "%html <img src='data:image/png;base64,$(base64 "DS_B.png")'>"
```

```
Linea_A-207.sgy
1+0 records in
6+1 records out
3240 bytes (3.2 kB, 3.2 KiB) copied, 0.00253489 s, 1.3 MB/s

/data/geoseas/extra/stripped_SU/bin/psimage: bclip=0.560535 wclip=-0.493127
```

Figure 12 Seismic Unix script executed in the Virtual Playground using the Specific Interprete

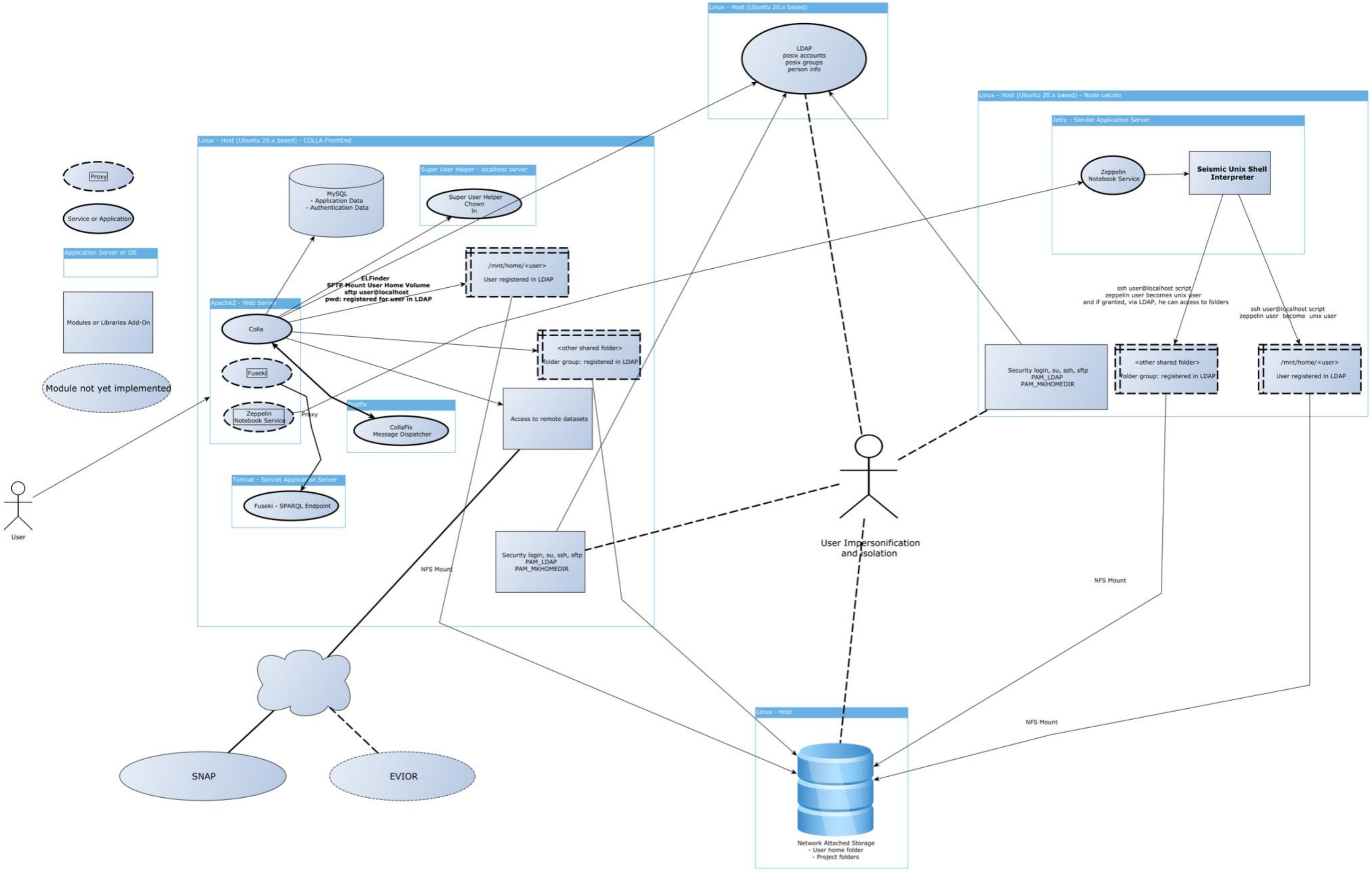


Figure 13 Detailed schema of the Virtual Playground ecosystem

Figure 15 shows the general architecture of the COLLA ecosystem at the moment. Datasets can be uploaded and processed to the system using the file manager. COLLA can be scaled to gather dataset to elaborate via a notebook directly to remote dataset sources.

COLLA it is a very modular system so new functionalities can be added or older can be updated with new features or new technologies, i.e. database. The modules are described in the next section

4 COLLA system architecture main components:

1. [*Collaboration PHP WebApp*] COLLA **Collaboration system server**

   COLLA implements the following main use cases:

   i. user registration;

   ii. project creation and its ontology definition;

   iii. project team discussion;

   iv. project ontology navigation;

   v. project administration

   vi. file manager to interact with user and project repository;

   vii. notebook interaction via REST api or directly via notebook webapp.

2. [*Notebook WebApp*] **Zeppelin Notebook Server**

   Zeppelin is a notebook server; it permits to user to write and execute notes based on the main language as Python, Spark, R, Sparql, Java

   It can be extended by writing a new interpreter as we did, implementing a shell customized for geophysics in order to execute Seismic Unix scripts.

   The authentication step is transparently managed byCOLLA portal through a smart authentication system based on project apache Shiro.

   Zeppelin supports user impersonation so every note can be executed by the user with his unix account and user home exploiting the LDAP component; this permits to enforce the isolation of user dataset by each other.

3. [*Persistent layer for Users and Projects application data*] **MySQL Server**

   It's the database chosen to save COLLA application data about registered users, projects, messages;

4. [*SPARQL Endpoint access to linked data*] **Jena-Fuseki SPARQL Endpoint Server**

   Every project is defined starting by an ontology which define semantically what project is; the ontology can be extended, preserving the ontology core, with new individuals;

5. [*Shared storage*] **NAS to share data between servers**

   Every node of architecture COLLA and one or more notebook server can be

This project has received funding from the EU H2020 research and innovation programme under Grant Agreement No 824077

harmonized using a common area where to save user home, project folder (which name maps a unix group); all of this is implemented using the Network Attached Storage;

6. [*Message dispatcher*] **CollaFix**

   CollaFix is a plugin writing by us extending the well known mail server Postfix in order to support team discussion messages dispatching; the messages type is the email; the one can be sent throw a form function offered by the portal or directly replying to the previous messages using the own preferred mail client; CollaFix keeps trace of the message updating consistently the application database; the email keeps the discussion strongly asynchronous and consents to the user to respond when he has the possibility and moreover , when he can do it free by distraction;

7. [*Unix user account and groups*] **LDAP server based on OpenLDAP**

   It's used to centralize and share unix account and user home and projects repository between servers. LDAP is a mature protocol to centralize in a hierarchical structure user account and groups information used by linux PAM modules to manage user permissions and groups;

8. [*HTTPS public frontend*] **Apache HTTP Server**

   Apache HTTP server is used in order to have a single point of access toCOLLA Virtual Playground via HTTPS keeping only standard port opened to Internet;

Currently all components are appropriately deployed into:

1. Server with Ubuntu 20.0 linux used by Colla
2. Computing-server with Ubuntu 20.0 linux used by Zeppelin, this host is used to execute effectively the notes written in the portal, for these reasons is more powerful than one used by COLLA portal; other computing servers can be added to the COLLA VRE.

References

- Diviacco,P., P. Fox, C. Pshenichny, & A. Leadbetter (Eds.), Collaborative Knowledge in Scientific Research Networks. Hershey, PA: IGI Global.
- Diviacco, P., & Busato, A. (2015). Maps, graphs and annotations as Boundary Objects in Knowl- edge Networks, Distributed Cognition and Collaborative e-Research. In P. Diviacco, P. Fox, C. Pshenichny, & A. Leadbetter (Eds.), Collabora- tive Knowledge in Scientific Research Networks. Hershey, PA: IGI Global.
- Diviacco, P. (2015). Reconciling Knowledge and Collaborative e-Research. In Collaborative Knowledge in Scientific Research Networks. Hershey, PA: IGI Global.
- Diviacco, P. Addressing Conflicting Cognitive Models in Collaborative E-Research: A Case Study in Exploration Geophysics" in Collaborative and Distributed E-Research: Innovations in Technologies, Strategies and Applications, IGI Global press, DOI: 10.4018/978-1-4666- 0125-3.ch012
- Diviacco P., Pshenichny C, Carniel R., Khrabrykh Z., Shterkun V. Mouromtsev D. Guzman S., Pascolo P. "Organization of a geophysical information space by using an event-bush- based collaborative tool" Earth Sci Inform, 2014, DOI 10.1007/s12145-014-0182-2
- Pshenichny C. Carniel R. and Diviacco P," Engineering of Dynamic Knowledge in Exact Sciences: First Results of Application of the Event Bush Method in Physics", in Proceedings of Engineering of Dynamic Knowledge in Exact Sciences: First Results of Application of the Event Bush Method in Physics
- Diviacco P Towards a collaborative research data-space in Geophysics" Mediterranean Marine Science 01/2011

Acronym list and glossary

COLLA	Collaborative tool developed by OGS
CWP SU	Center for Wave Phenomena at the Colorado School of Mines, Seismic Unix software
EF+	Eurofleets plus
EVIOR	European Virtual Infrastructure in Ocean Research
JRA	Joint Research Activity
RDF	Resource Description Framework
SPARQL	Semantic query language for databases that is able to retrieve and manipulate data stored in Resource Description Framework (RDF) format
URL	Uniform Resource Locator
VP	Virtual Playground
VRE	Virtual Research Environment

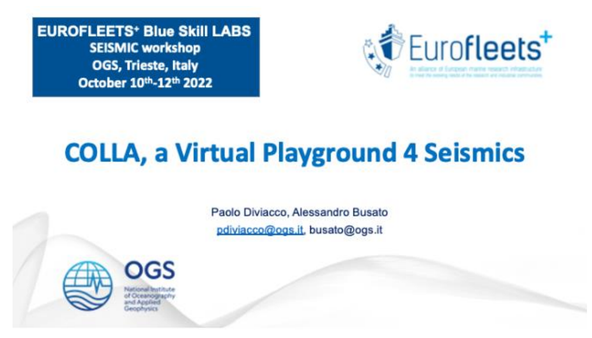
This project has received funding from the EU
H2020 research and innovation programme
under Grant Agreement No 824077

Annex: Presentation at EF+ Seismic workshop

The Virtual playground was presented at the EF+ Seismic workshop where four trainees used it with success.

This allowed both to test the system and the trainees to be provided with a tool to remotely access the resources and data shared within the workshop

In addition, using the messaging tool this allowed to create a small community of practice that can be used by the participants to share data, information and knowledge on the topic of Seismics and get in touch with experts that could help them in their career.

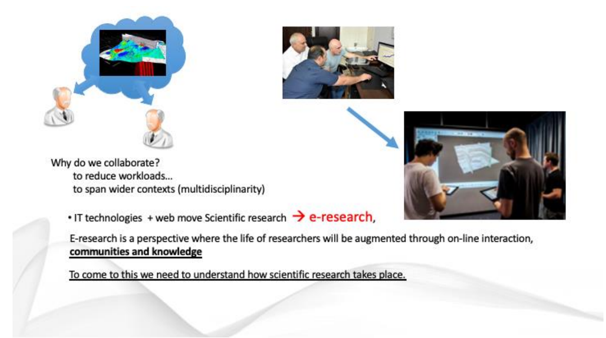


EUROFLEETS+ Blue Skill LABS
SEISMIC workshop
OGS, Trieste, Italy
October 10th-12th 2022

COLLA, a Virtual Playground 4 Seismics

Paolo Diviacco, Alessandro Busato
pdiviacco@ogs.it, busato@ogs.it

Eurofleets+
OGS
National Institute of Oceanography and Applied Geophysics

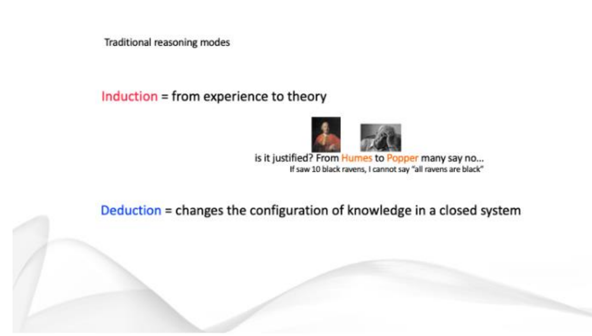


Why do we collaborate?
to reduce workloads…
to span wider contexts (multidisciplinarity)

- IT technologies + web move Scientific research → e-research,

E-research is a perspective where the life of researchers will be augmented through on-line interaction, **communities and knowledge**

To come to this we need to understand how scientific research takes place.

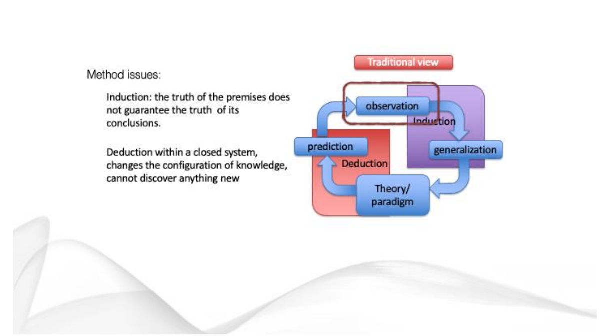


Traditional reasoning modes

Induction = from experience to theory

is it justified? From Humes to Popper many say no…
If saw 10 black ravens, I cannot say "all ravens are black"

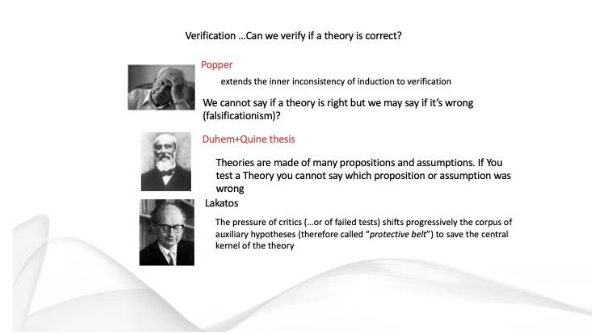
Deduction = changes the configuration of knowledge in a closed system



Method issues:

Induction: the truth of the premises does not guarantee the truth of its conclusions.

Deduction within a closed system, changes the configuration of knowledge, cannot discover anything new

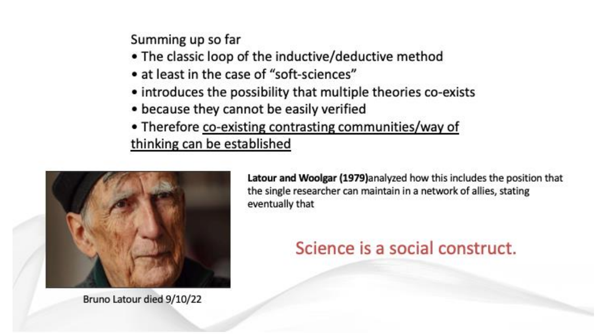


Verification …Can we verify if a theory is correct?

Popper
extends the inner inconsistency of induction to verification
We cannot say if a theory is right but we may say if it's wrong (falsificationism)?

Duhem+Quine thesis
Theories are made of many propositions and assumptions. If You test a Theory you cannot say which proposition or assumption was wrong

Lakatos
The pressure of critics (…or of failed tests) shifts progressively the corpus of auxiliary hypotheses (therefore called "*protective belt*") to save the central kernel of the theory

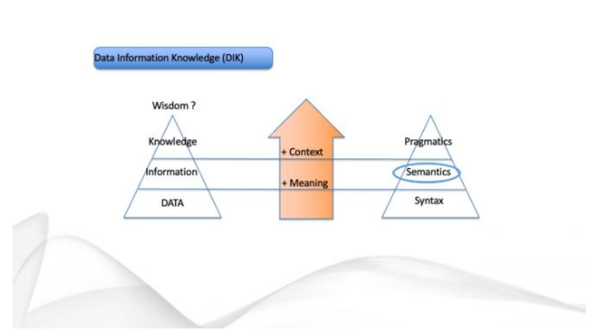


Summing up so far
- The classic loop of the inductive/deductive method
- at least in the case of "soft-sciences"
- introduces the possibility that multiple theories co-exists
- because they cannot be easily verified
- Therefore co-existing contrasting communities/way of thinking can be established

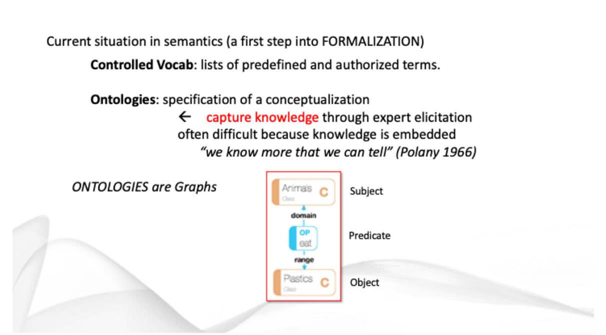
Latour and Woolgar (1979)analyzed how this includes the position that the single researcher can maintain in a network of allies, stating eventually that

Science is a social construct.

Bruno Latour died 9/10/22



Data Information Knowledge (DIK)



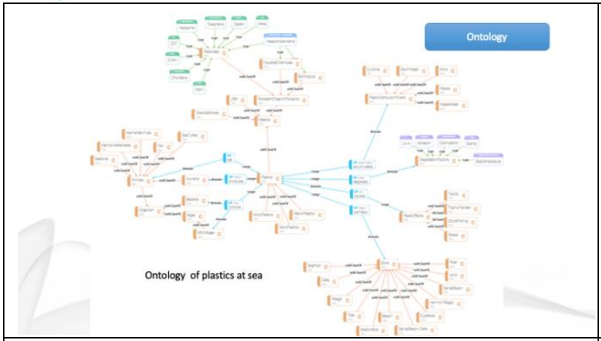
Current situation in semantics (a first step into FORMALIZATION)

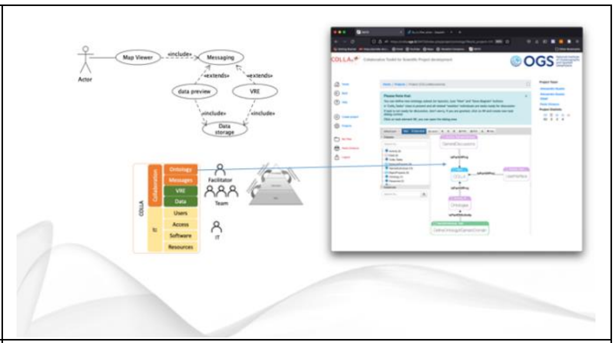
Controlled Vocab: lists of predefined and authorized terms.

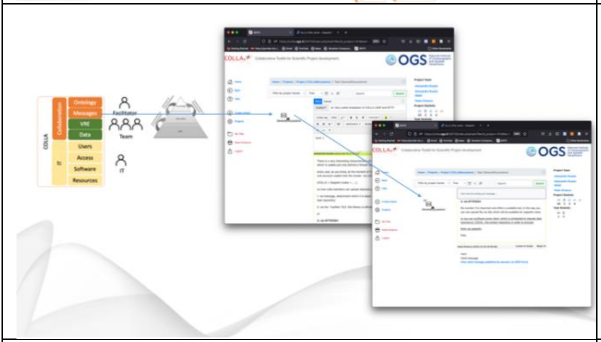
Ontologies: specification of a conceptualization
← capture knowledge through expert elicitation
often difficult because knowledge is embedded
"we know more that we can tell" (Polany 1966)

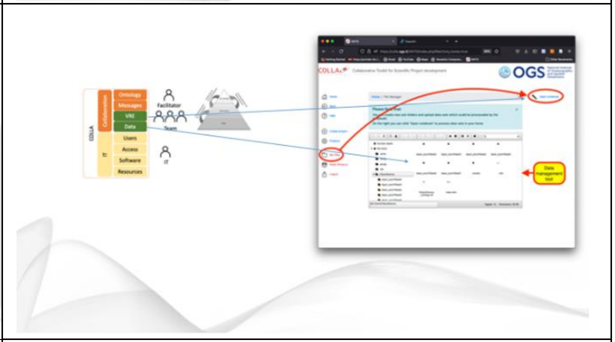
ONTOLOGIES are Graphs

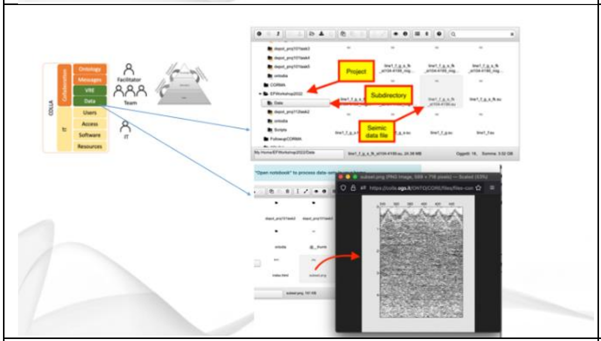
This project has received funding from the EU H2020 research and innovation programme under Grant Agreement No 824077

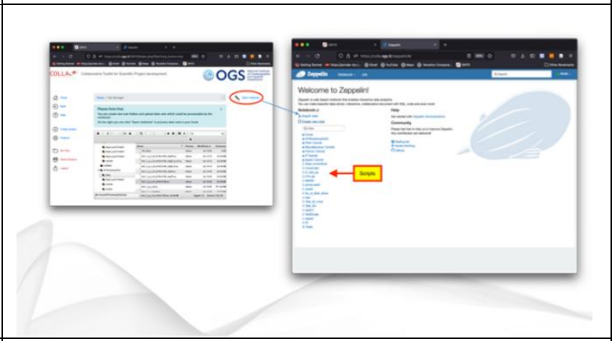


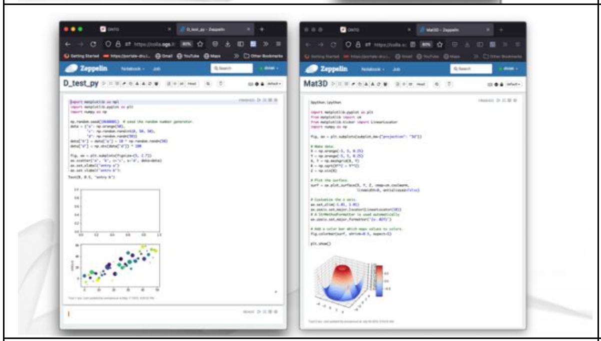


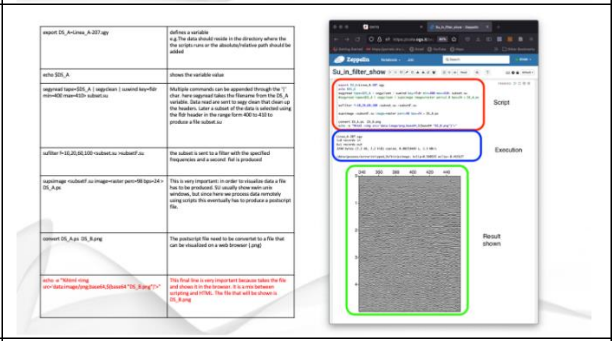




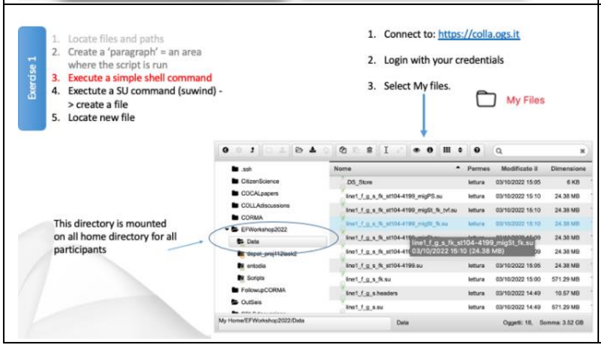








export DS_A=Linea_A-207.sgy	defines a variable e.g.The data should reside in the directory where the the scripts runs or the absolute/relative path should be added
echo $DS_A	shows the variable value
segyread tape=$DS_A | segyclean | suwind key=fldr min=400 max=410> subset.su	Multiple commands can be appended through the '|' char. here segyread takes the filename from the DS_A variable. Data read are sent to segy clean that clean up the headers. Later a subset of the data is selected using the fldr header in the range form 400 to 410 to produce a file subset.su
sufilter f=10,20,60,100 <subset.su >subsetF.su	the subset is sent to a filter with the specified frequencies and a second fiel is produced
supsimage <subsetF.su image=raster perc=98 bps=24 > DS_A.ps	This is very important: in order to visualize data a file has to be produced. SU usually show xwin unix windows, but since here we process data remotely using scripts this eventually has to produce a postscript file.
convert DS_A.ps DS_B.png	The postscript file need to be convertet to a file that can be visualized on a web browser (.png)
echo -e "%html <img src='data:image/png;base64,$(base64 "DS_B.png")'/>"	This final line is very important because takes the file and shows it in the browser. It is a mix between scripting and HTML. The file that will be shown is DS_B.png



Exercise 1

1. Locate files and paths
2. Create a 'paragraph' = an area where the script is run
3. Execute a simple shell command
4. Exectute a SU command (suwind) -> create a file
5. Locate new file

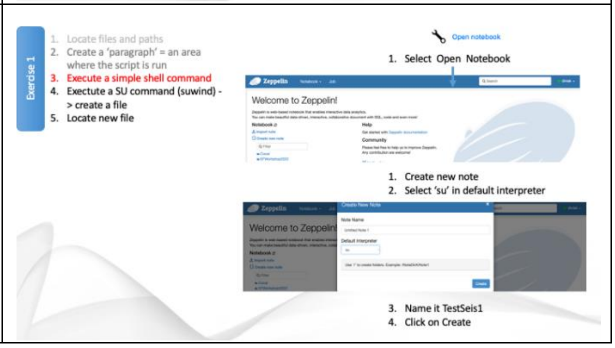
1. Connect to: https://colla.ogs.it
2. Login with your credentials
3. Select My files.

This directory is mounted on all home directory for all participants

Nome	Permes	Modificato il	Dimensione
.DS_Store	lettura	03/10/2022 15:05	6 KB
line1_f_g_s_fk_st104-4199_migPS.su	lettura	03/10/2022 15:10	24.38 MB
line1_f_g_s_fk_st104-4199_migSt_fk_tvf.su	lettura	03/10/2022 15:10	24.38 MB
line1_f_g_s_fk_st104-4199_migSt_fk.su	lettura	03/10/2022 15:10	24.38 MB
line1_f_g_s_fk_st104-4199_migSt...	lettura	03/10/2022 14:49	24.38 MB
line1_f_g_s_fk_st104-41...		...09	24.38 MB
line1_f_g_s_fk_st104-4199.su	lettura	03/10/2022 15:05	24.38 MB
line1_f_g_s_fk.su	lettura	03/10/2022 15:00	571.29 MB
line1_f_g_s.headers	lettura	03/10/2022 14:49	10.57 MB
line1_f_g_s.su	lettura	03/10/2022 14:49	571.29 MB

line1_f_g_s_fk_st104-4199_migSt_fk.su 03/10/2022 15:10 (24.38 MB)

My Home/EFWorkshop2022/Data Data Oggetti: 18, Somma: 3.52 GB



Exercise 1

1. Locate files and paths
2. Create a 'paragraph' = an area where the script is run
3. Execute a simple shell command
4. Exectute a SU command (suwind) -> create a file
5. Locate new file

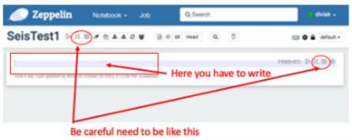
1. Select Open Notebook

1. Create new note
2. Select 'su' in default interpreter

3. Name it TestSeis1
4. Click on Create

Exercise 1

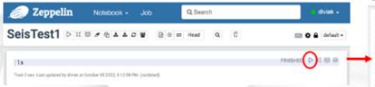
1. Locate files and paths
2. Create a ‘paragraph’ = an area where the script is run
3. Execute a simple shell command
4. Exectute a SU command (suwind) -> create a file
5. Locate new file



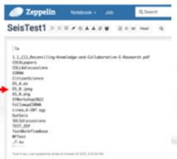
Here you have to write

Be careful need to be like this

1. Type in the paragraph area ‘ls’
2. Press the arrow
3. You will see a list of all the files in your home directory



```
ls
```

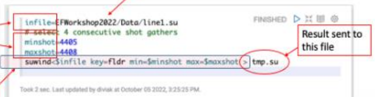


```
%ls
1.1_C(1)_Reconciling-Knowledge-and-Collaborative-E-Research.pdf
COCALpapers
COLLAdiscussions
CORMA
CitizenScience
DS_A.ps
DS_B.jpeg
DS_B.png
EFWorkshop2022
FollowupCORMA
Linea_A-207.sgy
OutSeis
SUSdiscussions
TEST_ZEP
TestWorkflowBase
WFTest
_f.su
```

Exercise 1

1. Locate files and paths
2. Create a ‘paragraph’ = an area where the script is run
3. Execute a simple shell command
4. Exectute a SU command (suwind) -> create a file
5. Locate new file

Gets data from the directory EFWORKSHOP2022/DATA
This is a shared read only directory monted to all home dirs for you convenience



```
infile=EFWorkshop2022/Data/line1.su
# select 4 consecutive shot gathers
minshot=4405
maxshot=4408
suwind<$infile key=fldr min=$minshot max=$maxshot > tmp.su
```

FINISHED

Took 2 sec. Last updated by diviak at October 05 2022, 3:25:25 PM.

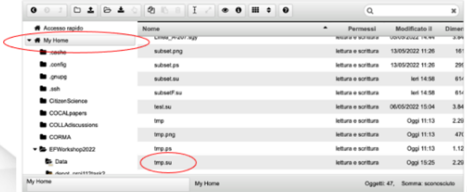
Variables defined in the script

Result sent to this file

Seismic Unix comand

Exercise 1

1. Locate files and paths
2. Create a ‘paragraph’ = an area where the script is run
3. Execute a simple shell command
4. Exectute a SU command (suwind) -> create a file
5. Locate new file



Nome	Permessi	Modificato il	Dimen
subset.png	lettura e scrittura	13/05/2022 11:26	161
subset.ps	lettura e scrittura	13/05/2022 11:26	29
subset.su	lettura e scrittura	Ieri 14:58	61
subsetF.su	lettura e scrittura	Ieri 14:58	61
test.su	lettura e scrittura	06/05/2022 15:04	3.84
tmp	lettura e scrittura	Oggi 11:13	2.29
tmp.png	lettura e scrittura	Oggi 11:13	47
tmp.ps	lettura e scrittura	Oggi 11:13	1.12
tmp.su	lettura e scrittura	Oggi 15:25	2.29

My Home — My Home — Oggetti: 47, Somma: sconosciuto

Exercise 2

1. Create a ‘paragraph’ = SeisTest2
2. Select subset of file (suwind)
3. Show subset
4. Filter tmp file
5. Show result

```
infile=EFWorkshop2022/Data/line1.su

# select 4 consecutive shot gathers
minshot=4405
maxshot=4408
suwind<$infile key=fldr min=$minshot max=$maxshot > subset.su

sufilter f=10,20,60,100 <subset.su >subsetF.su

supsimage <subset.su image=raster perc=98 bps=24 > subset.ps
convert subset.ps  subset.png

supsimage <subsetF.su image=raster perc=98 bps=24 > subsetF.ps
convert subsetF.ps  subsetF.png

echo -e "%html <img src='data:image/png;base64,$(base64 "subset.png")'>"
echo -e "%html <img src='data:image/png;base64,$(base64 "subsetF.png")'>"
```

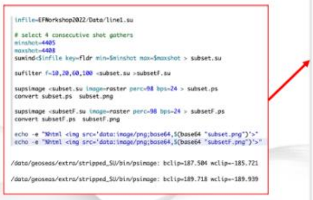
Since no interactivity no ximage
To produce ps files use supsimage

ps files cannot be used within a browser therefore we need to convert them to a png file

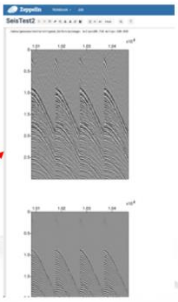
To show images within a web page we need to use HTML

Exercise 2

1. Create a ‘paragraph’ = SeisTest2
2. Select subset of file (suwind)
3. Show subset
4. Filter tmp file
5. Show result

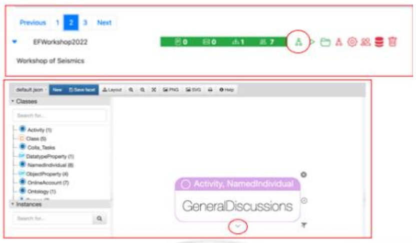


```
infile=EFWorkshop2022/Data/line1.su

# select 4 consecutive shot gathers
minshot=4405
maxshot=4408
suwind<$infile key=fldr min=$minshot max=$maxshot > subset.su

sufilter f=10,20,60,100 <subset.su >subsetF.su

supsimage <subset.su image=raster perc=98 bps=24 > subset.ps
convert subset.ps  subset.png

supsimage <subsetF.su image=raster perc=98 bps=24 > subsetF.ps
convert subsetF.ps  subsetF.png

echo -e "%html <img src='data:image/png;base64,$(base64 "subset.png")'>"
echo -e "%html <img src='data:image/png;base64,$(base64 "subsetF.png")'>"


/data/geoseos/extra/stripped_SU/bin/psimage: bclip=187.504 wclip=-185.721

/data/geoseos/extra/stripped_SU/bin/psimage: bclip=189.718 wclip=-189.939
```



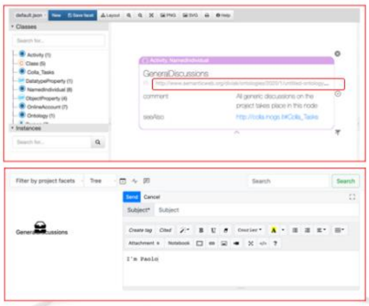
Exercise 3

1. Select node of ontology
2. Write message
3. Include notebook
4. Send message



Exercise 3

1. Select node of ontology
2. Write message
3. Include notebook
4. Send message



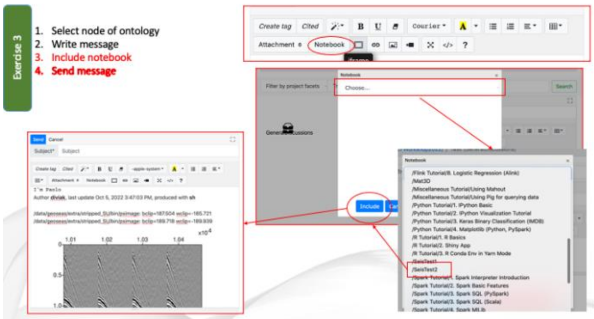
comment: All generic discussions on the project takes place in this node

seeAlso: http://colla.ingv.it#Colla_Tasks

I'm Paolo

Exercise 3

1. Select node of ontology
2. Write message
3. Include notebook
4. Send message



I'm Paolo

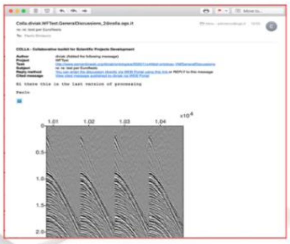
Author diviak, last update Oct 5, 2022 3:47:03 PM, produced with sh

/data/geoseos/extra/stripped_SU/bin/psimage: bclip=187.504 wclip=-185.721
/data/geoseos/extra/stripped_SU/bin/psimage: bclip=189.718 wclip=-189.939

Notebook

/Flink Tutorial/8. Logistic Regression (Alink)
/Mat3D
/Miscellaneous Tutorial/Using Mahout
/Miscellaneous Tutorial/Using Pig for querying data
/Python Tutorial/1. IPython Basic
/Python Tutorial/2. IPython Visualization Tutorial
/Python Tutorial/3. Keras Binary Classification (IMDB)
/Python Tutorial/4. Matplotlib (Python, PySpark)
/R Tutorial/1. R Basics
/R Tutorial/2. Shiny App
/R Tutorial/3. R Conda Env in Yarn Mode
/SeisTest1
/SeisTest2
/Spark Tutorial/1. Spark Interpreter Introduction
/Spark Tutorial/2. Spark Basic Features
/Spark Tutorial/3. Spark SQL (PySpark)
/Spark Tutorial/3. Spark SQL (Scala)
/Spark Tutorial/4. Spark MLlib

Received Colla mail



COLLA - Collaborative toolkit for Scientific Projects Development

Hi there this is the last version of processing

Paolo

This project has received funding from the EU H2020 research and innovation programme under Grant Agreement No 824077