Reliance

Research Lifecycle Management technologies for Earth Science Communities and Copernicus users in EOSC

Supporting reproducible OS in Earth Science with EGI and RELIANCE services

Raul Palma

RELIANCE Project Coordinator

Head of Data Analytics and Semantics Department

Poznan Supercomputing and Networking Center (PSNC)

Community Call for ARGOS and Data Management Plans (DMPs)

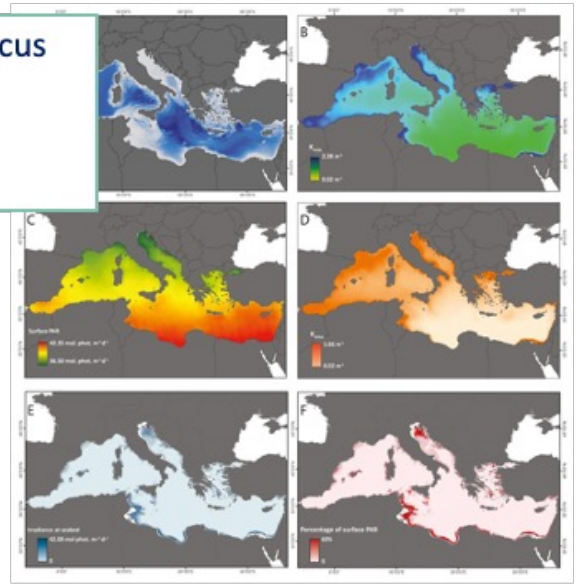
22th Februrary 2023

This project has received funding from the European research infrastructures (including e-Infrastructures) *under the European Union's Horizon 2020 research and innovation programme* under grant agreement No 101017501

How Earth Scientists usually work

- Remote sensing data

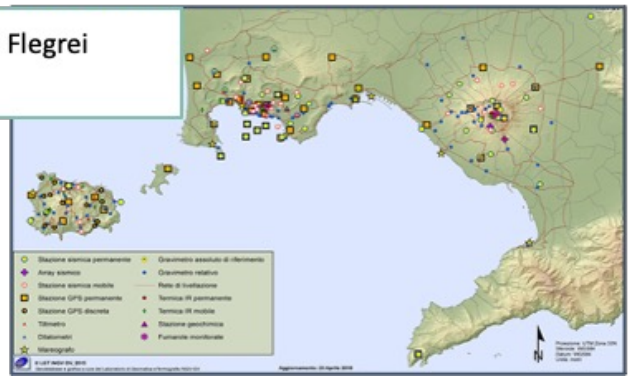
Copernicus
NASA
WOA
...



Data

- In-situ data from local networks

Vesuvius - Campi Flegrei (Italy)



Software/methods

- Data processing → Commercial and/or Personal tools (Matlab®, ENVI-IDL®, Python, R, etc.)
- Data modelling → Personal tools (Python, R, Matlab®, ENVI-IDL®, etc.)
- Post-processing → GIS (ArcGIS®, QGIS), GMT, etc.

Computing facilities

- Personal workstations
- Institutional HPC

Data – results management & dissemination

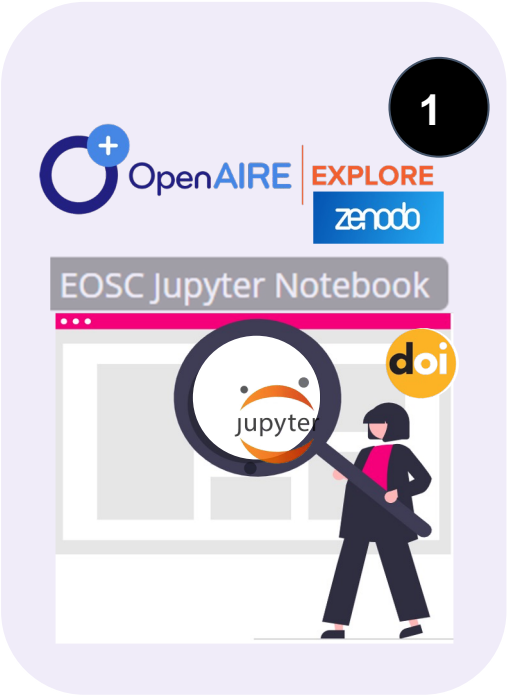
- Papers publication containing results and/or linked to data
- Results presented at meetings with ppt/poster stored
- Personal storage

Collaboration

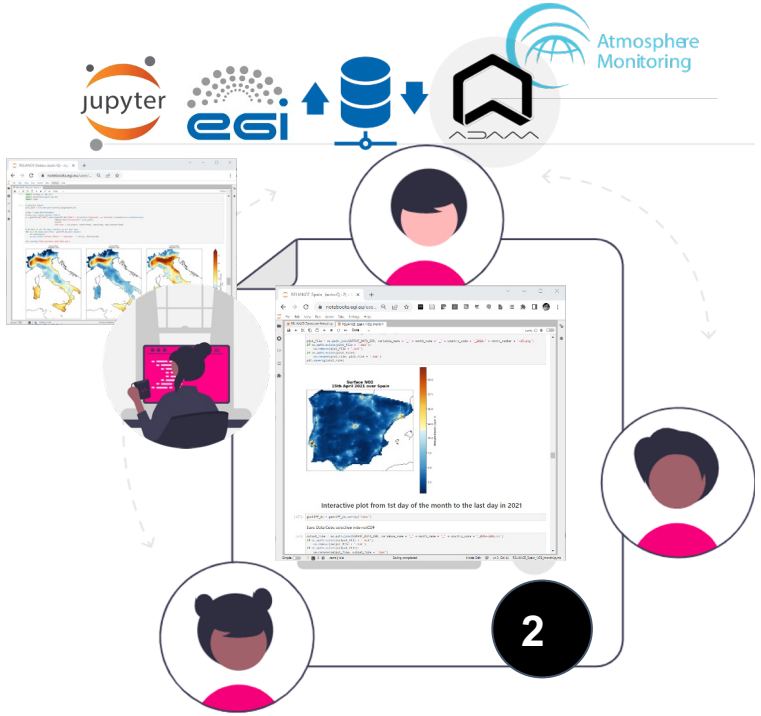
- In person, telecon
- Sharing by drive/cloud (Google Drive - institutional)

Reliance

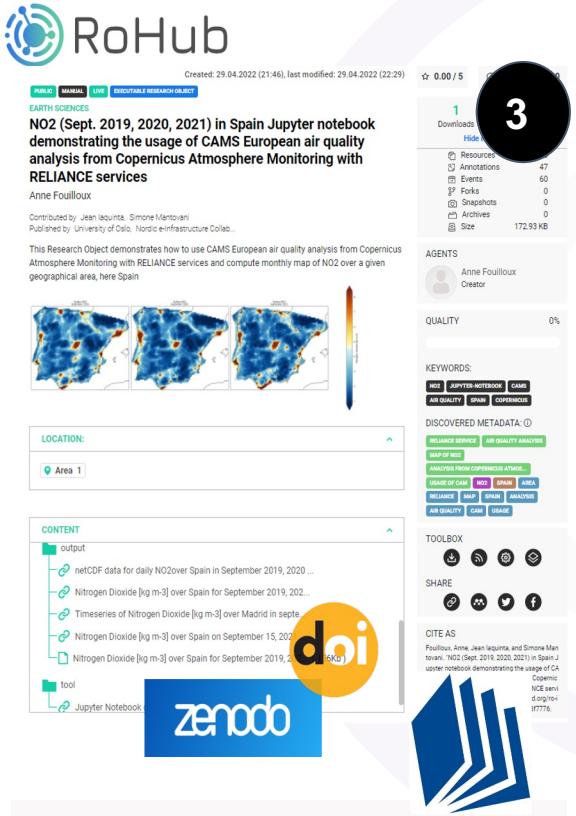
Motivation for FAIR and Open Science



Search, Find & Access



Reproducibility & Reuse



Publish & cite

Reliance

Motivation: challenges and vision

- The adoption of OS relies on having in place the mechanisms for FAIR sharing and re-using data and other research results to reproduce research and build upon it. This includes access to:
  - the data used to produce the results
  - the methods/code used to process the data and generate the results
  - documentation and explanation of the research
- On top of that, researchers should be able to easily reproduce the computing environment where these methods/codes were executed
  - Execution of research methods/code may depend on hundreds of different software packages, each with many different versions.
  - The reproducibility of results depends on having the correct version of these packages.

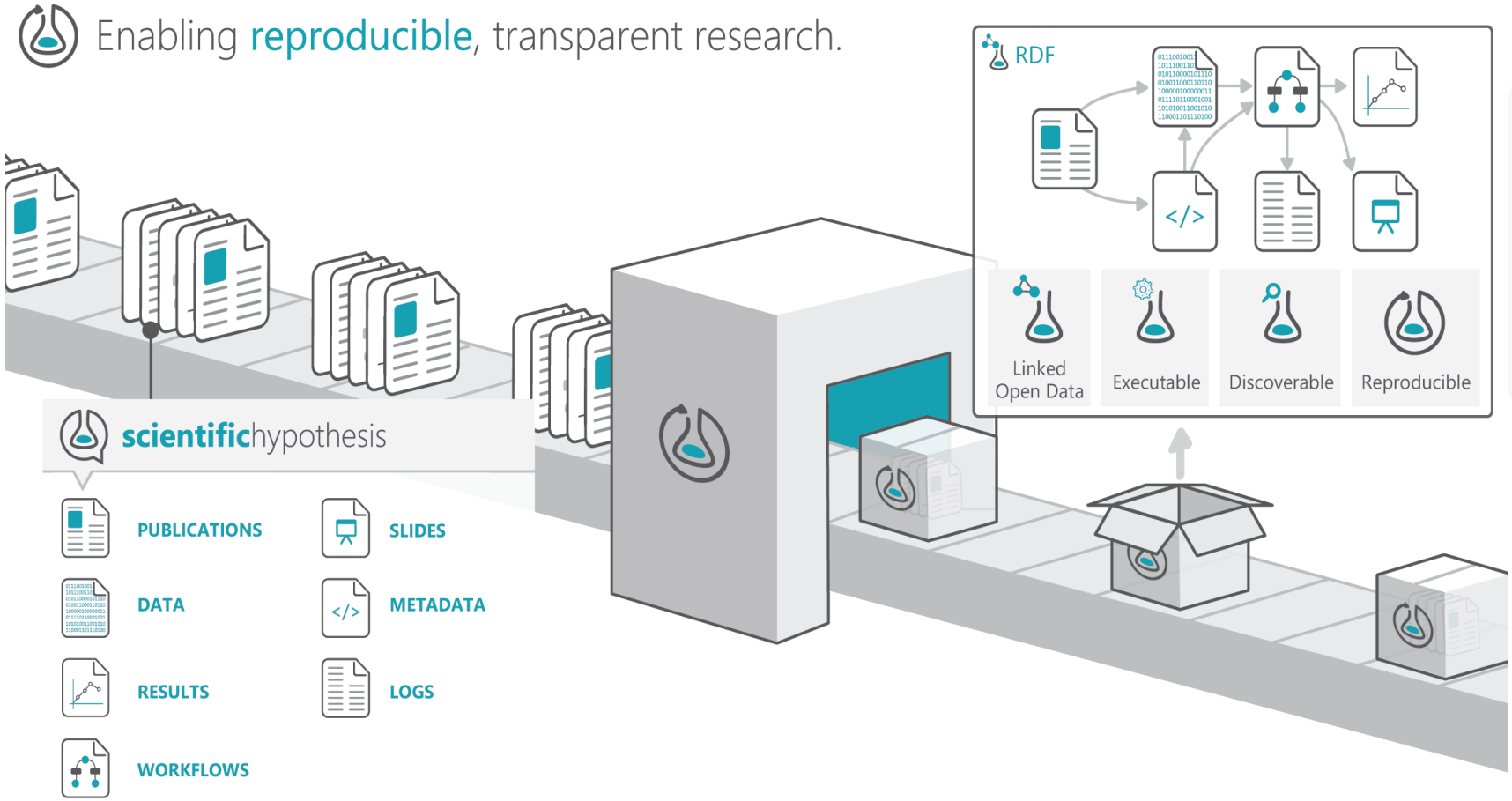
Reliance

Reliance

EGI-ACE

Research objects

Goal: Account, describe and share everything about your research, including how those things are related



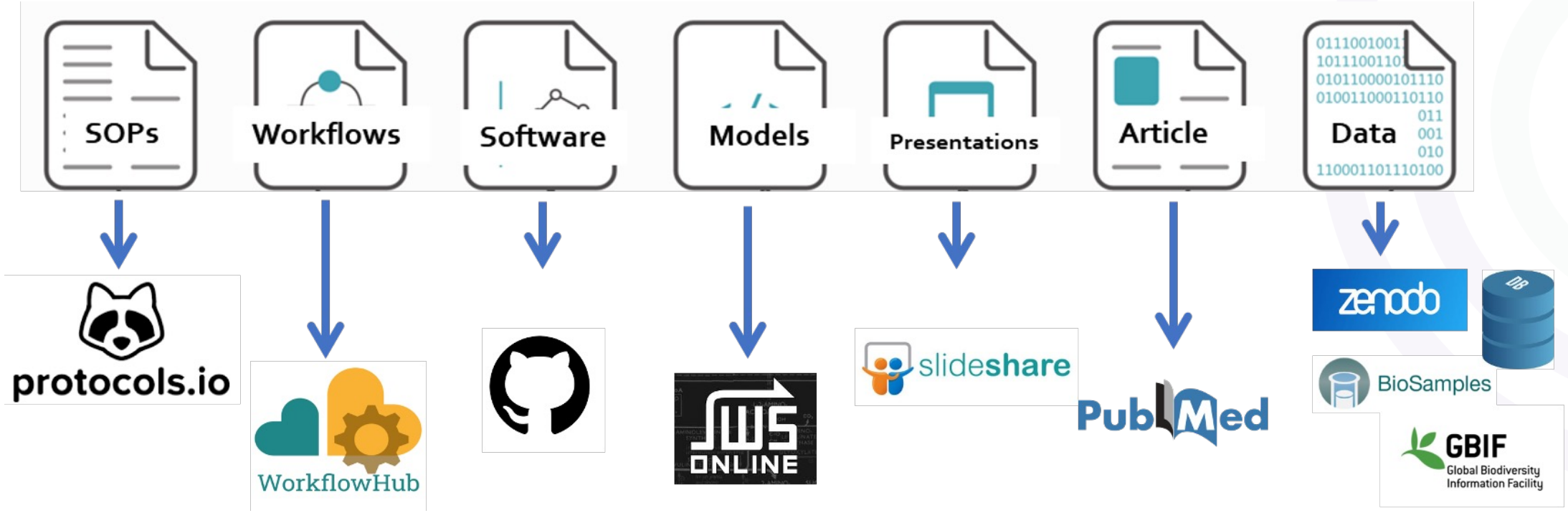
http://www.researchobject.org

Reliance

Research outcomes and related resources

All are first class citizens and are required to make research FAIR

Each object has its own metadata and repositories



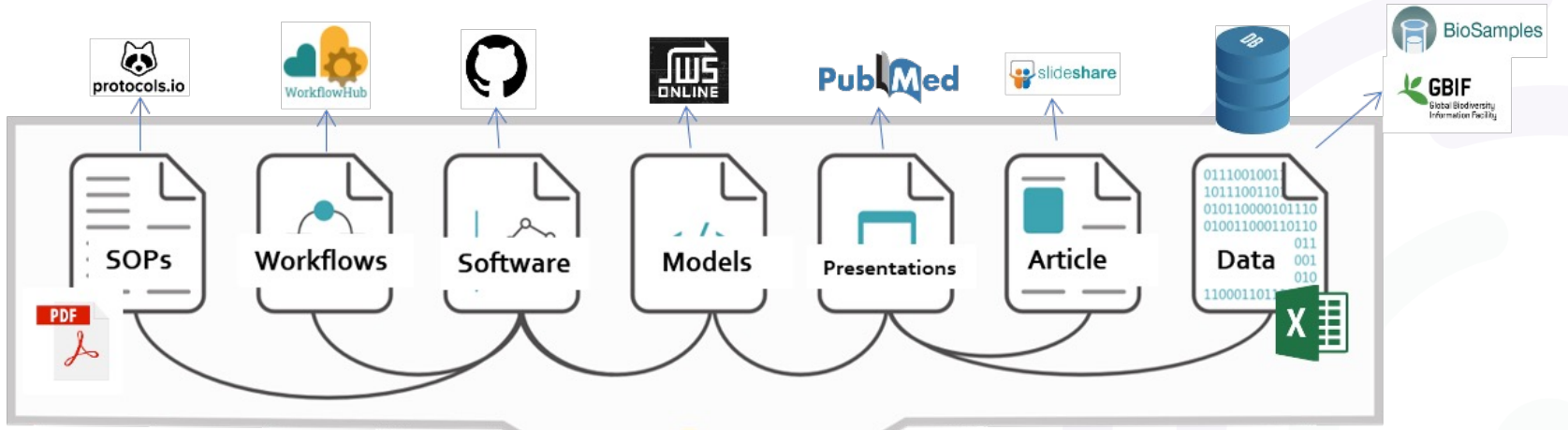
CC BY SA

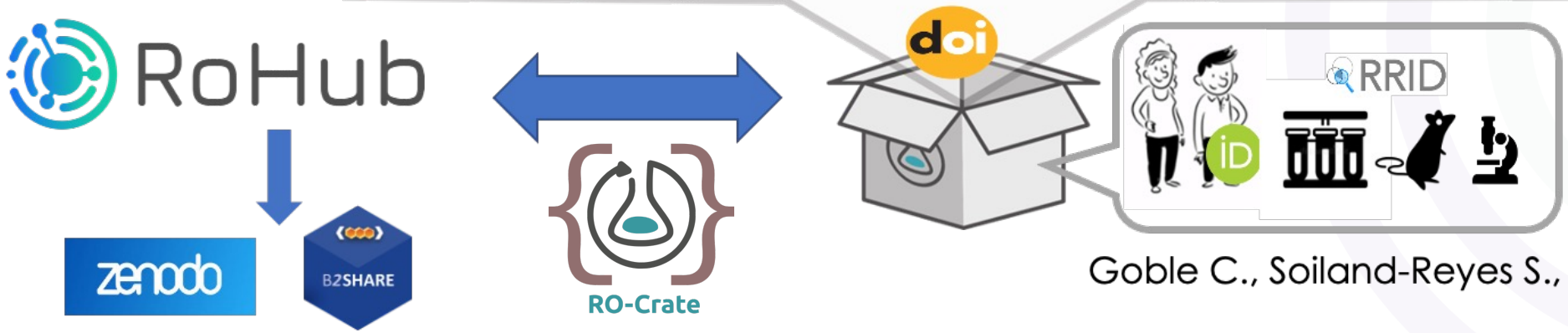
[source RO-Crate: A framework for packaging research products into FAIR Research Objects]

Reliance

Encapsulated content and references to external resources

integrated view over fragmented resources using PIDs and metadata





Goble C., Soiland-Reyes S.,

The RO has its own metadata, can be managed and evolved in its own right, and it can be packaged, deposited, transferred, accessed, and reproduced if appropriate

Workflow 4Ever

everest

http://www.researchobject.org

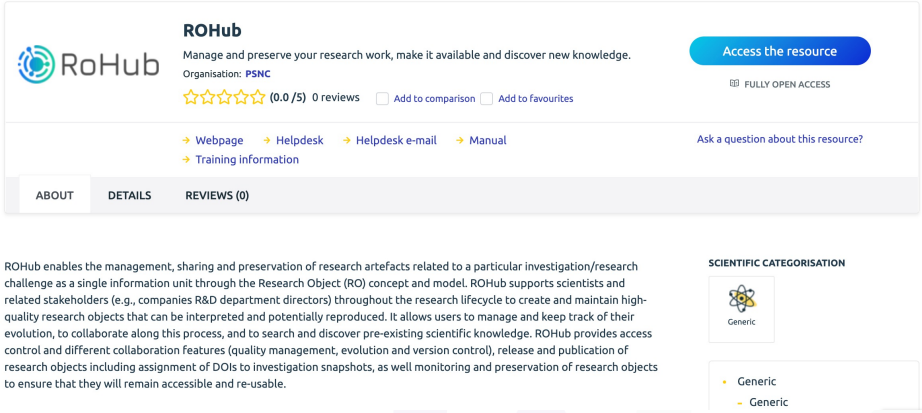
Reliance

ROHub overview

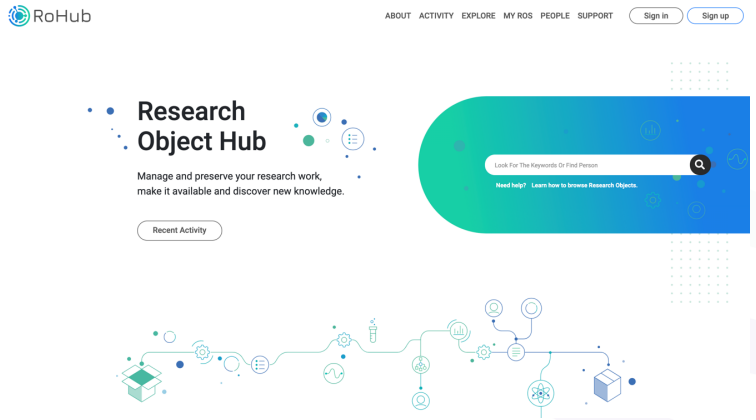
2010-2013 | 2014-2019 | 2020+

- Holistic solution for research object management, enabling
  - storage, lifecycle mgmt. & preservation of scientific outcomes
  - share and makes these resources available to others
  - publish and release them through a DOI
  - discover and reuse pre-existing scientific knowledge.
- Reference platform
  - implements natively the RO model and paradigm
  - support different stakeholders, with the primary focus on scientists, researchers, students and enthusiasts
  - provides the backbone to a wealth of RO-centric applications and interfaces across different scientific communities

Onboarded and integrated in EOSC

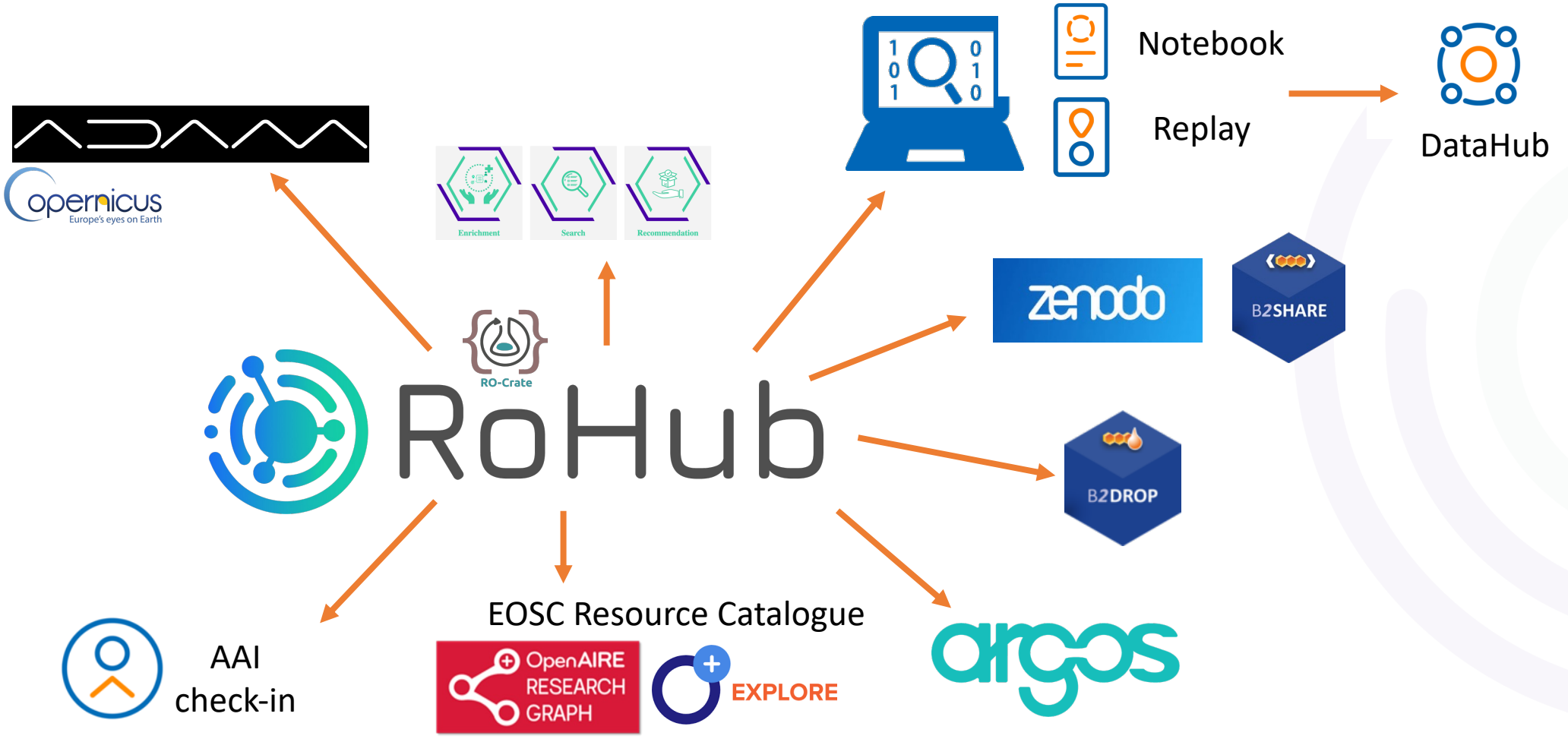


https://reliance.rohub.org/



3025	144	12498	51467
Research Objects	Users	Resources	Annotations

ROHub connections with EGI and other EOSC services

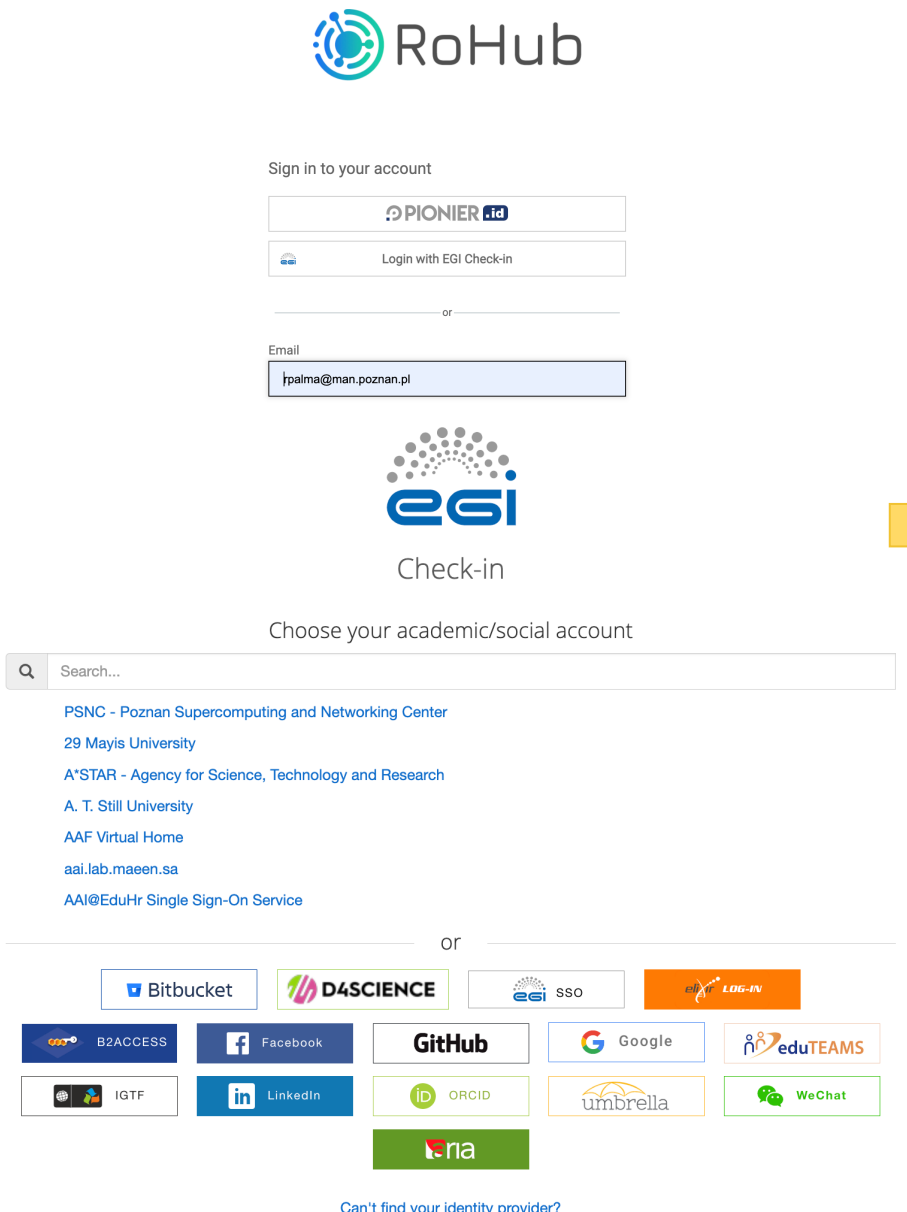


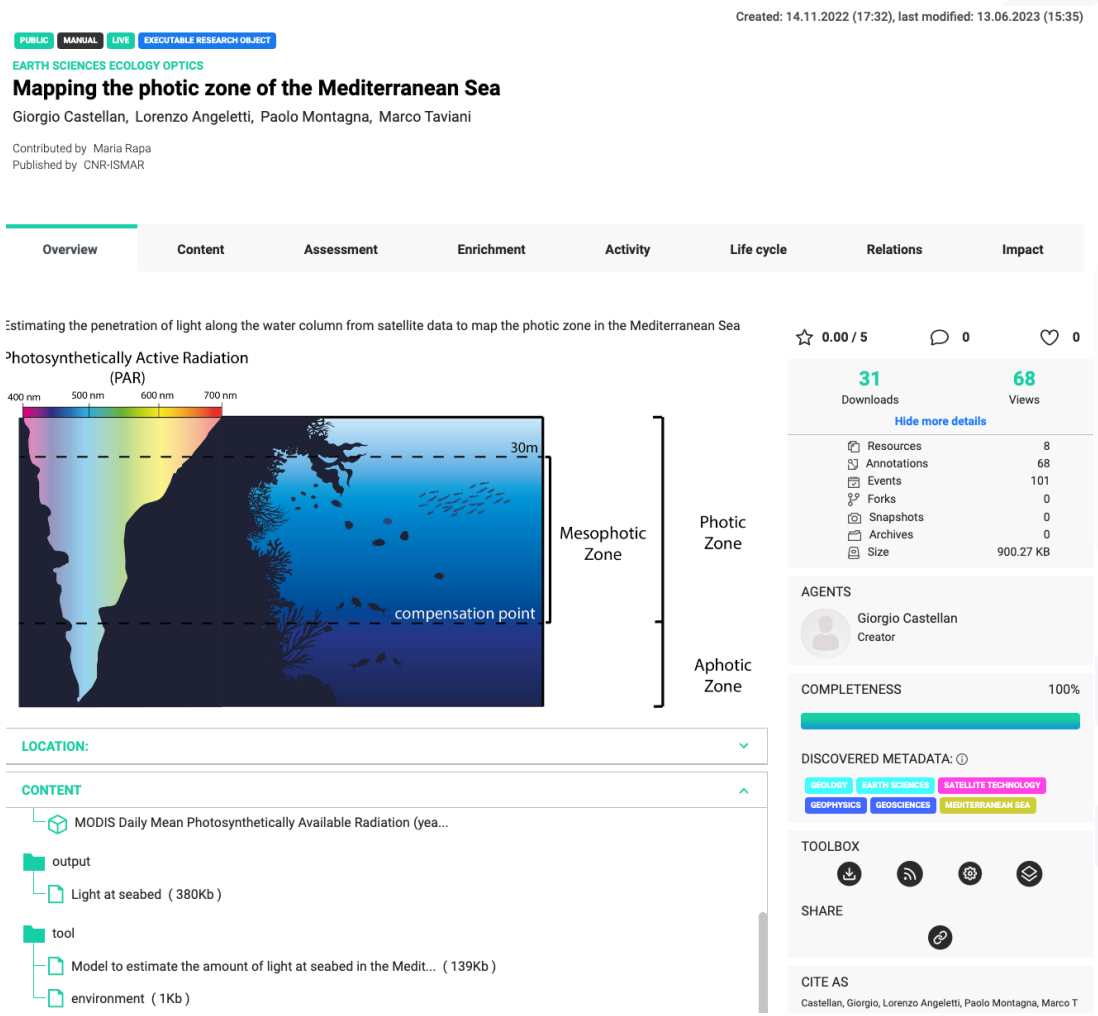
Reliance

EGI services integrated to support reproducible OS

- EGI check-in
  - allows users to access different EOSC services uniformly, using their preferred IdP
- EGI Notebook
  - a browser-based tool, based on JupyterHub, for interactive analysis of data using EGI storage and compute services.
- EGI Replay
  - a service, based on Jupyter Binder, to reproduce notebooks and their run-time environment.
- EGI DataHub
  - to bring data close to computing to exploit it efficiently and/or to publish and share data.

Services in practice





Reliance

Services in practice

Overview | Content | Assessment | Enrichment | Activity | Life cycle

Download resource

Open resource in EGI Notebook

Name	Details	Created
Model to estimate the amount of light at seabed in the Mediterranean Sea	139Kb	15.11.2022 (18:06)
environment	1Kb	16.11.2022 (13:40)

Created: 15.11.2022 (18:06), last modified: 15.11.2022 (18:06)

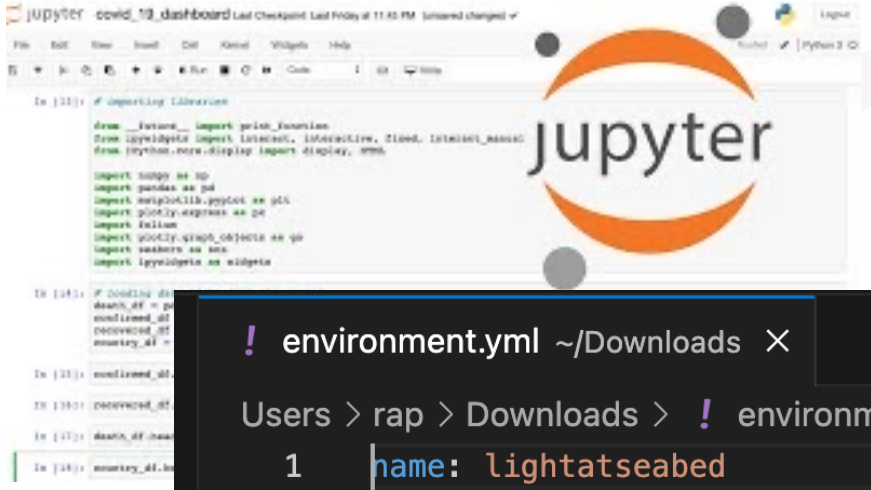
JUPYTER NOTEBOOK LOCAL

Model to estimate the amount of light at seabed in the Mediterranean Sea (139Kb)

Castellanetal_ScientificReports_Model.ipynb

Description

Jupyter Notebook reporting the model to estimate the amount of light at seabed



environment.yml ~/Downloads

Users > rap > Downloads > environment.yml

```yaml
name: lightatseabed
dependencies:
  - numpy
  - gdal
  - pip
  - pip:
    - matplotlib
```

Relations

Sort by: Sources

Filter relations by: All

List by: Date

nce

Services in practice

binder

Starting repository: rohub/notebooks/link_da4d3c23-c1b9-3cc1-9361-ce91de445e95

Your launch may take longer the first few times a repository is used. This is because our machine needs to create your environment.

Build logs

```
 ---> Running in c63e2afd60ec
Removing intermediate container c63e2afd60ec
 ---> 495054718b3d
Step 47/51 : ENV PYTHONUNBUFFERED=1
 ---> Running in 894f532d90b5
Removing intermediate container 894f532d90b5
 ---> 0890aee72e99
Step 48/51 : COPY /python3-login /usr/local/bin/python3-login
 ---> 9fc39b532633
Step 49/51 : COPY /repo2docker-entrypoint /usr/local/bin/repo2docker-entrypoint
 ---> b8690e194edc
Step 50/51 : ENTRYPOINT ["/usr/local/bin/repo2docker-entrypoint"]
 ---> Running in a50a9a9baf6a
Removing intermediate container a50a9a9baf6a
 ---> 3345148c09e
Step 51/51 : CMD ["jupyter", "notebook", "--ip", "0.0.0.0"]
 ---> Running in 1d85e8ff28da
Removing intermediate container 1d85e8ff28da
 ---> 594b8811620d
{"aux": {"ID": "sha256:594b8811620d811a06eafbbf5975a56c96afiad12f178becf8337523ca9e555e"}}Successfully built 594b8811620d
Successfully tagged docker-notebooks.fedcloud-tf.fedcloud.eu/binder-rohub-2dnotebooks-1900d1:0d4bceda46a27d517eeb853cdd703770d729ec1a
Pushing image
```

Castellanetal_ScientificReports_Model.ipynb

File Edit View Run Kernel Tabs Settings Help

Download GitHub Binder Markdown

Reliance

Estimating the amount of sunlight reaching the seabed from satellite data

[6]:
```
import matplotlib
import matplotlib.pyplot as pyplot
from osgeo import gdal
import numpy as np
import warnings
```

Import satellite data

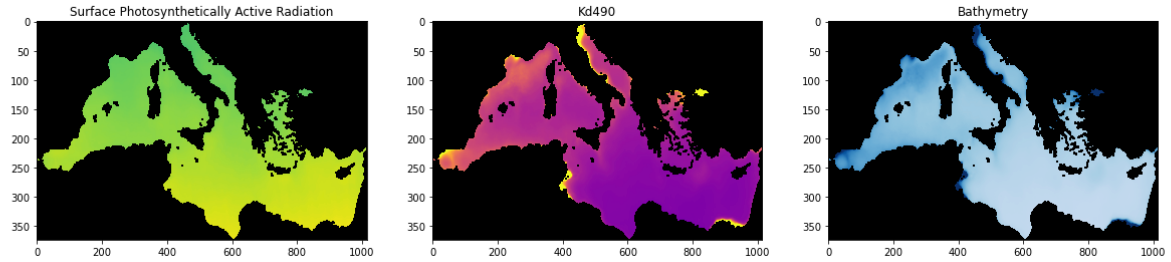
[2]:
```
par_18 = gdal.Open("NASA_2002_2018_avg_par.tif")
par_18_band = par_18.GetRasterBand(1)
par_18_array = par_18_band.ReadAsArray()

kd490_18 = gdal.Open('NASA_avg_2002-2017_kd490.tif')
kd490_18_band = kd490_18.GetRasterBand(1)
kd490_18_array = kd490_18_band.ReadAsArray()

bathy_med = gdal.Open('EMODnet_bathymetry_rev.tif')
bathy_med_band = bathy_med.GetRasterBand(1)
bathy_med_array = bathy_med_band.ReadAsArray()
```

Set null data

[ ]:
```
par_18_array[par_18_array<0]=np.nan
kd490_18_array[kd490_18_array<0]=np.nan
bathy_med_array[bathy_med_array<0]=np.nan
```

Plot the data

[214]:
```
pyplot.figure(figsize=(20,4))
pyplot.subplot(1,3,1)
cmap1 = plt.cm.get_cmap("viridis").copy()
cmap1.set_bad(color='black')
pyplot.imshow(par_18_array, vmin=0, vmax=43, cmap=cmap1, aspect='auto')
plt.title("Surface Photosynthetically Active Radiation")
pyplot.subplot(1,3,2)
cmap2 = plt.cm.get_cmap("plasma").copy()
cmap2.set_bad(color='black')
pyplot.imshow(kd490_18_array, vmin=0, vmax=0.1, cmap=cmap2, aspect='auto')
plt.title("Kd490")
pyplot.subplot(1,3,3)
cmap3 = plt.cm.get_cmap("Blues").copy()
cmap3.set_bad(color='black')
pyplot.imshow(kd490_18_array, vmin=0, vmax=0.1, cmap=cmap3, aspect='auto')
plt.title("Bathymetry")
pyplot.show()
```



Reliance

Research Lifecycle Management technologies for Earth Science Communities and Copernicus users in EOSC

RoHub

Thanks!

Raul Palma
rpalma@man.poznan.pl

Useful Links:

ROHub in EOSC marketplace: https://marketplace.eosc-portal.eu/services/psnc.rohub

ROHub portal https://reliance.rohub.org/

ROHub tutorial: https://reliance-eosc.github.io/ROHUB-API_documentation/html/tutorials.html

ROHub portal documentation: https://reliance-eosc.github.io/rohub-portal-documentation/

ROHub API library documentation : https://reliance-eosc.github.io/ROHUB-API_documentation/html/index.html

ROHub API library example Jupyter Notebooks: https://github.com/RELIANCE-EOSC/sample-notebooks

ROHub helpdesk: https://support.pcss.pl/servicedesk/customer/portal/27 or support email:support@rohub.org

This project has received funding from the European research infrastructures (including e-Infrastructures) *under the European Union's Horizon 2020 research and innovation programme* under grant agreement No 101017501