INDIGO-DataCloud

INITIAL REQUIREMENTS FROM RESEARCH COMMUNITIES ANNEX 1.P24(INGV/EMSO): SELECTED CASE STUDY FROM MOIST - MULTIDISCIPLINARY OCEANIC INFORMATION SYSTEM

INPUT TO EU DELIVERABLE: D 2.1

Document identifier:	INDIGO-WP2-D2.1-ANNEX-1P0-V7
Date:	**27/05/2015**
Activity:	**WP2**
Lead Partner:	**EGI.eu**
Document Status:	**DRAFT**
Dissemination Level:	**CONFIDENTIAL (INTERNAL)**
Document Link:	

Abstract

This report summarizes the findings of T2.1 and T2.2 **for partner P24** along the first three months of the project. It is an integrated document including a general description of the research communities involved and the selected Case Studies proposed, in order to prepare deliverable D2.1, where the requirements captured will be prioritized and grouped by technical areas (Cloud, HPC, Grid, Data management, etc). The report includes an analysis of DMP (Data Management Plans) and data lifecycle documentation aiming to identify synergies and gaps among different communities.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	Partner/Activity	Date
From	Massimiliano Rossi, Laura Beranzoli **/editor in Px**>>	**P24**/WP2	June 3rd, 2015
Reviewed by	**Moderators:** P.Solagna, F.Aguilar, J.Marco **Internal Reviewers:** <<To be completed by project office on submission to PMB>>		
Approved by	**PMB** <<To be completed by project office (no submission)>>		

III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	5-may-2015	First draft, v01	J.Marco, F.Aguilar CSIC
2	7-may-2015	Initial feedback on structure from all partners	F.Aguilar CSIC, A.Bonvin UUtrecht
3	18-may-2015	Draft discussed in f2f meeting in Lisbon	P.Solagna, EGI.eu F.Aguilar, CSIC
4-7	28-may-2015	Draft ready for initial community input, to be iterated with JRA, v07	P.Solagna, EGI.eu J.Marco, F.Aguilar, CSIC, I.Blanquer UPV
8	4-june-2015	Draft after input from community, v08	JRA?
9	7-june-2015	Draft revised also with JRA, v09	P.Solagna, EGI.eu F.Aguilar, CSIC
10	10-june-2015	Draft to be circulated for internal review, v10	P.Solagna, EGI.eu
11	20-june-2015	Comments included, version for release v11	P.Solagna, EGI.eu

TABLE OF CONTENTS

0 INTRODUCTION AND CONVENTIONS

PLEASE, READ CAREFULLY BEFORE COMPLETING THE ANNEX:

*This Annex is an example of compilation of the information needed to support adequately a **Case Study** of interest in a Research Community. Each partner in INDIGO WP2 is expected to provide such information along the first three months of the project (i.e. by June 2015), and it will be used to compile Deliverable D2.1 on Initial Requirements from Research Communities.*

There will be around 10 Annexes, for example Annex 1.P1 for partner 1 in WP2 (i.e. UPV), will cover Case Studies from EuroBioImaging research community.

The initial version will be discussed with INDIGO Architectural team to agree on a list of requirements.

Some relevant definitions:

***A Case Study** is an implementation of a research method involving an up-close, in-depth, and detailed examination of a subject of study (the case), as well as its related contextual conditions.*

Comentario [Laura B1]: Sfruttare emso per I daTI E INDIGO PER I SAAS E PAAS PER I RICERCATORI

We should focus on Case Studies that are representative both of the research challenge and complexity but also of the possibilities offered by INDIGO-DataCloud solutions on it!

*The Case Study will be based on a set of User Stories, i.e. how the researcher describes the steps to solve each part of the problem addressed. **User Stories** are the starting point of **Use Cases**, where they are transformed into a description using software engineering terms (like the actors, scenario, preconditions, etc). Use Cases are useful to capture the Requirements that will be handled by the INDIGO software developed in JRA workpackages, and tracked by the Backlog system from the OpenProject tool.*

The User Stories are built by interacting with the users, and a good way is to do it in three steps (CCC): Card, Conversation and Confirmation[1].

Use Cases can benefit from tools like "mock-up" systems where the user can describe virtually the set of actions that implement the User Story (i.e. by clicking or similar on a graphical tool).

Different parts of this document should be completed with the help/input of different people:

RESEARCH MANAGERS

-Section 1, SUMMARY, is to be reviewed/agreed with them as much as possible

RESEARCHERS

*-Section 2, INTRODUCTION is designed to be filled with direct input from (senior) researchers describing the interest of the application, and written in such a way that it can be included in related technical papers. It is likely that such introduction is already available for some communities (for example, for several research communities in WP2 like DARIAH, CTA,EMSO, Structural Biology, one may start from the **Compendium of e-Infrastructure requirements for the digital ERA[2] from EGI***

APPLICATION DEVELOPERS AND INTEGRATORS WITHIN THE RESEARCH COMMUNITIES

-Sections 3, 4, 5, 6: should be discussed from their technical point of view (including data management as much as possible).

MIDDLEWARE DEVELOPERS AND E-INFRASTRUCTURE MANAGERS

-Sections 7, 8: should be discussed with them

[1] For a nice intro, see: https://whyarerequirementssohard.wordpress.com/2013/10/08/when-to-use-user-stories-use-cases-and-ieee-830-part-1/ , and also https://whyarerequirementssohard.wordpress.com/2015/02/12/how-do-we-write-good-user-stories/ etc.

[2] *https://documents.egi.eu/public/ShowDocument?docid=2480*

The logical order to fill the sections is: 2,3,4,5,6,1,7,8. Sections 1 and 8 will go into deliverable D2.1.

Other conventions and instructions for this document:

As this document/template is to be reused, the convention to use it as a questionnaire is that:

1) -text in italics provides its structure and questions,

2) -input/content should be written using normal text, replacing **<input here>**

Also the following conventions are used to identify the purpose of some parts of the questionnaire:

Bold text in blue corresponds to indications/suggestions to complete the questionnaire

Bold text in dark red marks technical issues particularly relevant that should be carefully considered for further analysis of requirements

Text in red indicates pending issues or ad-hoc warnings to the reader

1 EXECUTIVE SUMMARY ON THE CASE STUDY

Summarize the research community applications/plans/priorities (max length 2 pages).
To be completed after section 2 and reviewed later. Supervision by a senior researcher is required.

1.1 Identification

- *Community Name*: *Multidisciplinary Oceanic Information SysTem (MOIST) i.e. Earth Science*
- *Institution/partner representing the community in INDIGO:* **INGV**
- *Main contact person*: **Laura Beranzoli, Massimiliano Rossi**
- *Contact email:* **laura.beranzoli@ingv.it, massimiliano.rossi@ingv.it**
- *Specific Title for the Case Study:* **MOIST**

1.2 Brief description of the Case Study and associated research challenge

Please include also a brief description of the community regarding this Case Study: partners collaborating, legal framework, related projects, etc.

MOIST is a data management system for multi-parametric observatories focused on standards, open accessibility and web services.

This web interface allows to discover, visualize and download the metadata and data related to seafloor campaigns with GEOSTAR-type observatories from 1998 to present.

The datasets discovery can be done through the navigation menu (in particular by sites, projects, instruments)

Describe the research/scientific challenge that the community is addressing in the Case Study

<input here>

1.3 Expectations in the framework of the INDIGO-DataCloud project

What do you think could be your main objectives to be achieved within the INDIGO project in relation to this Case Study?

Enhance the system to cloud with increasing of general performance

1.4 Expected results and derived impact

Describe the research results and impact associated to this Case Study.

<input here> ?

1.5 References useful to understand the Case Study

Include previous reports, articles, and also presentations describing the Case Study

- **De Caro, M., et al. "Seafloor Seismic Noise at Central Eastern Mediterranean Sites." Seismological Research Letters 85.5 (2014): 1019-1033.,DOI: 10.1785/0220130203**
  SITE: Marmara Sea
- **Sgroi, T., et al. "Geohazards in the Western Ionian Sea: Insights from non-earthquake signals recorded by the NEMO-SN1 seafloor observatory". Oceanography 27(2):154–166, doi: 10.5670/oceanog.2014.51.**
  CAMPAIGN: SMO 1
- **Embriaco, D., Marinaro, G., Frugoni, F., Monna, S., Etiope, G., Gasperini, L., ... & Favali, P. (2014). Monitoring of gas and seismic energy release by multiparametric benthic observatory along the North Anatolian Fault in the Sea of Marmara (NW Turkey). Geophysical Journal International, 196(2), 850-866.**
  CAMPAIGN: MARMARA-DM
- **Etiope, G., et al. "Offshore and onshore seepage of thermogenic gas at Katakolo Bay (Western Greece)." Chemical Geology, DOI: 10.1016/j.chemgeo.2012.08.011**
  CAMPAIGN: HYPOX 1
- **Favali, P., et al. "NEMO-SN1 Abyssal Cabled Observatory in the Western Ionian Sea." IEEE journal of oceanic engineering (2012).,DOI: 10.1109/JOE.2012.2224536**
  CAMPAIGN: SMO 1
- **Gasperini, L., et al. "Cold seeps, active faults and the earthquake cycle along the North Anatolian Fault system in the Sea of Marmara (NE Turkey)". Bollettino di Geofisica Teorica ed Applicata Vol. 53, n.4, 2012**
  SITE: Marmara Sea
- **Gasperini, L., et al. "Gas seepage and seismogenic structures along the North-Anatolian Fault in the eastern Sea of Marmara." Geochemistry, Geophysics, Geosystems, DOI: 10.1029/2012GC004190**
  SITE: Marmara Sea
- **Favali, Paolo, et al. "NEMO-SN1 observatory developments in view of the European Research Infrastructures EMSO and KM3NET." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 626 (2011): S53-S56., DOI: 10.1016/j.nima.2010.04.139**
- **Geissler, W. H., et al. (2010), Focal mechanisms for sub-crustal earthquakes in the Gulf of Cadiz from a dense OBS deployment, Geophys. Res. Lett., 37, L18309, doi:10.1029/2010GL044289.**
  DATASET: NEAREST 1 - Seismometer
- **Gasperini, L., et al. "The entrance of the Izmit Gulf: a key site for monitoring gas emissions and seismicity in the Sea of Marmara." EGU General Assembly Conference Abstracts. Vol. 12. 2010.**
  SITE: Marmara Sea
- **Embriaco D, Giovanetti G, Marinaro G (2010). Geophysical Data Acquisition and Transmission for the SN-1 observatory to be deployed off shore Catania. In: Rapporti Tecnici INGV CAMPAIGN: SMO 1**

2 INTRODUCTION TO THE RESEARCH CASE STUDY

Summarize the Case Study from the point of view of the researchers (max length 3 pages + table). Input by the research team in the community addressing the Case Study is required.

The scientific management of time-increasing quantity of data has recently become a big challenge in terms of storage capacity, preservation, interoperability and data access in many disciplinary sectors.

In the environmental Science sectors, the analysis of large amount of time series (sustained measurements over time) is considered necessary for any predicting modelling in reply to urgent questions on global changes at different space and time scales.

As an example, the development and use of multiparameter seafloor and water column observatories enabling a multidisciplinary approach to investigate the deep sea processes with different time scales (from seconds to decades), has posed the need to collect, organise and maintain in the long-term a variety of long time series.

MOIST, Multidisciplinary Oceanic Information SysTem, is presently a data and metadata provider initiated within the ESONET NoE project and under implementation and development as part of EMSO (European Multidisciplinary Seafloor and water column Observatory) also in link with funded projects such as EC Genesi-DEC, EC SCIDIP-es, EC ENVRI, EC CoopEUS, EC ENVRI-Plus, EC EMSODEV.

MOIST is designed to make available scientists and users multidisciplinary data obtained by means of fixed-point observatories managed by INGV in some EMSO key-sites. The MOIST configuration underpins the observatory data flow from the sensor acquisition to the dissemination.

MOIST is developed by adopting the most common data standards (e.g., OGC, NASA, INSPIRE) organising, indexing and converting the data into a unique data scheme and supports some EMSO observatories node regardless of their specific suite of sensors and sensor configuration and operational status.

Visualisation functions are implemented to inspect and qualitatively compare the time series of different parameters and/or sensors.

Raw data quality check tools are going to be designed to enrich the metadata with additional information about data completeness, consistency and coherence.

MOIST represents a significant example of local data management systems of a typical EMSO observatory node.

Comentario [Laura B2]: A special attention devoted to all standardisation aspects in terms of f formats, metadata, interoperability, transport protocols and controlled vocabularies for keyword and parameters…
This guarantees the quality, completeness and availability for different sciences, anticipating the future by initiating a long-term data preservation strategy.

2.1 Presentation of the Case Study

Describe the Case Study from the research point of view

Presently MOIST relies upon servers located in INGV (Rome) with the following disadvantages:

- **Limited bandwidth for a smooth and quick access: a 600 Mb band width is used for about 100 servers storage devices and users from INGV;**
- **Limited storage capacity**

- **Poor redundancy and exposure to risk of loss of data**
- **Mirror facilities unavailable yet**
- **Convergence layer unavailable yet**
- **Lack of a federated authentication and authorisation infrastructure (AAI)**

The proposed use case shall implement appropriate solutions for overcoming these limitations exploiting the benefits of a federated DATA CLOUD approach. In particular

- **Very high connection throughput**
- **High-availability of whole systems**
- **Increased cyber security**

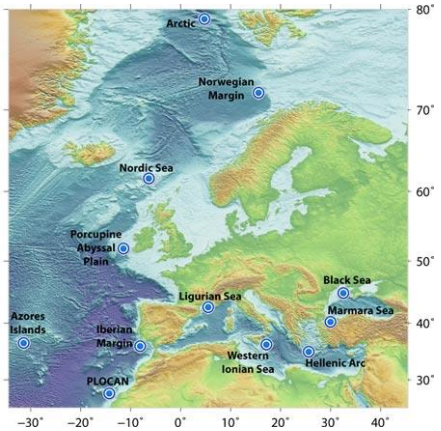
Through the use case MOIST will become compliant with state-of-art technology.

2.2 Description of the research community including the different roles

Please include a description of the scientific and technical profiles, and detail their institutions

Describe the research community specifically involved in this Case Study

EMSO is a large-scale European Research Infrastructure in the field of marine environmental sciences supported by EC through ESFRI- European Strategy Forum on Research Infrastructures and by 10 Member States. EMSO is based on a European-scale network of fixed-point seafloor and water column observatories with the basic scientific objective of long-term monitoring, mainly in real-time, of processes related to the interaction between the geosphere, biosphere, and hydrosphere, including natural hazards. EMSO is a geographically distributed infrastructure composed of several deep-sea monitoring systems deployed on specific sites in the European seas, from the Arctic to the Black Sea through the Atlantic Ocean and the Mediterranean, thus forming a widely distributed pan-European infrastructure. The map below shows the current location of EMSO observatory nodes around Europe.



List of EMSO Observatory nodes

- Arctic
- Celtic/Porcupine
- Azores Islands
- PLOCAN
- Norwegian Margin
- Iberian Margin
- Galway Bay, Irish West Coast
- Iroise Sea - Molene Island
- Ligurian Sea
- Western Ionian
- Hellenic Arc
- Marmara Sea
- Black Sea
- OBSEA
- Koljoe Fjord

The research community addressed by EMSO includes scientists belonging to very diverse marine environmental sectors (geophysicists, oceanographers, marine-biologists, ecologists, bio-geo-chemists, etc) with interest in revealing transient signals and trends from long time series of local and regional processes at nodes. Recovery of homogeneous time-series from a number EMSO nodes referred to the same time period is of interest for global process modelling. In addition, single data user access to EMSO nodes is expected to follow-up experiments and sensor testing.

Comentario [Laura B3]: Usare come case st

EMSO infrastructure is on the way to get a legal status and will be managed by a European Consortium, EMSO ERIC, as per EC regulations.

2.3 Current Status and Plan for this Case Study

Please indicate if the Case Study is already implemented or if it is at design phase.

Describe the status of the Case Study and its short/mid term evolution expected

The MOIST Case Study is presently operational and exposes data from some EMSO nodes of the Mediterranean and Atlantic Ocean. MOIST will undergo to enhancements within the EMSODEV EC project to become part of a larger integrated Data Management Plan for the whole EMSO infrastructure.

2.4 Identification of the KEY Scientific and Technological (S/T) requirements

Please try to identify what are the requirements that could make a difference on this Case Study (thanks to using INDIGO solutions in the future) and that are not solved by now.

Indicate which are the KEY S/T requirements from your point of view

Approach to the distributed storage and network through

- Very high connection throughput
- High-availability of whole systems

- **Increased cyber security**

2.5 General description of e-Infrastructure use

Please indicate if the current solution is already using an e-Infrastructure (like GEANT, EGI, PRACE, EUDAT, a Cloud provider, etc.) and if so what middleware is used. If relevant, detail which centres support it and what level of resources are used (in terms of million-hours of CPU, Terabytes of storage, network bandwidth, etc.) from the point of view of the research community.

Detail e-Infrastructure resources being used or planned to be used.

MOIST relies upon local INGV servers and on a commercial network provider and in the next future will switch to GARR (the Italian Education and Research Network).

At the moment external resources are not used.

2.6 Description of stakeholders and potential exploitation

Please summarize the potential stakeholders (public, private, international, etc.) and relate them with the exploitation possibilities. Provide also a realistic input to table on KPI.

Describe the exploitation plans related to this Case Study

EMSO ERIC is going to be established to serve a wide variety of user groups and stakeholders in both the public and private sectors and in multiple scientific, technological, educational and policy-making fields, including:

- ***national and international scientific communities*** **- marine researchers and scientists across numerous fields are EMSO ERIC prime clients (see below);**
- ***large and small enterprises*** **whose activities are directly or indirectly related to the use of the sea as a resource, including:**
  - **oil & gas companies;**
  - **renewable energy plants/industries - wind, tidal and wave energy, geothermal;**
  - **medical and pharmaceutical companies**
  - **shipping and maritime transportation companies and regulators**
  - **insurance companies**
  - **mine & mineral industries (increasing demand for sources of critical rare minerals and materials to support a variety of hi-tech sectors, most notably telecoms, security and defence;**
  - **underwater sensors and instrumentation developers and manufacturers;**
  - **industrial organizations and advanced materials companies and associations active in carrying out specific tests (testing services at high pressures: e.g. testing of materials);**
- ***public administration organisations*** **responsible for monitoring the (marine) environment;**

- ***international environmental organisations*, public interest groups and NGOs active in ecology issues, climate change, environmental monitoring (pollution, contamination);**
- ***international media*: as corroborating reference authority on marine science matters;**
- ***geo-hazards monitoring and prevention organizations* - disaster management and mitigation, tsunami and seismic risk, metoceaneo- and climate change modelling and forecasting;**
- ***port authorities* and other bodies responsible for monitoring and security;**
- ***tourist industries* for whom maintaining the quality of waters is paramount;**
- ***fisheries* resources, including stock assessment, changes in distribution, response to stressors3;**
- ***ocean litter clean-up companies and organizations*;**
- ***defence industry* - war and military litter and debris detection and disposal;**
- ***academia* and developers/suppliers/users of education & training content and aides;**
- ***the general public* ever-more enabled, eager and active in pursuing personal-development, life-long-learning**

Please indicate (as realistic as possible) the expected impact for each topic in the following table:

Area	*Impact Description*	*KPI Values*
Access	*Increased access and usage of e-Infrastructures by scientific communities, simplifying the "embracing" of e-Science.*	• *Number of ESFRI or similar initiatives adopting advanced middleware solutions* ESFRIs: EMSO is expected to use • *Number of production sites supporting the software* **3 EMSO nodes are exopected**
Usability	*More direct access to state-of-the art resources, reduction of the learning curve. It should include analysis platforms like R-Studio, PROOF, and Octave/Matlab, Mathematica, or Web/Portal workflows like Galaxy.* *Use of virtualized GPU or interconnection (containers).* *Implementation of elastic scheduling on IaaS platforms.*	• *Number of production sites running INDIGO-based solutions to provide virtual access to GPUs or low latency interconnections* **3 EMSO nodes are exopected** • *Number/List of production sites providing support for Cloud elastic scheduling* **Not applicable** • *Number of popular applications used by the user communities directly integrated with the project products:* **3** • *Number of research communities using the developed Science Gateway and Mobile Apps:* **EMSO is expected to use** • *Research Communities external to INDIGO using the software products*: **EMSO is expected to use**
Impact on Policy	*Policy impact depends on the successful generation and*	• *Number of contributions to roadmaps, discussion papers:* **For EMSO: 4 (2european,+ 2 national)**

	dissemination of relevant knowledge that can be used for policy formulation at the EU, or national level.	
Visibility	*Visibility of the project among scientists, technology providers and resource managers at high level.*	• *Number of press releases issued*: **1** • *Number of download of software from repository per year:* **Not Applicable** • *List of potential events/conferences/workshops:* **3 for EMSO** • *Number of domain exhibitions attended* **4 for EMSO** • *Number of communities and stakeholders contacted* **3 for EMSO**
Knowledge Impact	*Knowledge impact creation: The impact on knowledge creation and dissemination of knowledge generated in the project depends on a high level of activity in dissemination to the proper groups.*	• *Number of journal publications:* **2 for EMSO** • *Number of conference papers and presentations*: **2 for EMSO**

Table 1 Key Performance Indicators (KPI) associated to different areas. Add in this table how your community would contribute to the KPIs. Note: this table will NOT be included in the deliverable.

3 TECHNICAL DESCRIPTION OF THE CASE STUDY

Describe the Case Study from the point of view of developers (4 pages max.)
Assemble it using preferably an AGILE scheme based on User Stories.

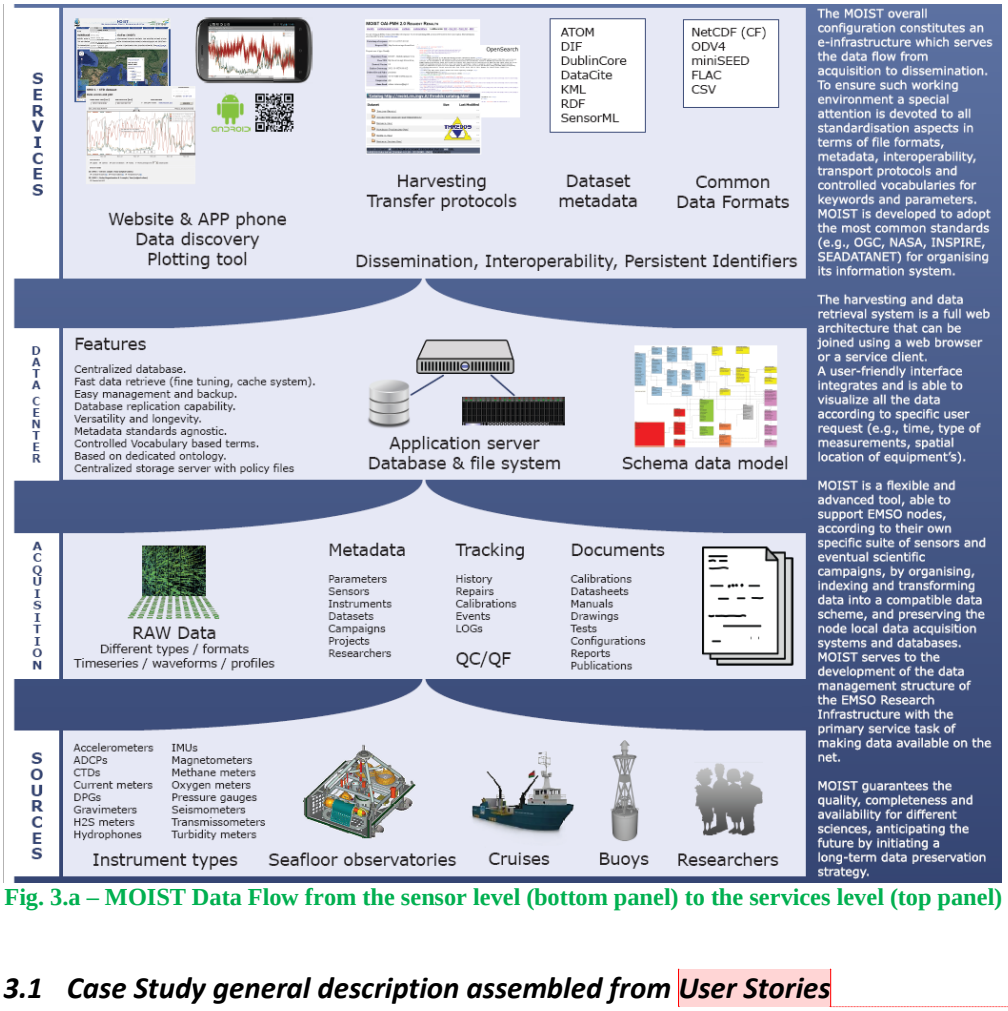


Fig. 3.a – MOIST Data Flow from the sensor level (bottom panel) to the services level (top panel)

3.1 Case Study general description assembled from User Stories

Please describe here globally the Case Study. If possible use as input "generic" User Stories built according to the scheme: short-description (that fits in a "card") + longer description (after "conversation" with the research community). Provide links to presentations in different workshops describing the Case Study when available. Include schemes as necessary.

Comentario [Laura B4]: Refer to pag 13

Comentario [Laura B5]: It's important to remember the famous 3 Cs of a user story: Ca Conversation and Confirmation.
▪The **card** is the thing you use to track the status of the user story. Using a card, rather than a spreadsheet or word processing tool, is also useful because its size limits the amount you can write.
▪**Conversation**, which should be plural in m mind, because you're going to have many conversations about the user story. You're goi to have an initial conversation when it's creat more conversations when you estimate it, prioritize it and break it into smaller stories, a then even more conversations in the sprint planning meeting, the sprint review meeting, and throughout the sprint as you're satisfying However, the detail of these conversations are not recorded on the card, other than any acceptance criteria that are uncovered.
▪**Confirmations** are the criteria you're using to assess whether the user story has been met These are written as they discovered, and can roughly approximated as acceptance tests. These are one of the means that the team uses to evaluate whether the user story is done or not, and how the customer measures the succ of the sprint.

*Describe the **Case Study showing the different actors** and the basic components (data, computing resources, network resources, workflow, etc.). Reference relevant documentation.*

Seafloor instability resulting from both tectonics and volcanism affects the Western Ionian Sea. The EMSO observatory in the Western Ionian Sea, NEMO-SN1, deployed 25 km offshore eastern Sicily at 2100 m water depth, records a variety of geophysical and environmental signals over the long-term, including non-earthquake seismic signals that can also provide hints for geohazards assessment. A 28 km long electro-optical cable from a shore laboratory managed by INFN-LNS in the Catania's harbor to NEMO-SN1, provides power and bi-directional real-time communication for the whole observatory. Being NEMO-SN1 a cabled seafloor observatory, all the data collected by the instruments are transferred in real time to an onshore laboratory. Users can access the real-time diagnostic information on the observatory and send commands to the observatory control system as well as to each device. It is also possible to reconfigure the observatory mission by switching each instrument on and off and to change the sampling frequencies and other operation parameters of the different devices.

A representative user story for MOIST is about the study of Etna Mount dynamics located nearby NEMO-SN1 and includes the analysis of:

- seismic signals associated with submarine landslides caused by slope instabilities of the volcano's submerged flanks, moreover tectonic shifts can lead to submarine landslides.
- volcanic tremor signals, it is the result of sustained pressure fluctuations, probably related to stress variations induced by magma movement
- short duration ground-motion events (SDEs).

These signals are analyzed together with water pressure data acquired by pressure gauges and hydrophones.

Increased tremor amplitudes recorded at NEMO-SN1 during lava fountain episodes suggest the presence of an east-southeast offshore location of the roots of Mt. Etna's magma feeding system. SDEs are thought to result from hydro-fracturing of carbonate outcroppings at the base of the Malta Escarpment, an important North-South tectonic structure south of Mt. Etna, that is possibly induced by changes in the stress field associated with magma movement.

The user story is implemented making use of R software in order to evaluate cumulative energy released by the different studied processes.

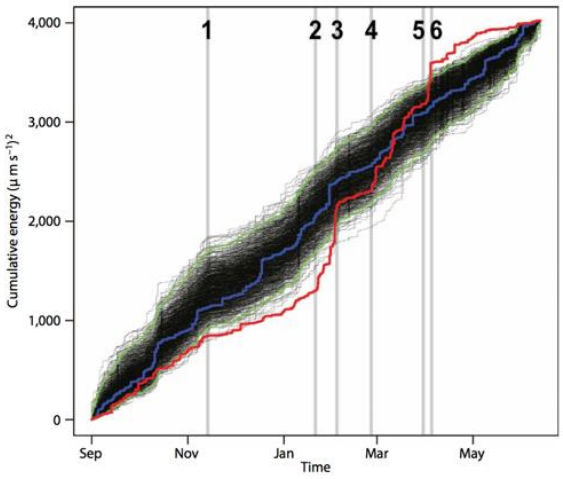


Figure 3.1.a - Statistical distribution function evaluated using *R software*: the plot shows the cumulative sum of the SDE energy over nine months (red curve) compared to a set of 1,000 randomly generated energy cumulative sums (black curves).

The use of the *R software* allowed show that the accumulated SDE energy is the result of some driving mechanism and not of a random process. For the sake of completeness, the observation period can be divided into two parts by comparing the slope of the observed cumulative sum (red curve) and of the median cumulative sum (blue curve) of the random processes: the first part up to point 2, and the second part between points 2 and 6.

3.2 User categories and roles

Describe in more detail the different user categories in the Case Study and their roles, considering in particular potential issues (on authorization, identification, access, etc.)

Comentario [MR6]: Seems we don't have this structure at the moment

MOIST access does not assume any user category yet. However the following user categories will be surely implemented:

- **Campaign manager: responsible for collecting requirements of the scientific scenario including expected equipment's and data;**
- **Generic scientific end-user interested in accessing and downloading selected data according to specific requirements.**

3.3 General description of datasets/information used

List the main datasets and information services used (details will be provided in next section)

Please refer to Fig.3.a

3.4 Identification of the different Use Cases and related Services

Identify initial Use Cases based on User Stories, and describe related (central/distributed) Services

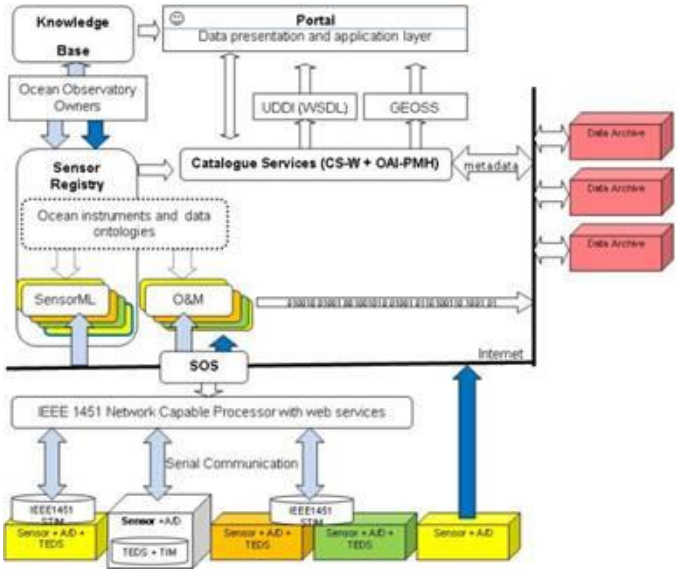
With reference to Fig.3.a, the following Services are operational in MOIST:

- **Website and mobile App**
- **Data discovery**
- **Plotting tools**
- **Standard Data Formats**

3.5 Description of the Case Study in terms of Workflows

Summarize the different Workflows within the Case Study, and in particular Dataflows. Include the interaction between Services.

EMSO Case Study workflow is outlined in figure 3.5.a



3.6 Deployment scenario and relevance of Network/Storage/HTC/HPC

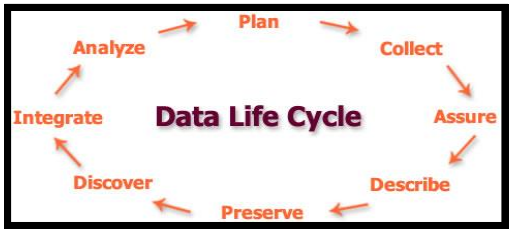
Indicate the current deployment framework (cluster, Grid, Cloud, Supercomputer, public or private) and the relevance for the different Use Cases of the access to those resources.

EMSO does not make use of any scenario yet but the use of Cloud is really appropriate given the distributed nature of the infrastructure.

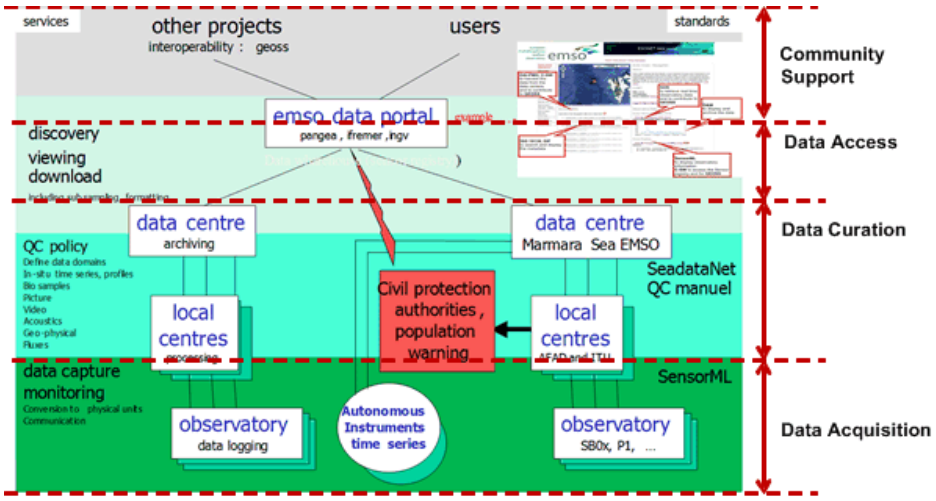
4 DATA LIFE CYCLE

INDIGO-DataCloud is a DATA oriented project. So the details provided in this complex section are KEY to the project. Please try to be as complete as possible with the relevant information.

Using the DataONE scheme, shown below, the different stages in the data life cycle are considered under the perspective of preparation of a DMP (Data Management Plan) following the recommendations of the UK DCC and H2020 guidelines.



EMSO detailed data life cycle is shown below



BEFORE FILLING NEXT SECTIONS, CONSIDER CONSULTING:
https://www.dataone.org/all-best-practices-download-pdf and https://dmponline.dcc.ac.uk/

4.1 Data Management Plan (DMP) for this Case Study

According to EU H2020 indications[4], following UK DCC tool indications

The EMSO DMP is going to be deployed in EMSODEV project which runs in parallel to INDIGO.

4.1.1 Identification of the DMP

Plan identification: *<Code, ID>* **To be Defined**

Associated grants: <Funded Projects, other grants> **EC EMSODEV project**

Principal Researcher: **Paolo Favali (coordinator)**

DMP Manager: **To be Defined (provisional: Massimiliano Rossi)**

Description:

the DMP will include:

- **Roles and Responsibilities**
- **Types of Data**
- **Data Formats and Metadata**
- **Access, Sharing and Privacy**
- **Policies and Provisions for Re-use & Re-distribution**
- **Data Storage and Preservation**
- **Costs**
- **DMPTool: Create ready-to-use plans**

[4] *In Horizon 2020 a limited pilot action on open access to research data will be implemented. Projects participating in the Open Research Data Pilot will be required to develop a Data Management Plan (DMP), in which they will specify what data will be open. Other projects are invited to submit a Data Management Plan if relevant for their planned research. The DMP is not a fixed document; it evolves and gains more precision and substance during the lifespan of the project. The first version of the DMP is expected to be delivered within the first 6 months of the project. More elaborated versions of the DMP can be delivered at later stages of the project. The DMP would need to be updated at least by the mid-term and final review to fine-tune it to the data generated and the uses identified by the consortium since not all data or potential uses are clear from the start. The templates provided for each phase are based on the annexes provided in the Guidelines on Data Management in Horizon 2020 (v.1.0, 11 December 2013).*

4.1.2 DMP at initial stage (to be prepared before data collection)

The DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

Not yet available

For each data set provide:

Description of the data that will be generated or collected; indicate its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

Data set reference and name < **Not yet available**

>

Data set description < **Not yet available**

>

Standards and metadata < **Not yet available**

>

Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created (see also below).

Connection to Instrumentation,

Sensors, Metadata, Calibration, etc (pending definitive form, see next sections)

< **Not yet available**

>

Vocabularies and Ontologies

Are they relevant? Internal vocabularies related to the specific fields. RDA groups.

(pending definitive form, see next sections)

< **Not yet available** >

Data Capture Methods

Outline how the data will be collected / generated and which community data standards (if any) will be used at this stage. Indicate how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier for the research team to find, understand and reuse.

• *How will the data be created?* **Sensors output**

• *What standards or methodologies will you use?* **EMSO internal Best Practices**

• *How will you structure and name your folders and files?* **Time and space (geographical coordinates)**

• *How will you ensure that different versions of a dataset are easily identifiable?* **Not yet available**

Metadata

Metadata should be created to describe the data and aid discovery. Consider how you will capture this information and where it will be recorded e.g. in a database with links to each item, in a

'readme' text file, in file headers etc. Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. The UK Data Curation Center offers a catalogue of disciplinary metadata standards.

• *How will you capture / create the metadata?* **< Not yet available >**

• *Can any of this information be created automatically?* **< Not yet available >**

• *What metadata standards will you use and why?* **< Not yet available>**

Data sharing

Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.). In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

<input here> Trought data portal

Method for Data Sharing

Consider where, how, and to whom the data should be made available. Will you share data via a data repository, handle data requests directly or use another mechanism? The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Mention earlier examples to show a track record of effective data sharing.

• *How will you make the data available to others?* **< Not yet available >**

• *With whom will you share the data, and under what conditions?* **< Not yet available >**

Restrictions on Sharing

Comentario [MR7]: Non abbiamo un data acc policy

Outline any expected difficulties in data sharing, along with causes and possible measures to overcome these. Restrictions to data sharing may be due to participant confidentiality, consent agreements or IPR. Strategies to limit restrictions may include: anonymising or aggregating data; gaining participant consent for data sharing; gaining copyright permissions; and agreeing a limited embargo period.

• *Are any restrictions on data sharing required? e.g. limits on who can use the data, when and for what purpose.* **NA**

• *What restrictions are needed and why?* **NA**

• *What action will you take to overcome or minimise restrictions?* **NA**

Data Repository

Most research funders recommend the use of established data repositories, community databases and related initiatives to aid data preservation, sharing and reuse. An international list of data repositories is available via Databib or Re3data.

• *Where (i.e. in which repository) will the data be deposited?* **Present storage on SAN**

Archiving and preservation (including storage and backup)

Questions to consider before answering:

•What is the long-term preservation plan for the dataset? e.g. deposit in a data repository

•Will additional resources be needed to prepare data for deposit or meet charges from data repositories?

Researchers should consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.

• What additional resources are needed to deliver your plan?

• Is additional specialist expertise (or training for existing staff) required?

• Do you have sufficient storage and equipment or do you need to cost in more?

• Will charges be applied by data repositories?

• Have you costed in time and effort to prepare the data for sharing / preservation?

Carefully consider any resources needed to deliver the plan. Where dedicated resources are needed, these should be outlined and justified. Outline any relevant technical expertise, support and training that is likely to be required and how it will be acquired. Provide details and justification for any hardware or software which will be purchased or additional storage and backup costs that may be charged by IT services. Funding should be included to cover any charges applied by data repositories, for example to handle data of exceptional size or complexity. Also remember to cost in time and effort to prepare data for deposit and ensure it is adequately documented to enable reuse. If you are not depositing in a data repository, ensure you have appropriate resources and systems in place to share and preserve the data.

Describe the procedures that will be put in place for long-term preservation of the data.

We use a data storage for primary backup and mass storage and a tool named HAPPI developed in EC SCIDIP-es project. The tool is able to provide the following services:

- **preservation, integrity and identity of digital information,**
- **tracking of transformations/changes carried out on digital information**
- **authenticity assessment**
- **PDI assignment according to OAIS**

Indicate how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered. **We don't have a time scheduled preservation plan at the moment**

4.1.3 DMP at final stage (to be ready when data is available)

*SCIENTIFIC RESEARCH DATA SHOULD BE EASILY **DISCOVERABLE***

Questions to consider:

• How will potential users find out about your data?

• Will you provide metadata online to aid discovery and reuse?

Guidance: Indicate how potential new users can find out about your data and identify whether they could be suitable for their research purposes. For example, you may provide basic discovery metadata online (i.e. the title, author, subjects, keywords and publisher).

Are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. ***Digital Object Identifier****)?* **<NA>**

SCIENTIFIC RESEARCH DATA SHOULD BE **ACCESIBLE**

Questions to consider:

• Who owns the data?

• How will the data be licensed for reuse?

• If you are using third-party data, how do the permissions you have been granted affect licensing?

• Will data sharing be postponed / restricted e.g. to seek patents?

State who will own the copyright and IPR of any new data that you will generate. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. If purchasing or reusing existing data sources, consider how the permissions granted to you affect licensing decisions. Outline any restrictions needed on data sharing e.g. to protect proprietary or patentable data. See the DCC guide: How to license research data.

Are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses? (e.g. licencing framework for research and education, embargo periods, commercial exploitation, etc) **<NA>**

SCIENTIFIC RESEARCH DATA SHOULD BE ***ASSESSABLE*** *AND INTELLIGIBLE*

• What metadata, documentation or other supporting material should accompany the data for it to be interpreted correctly?

• What information needs to be retained to enable the data to be read and interpreted in the future?

Describe the types of documentation that will accompany the data to provide secondary users with any necessary details to prevent misuse, misinterpretation or confusion. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data.

Are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review?, e.g. are the minimal datasets handled together with scientific papers for the purpose of peer review, are data is provided in a way that judgments can be made about their reliability and the competence of those who created them **<NA>**

USABLE *BEYOND THE ORIGINAL PURPOSE FOR WHICH IT WAS COLLECTED*

• What is the long-term preservation plan for the dataset? e.g. deposit in a data repository

• Will additional resources be needed to prepare data for deposit or meet charges from data repositories?

Researchers should consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.

Guidance on Metadata:

• How will you capture / create the metadata?

• Can any of this information be created automatically?

• What metadata standards will you use and why?

Metadata should be created to describe the data and aid discovery. Consider how you will capture this information and where it will be recorded e.g. in a database with links to each item, in a 'readme' text file, in file headers etc.

Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. The DCC offers a catalogue of disciplinary metadata standards.

Are the data and associated software produced and/or used in the project useable by third parties even long time after the collection of the data? e.g. is the data safely stored in certified repositories for long term preservation and curation; is it stored together with the minimum software, metadata and documentation to make it useful; is the data useful for the wider public needs and usable for the likely purposes of non-specialists? **<NA>**

INTEROPERABLE** TO SPECIFIC QUALITY **STANDARDS

• What format will your data be in?

• Why have you chosen to use particular formats?

• Do the chosen formats and software enable sharing and long-term validity of data?

Outline and justify your choice of format e.g. SPSS, Open Document Format, tab-delimited format, MS Excel. Decisions may be based on staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community. Using standardised and interchangeable or open lossless data formats ensures the long-term usability of data?

See the UKDS Guidance on recommended formats

Are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc?, e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins

<NA>

4.2 Data Levels, Data Acquisition, Data Curation, Data Ingestion

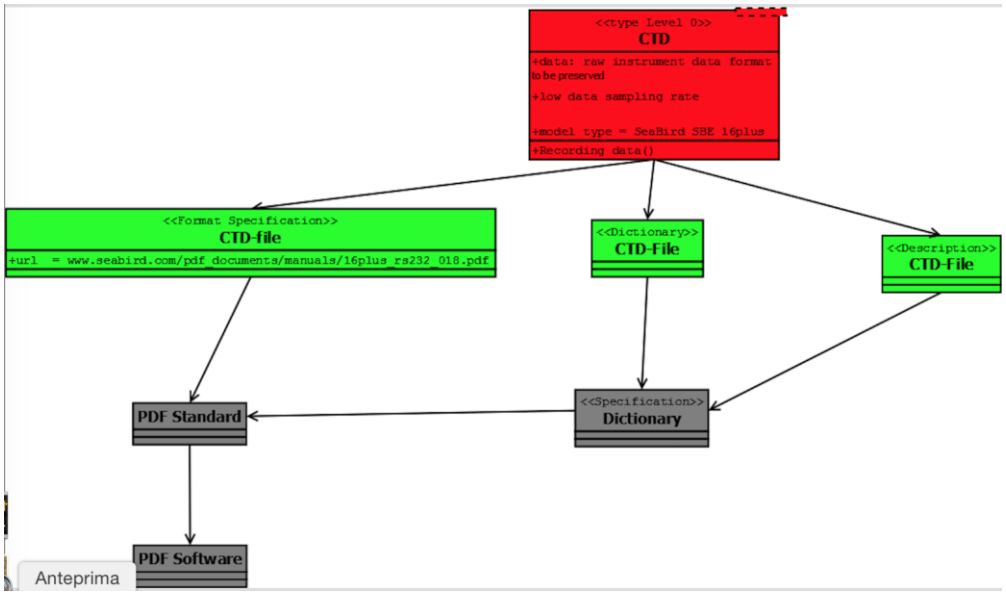
4.2.1 General description of data levels

Indicate if the DATASETS are organized into different levels (LEVEL-0, 1, 2, 3,4) and if so what are the relevant definitions and how DOI are provided. **<input here>**

4.2.2 Collection/Acquisition

Gathering RAW data

Specify how do you gather/collect your data (e.g. sensors, observations, satellites, etc.)?
MOIST Data Level 0 (Raw Data) is shown in figure 4.2.1.a below

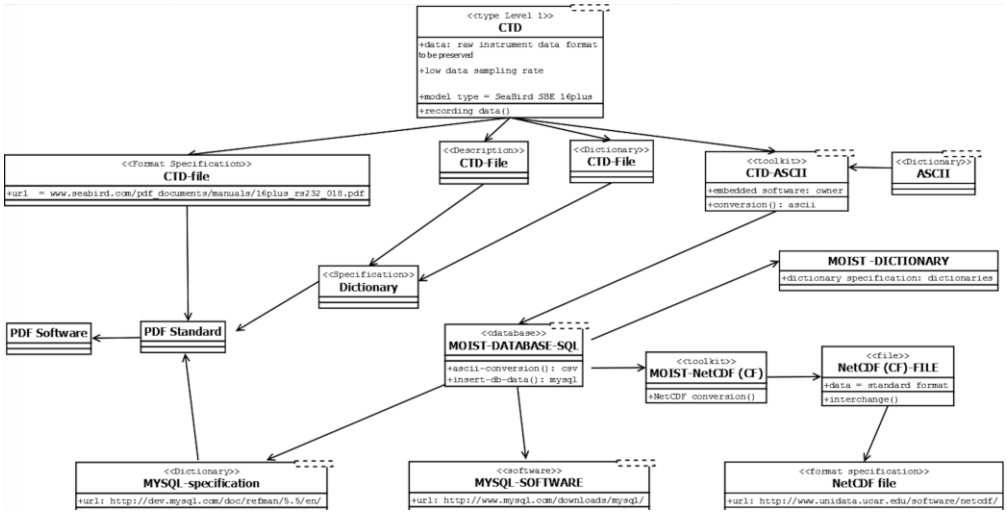


How do you pre-process, transfer and store your RAW data? **<input here>**

From RAW Data to Calibrated Data

Describe the processes applied for Data Calibration, Validation, Filtering, etc.

EMSO Level 1 (from sensor calibration to the data dissemination)



4.2.3 Access to external data

Describe the identification and access to External Data **<input here>www.moist.it**
Indicate if there is a procedure for validation of External Data **<input here>NA**

4.2.4 Data curation

Specify any automatic check applied, like completing series, detecting outlier **<input here>**
Describe manual quality checks **<input here> Visual inspection only**
Are there quality flags applied to the data? **<input here>NA**

4.2.5 Data ingestion / integration

Describe transformations applied to data taking into account ontologies/metadata. Indicate also if there is any "harmonization procedure" (to share/integrate data) and how linking internal and external data is made if relevant. **Ingestion to the Relational Database portal**

4.2.6 Further data processing

Describe, if relevant, the different additional processing steps (and the associated software and resources) applied to the (collected/curated) datasets to provide a "final" dataset collection that can be used in the analysis **<input here> Data series Plot**

4.3 Analysis

4.3.1 Basic analysis and standard analysis suites

Describe usual examples of basic analysis in the Case Study **<NA>**
Specify if software packages/tools like MATLAB, R-Studio, iPython,etc. are used **R Studio**

4.3.2 Data analytics and Big Data

Describe relevant examples of advanced analysis in the Case Study (like for example application of neural networks, series analysis, etc.) **NA at the moment, this issue will be addressed in EMSODEV.**
Specify the resources and additional software required **NA**
Identify analysis challenges that can be classified as "Big Data" **NA**
List Big Data driven workflows **NA**

4.3.3 Data visualization and interactive analysis

Indicate the need for data and analysis results visualization **<Visualisation of multiple signal related to different time series is needed for amultidisciplinary data analysis.**
Indicate how visualization is made and if interactivity/steering is needed **<Interactivity is a basic requirement>**
Specify the User Interfaces (web, desktop, mobile, etc.) < web, desktop, mobile >

4.4 Data Publication

Describe the information flow from the analysis to the publication **<NA>**

Indicate the requirements from publishers/editors to access data, and how it is made available (open data?) **<NA>**

5 SIMULATION/MODELLING

Describe the Simulation/Modelling requirements in this Case Study. Please identify also any other intensive CPU mainly activity as required.

5.1 General description of simulation/modelling needs

Describe the different models used (including references) **<input here>**

*Indicate the type and quantity of simulations needed in the Case Study, and how they are incorporated in the general workflow of the solution***<input here>**

5.2 Technical description of simulation/modelling software

For each simulation package:

Identify the simulation software **<input here>**

Provide a link to its documentation, and describe its maturity and support level **<input here>**

Indicate the requirements of the simulation software (hardware: RAM, processor/cores, extended instruction set, additional software and libraries, etc.) **<input here>**

Tag the simulation software as HTC or HPC **<input here>**

*List the input files required for execution and how to access them***<input here>**

Describe the output files and how they will be stored **<input here>**

Reference an existing installation and performance indicators **<input here>**

Specify if the simulation software is parallelized (or could be adapted) **<input here>**

Specify if the simulation software can exploit GPUs **<input here>**

Specify how the simulation software exploits multicore systems **<input here>**

Specify if parametric runs are required **<input here>**

Estimate the use required of the resources (million-hours, # cores in parallel, job duration, etc) **<input here>**

5.3 Simulation Workflows

Describe if there are workflows combining several (HTC/HPC) simulations or simulations and data processing **<input here>**

6 DETAILED USE CASES FOR RELEVANT USER STORIES

This section tries to put the focus on the preparation of detailed Use Cases starting from User Stories most relevant to the Case Study considered.

6.1 Identification of relevant User Stories

Examples of relevant User Stories linked to roles like for example Final User, Data Curator, etc.

List User Stories based on data collection, curation, processing, analysis, simulation, etc, that are considered most relevant for the Case Study being analyzed **<input here>**

For each relevant User Story:

Draft a basic card **<input here>**

Provide details from conversation with the researchers' teams **<input here>**

Draft as a Use Case **<input here>**

Analyze tools to support the definition of the Use Case (like mockups). Integrate in the analysis the requirements on user interfaces (like the use of mobile resources, under different flavours, access through web interfaces, etc.) **<input here>**

Describe the way to extract requirements and define acceptance criteria **<input here>**

Include if possible an example of support for Big Data driven workflows for e-Science, with requirements for scientific workflows management, under a "Workflow as a Service" model, where the proper workflow engines will be selected according to user needs and requirements.

In such case please describe the scenario for Big Data analysis, and assure that the Use Case considers which levels of workflow engines are needed (e.g., "coarse gran", which targeting distributed (loosely coupled) experiments, through workflow orchestration across heterogeneous set of services; "fine grain", which targeting high performance (tightly coupled) data analysis through workflows orchestration on big data analytics frameworks)

7 INFRASTRUCTURE TECHNICAL REQUIREMENTS

Describe the Case Study from the point of view of the required e-infrastructure support.
INDIGO Data-Cloud will support the use of heterogeneous resources.

7.1 Current e-Infrastructures Resources

Start from the current use of e-infrastructures.

7.1.1 Networking

Describe the current connectivity **<input here>**
Describe the key requirements (availability, bandwidth, latency, privacy, etc) **<input here>**
Specify any current issue (like last mile, or access from commercial, etc) **<input here>**

7.1.2 Computing: Clusters, Grid, Cloud, Supercomputing resources

Describe the current use of each of these type of resources: size and usage **<input here>**
Indicate if there is any mode of "orchestration" between them **<input here>**

7.1.3 Storage

Describe the current resources used **<input here>**
Discuss the key requirements (I/O performance, capacity, availability, reliability, any other QoS indicator) **<input here>**

7.2 Short-Midterm Plans regarding e-Infrastructure use

Plans for next year (2016) and in 5 years (2020).

7.2.1 Networking

Describe the proposed connectivity **<input here>**
Describe new/old key requirements (availability, bandwidth, latency, QoS, private networking, etc) **<input here>**
Specify any potential solution/technique (for example SDN) **<input here>**

7.2.2 Computing: Clusters, Grid, Cloud, Supercomputing resources

Describe the evolution expected: which infrastructures, total "size" and usage **<input here>**
Detail potential "orchestration" solutions **<input here>**

7.2.3 Storage

Describe the resources required **<input here>**
Discuss the key requirements (I/O performance, capacity, availability, reliability, any other QoS indicator) **<input here>**

7.2.4 SPECIFIC QUESTIONS REGARDING USE OF EGI.eu (FROM EGI DOC 2478)

Sample questions to capture details of a support case

These questions can help case supporters interview the case submitter and the NGIs to refine the technical details of the case and ultimately to move towards a suitable technical setup. These questions aim at understanding the user's need, the technical and other requirements/constrains of the case, and the impact that a solution would bring to the scientific community. These questions provide only guidance – Ticket owners can use other questions or even other methods to identify details of their support case(s).

- *What does the user/community want to achieve? (What's the user story?)*
- *For who does the case request resources for? (CPU/storage capacity, SW tools, consultant time, etc.) For a group? For a project? For a collaboration? Etc.*
- *What is the size of the group that would benefit from these resources, and where these people are? (which country, institute)*
- *Approximately how much compute and storage capacity and for how long time is needed? (may be irrelevant if the activity is for example assessment of an EGI technology)*
- *Does the user need access to an existing allocation (→ join existing VO), or does he/she needs a new allocation? (→ create a new VO)*
- *What is the scientific discipline?*
- *Which institute does the contact work for (or those he/she represents)?*
- *Does the case include preferences on specific tools and technologies to use?*
    - *For example: grid access to HTC clusters with gLite; Cloud access to OpenStack sites; Access to clusters via standard interdafaces; Access to image analysis tools via Web portal*
- *Does the user have preferences on specific resource providers? (e.g. in certain countries, regions or sites)*
- *Does the user (or those he/she represents) have access to a Certification Authority? (to obtain an EGI certificate)*
- *Does the user (or those he/she represent) have the resources, time and skills to manage an EGI VO?*
- *Which NGIs are interested in supporting this case? (Question to the NGIs)*

7.3 On Monitoring (and Accounting)

Please outline any requirements for monitoring of the platforms and the applications.

If you have specific tools already in use, please outline them.

Please also specify monitoring, metrics at different levels: system, performance, availability, network QoS, website, security, etc.

<input here>

7.4 On AAI

(From EGI, revise and check with WP4/5/6)

Describe the current AAI status of your community/research infrastructure

• Does your community/research infrastructure already use AAI solutions? **<input here>**

• Can you describe the solutions you have adopted highlighting as applicable: Technology adopted (e.g. X509, SAML Shibboleth,...), Identity Providers (IdP) federations integrated (e.g. eduGAIN) or approximate number of individual IdPs integrated, Solution for homeless users (users without an insitutional IdP), Solutions to handle user attributes **<input here>**

Describe the potential needs and expectations from an AAI integration in the ***services and platforms provided by INDIGO***

- *Type of IdP to be integrated (e.g. institutional IdP part of national federations and eduGAIN or non federated, social media credentials, dedicated research community catch-all IdP, ...)* **<input here>**
- *Preferred authentication technology, and requirements for support of multiple technology and credential translation services (e.g. SAML -> X509 translation)* **<input here>**
- *Community level authorization/attribute based authorization to support different authorization levels for the users* **<input here>**
- *Web access and/or non-web access* **<input here>**
- *Need for delegation (e.g. execute complex workflows on behalf of the user)* **<input here>**
- *Support for different level of assurance credentials, and need to use the information about users with lower level of assurance credentials to limit their capability* **<input here>**
- *Requirements for high level of assurance credentials (e.g. to access confidential/sensitive data)* **<input here>**

7.5 On HPC

Describe any specific issue related to the use of supercomputers.

<input here>

7.6 Initial short/summary list for "test" applications (task 2.3)

Software used	*Software/applications/services required, configuration, dependencies (Describe the software/applications/services name, version, configuration, and dependencies needed to run the application, indicating origin and requirements.)* **<input here>**
Operating system requirements	**<input here>**
Run libraries requirements	*Run API/libraries requirements (e.g., Java, C++, Python, etc.)* **<input here>**
CPU requirements (multithread,MPI, "wholenode")	**<input here>**
Memory requirements	**<input here>**
Network requirements	**<input here>**
Disk space requirements (permanent, temporal)	*Include the requirements for data transferring (upload and download of data objects: files, directories, metadata, VM/container images, etc.)* **<input here>**
External data access requirements	**<input here>**
Typical processing time	**<input here>**
Other requirements	*Requirements for data synchronization* *Requirements for data publication* *Requirements for depositing data to archives and referring them* *Requirements for mobile application components for data storage and access* *Requirements for data encryption and integrity control-related functionality* **<input here>**
Other comments	**<input here>**
Relevant references or URLs	**<input here>**

8 CONNECTION WITH INDIGO SOLUTIONS

<To be filled by INDIGO JRA >

8.1 IaaS / WP4

8.2 PaaS / WP5

8.3 SaaS / WP6

8.4 Other connections

9 FORMAL LIST OF REQUIREMENTS

<this will be further edited within WP2>

10 REFERENCES

R 1	
R 2	
R 3	
R 4	
R 5	