RIA-653549

WP2
Intro to Project & Partners

Compiled by
Jesús Marco (marco@ifca.unican.es)
Fernando Aguilarf (aguilarf@ifca.unican.es

CSIC, National Research Council, SPAIN

Kick-off meeting,
Bologna 22 April 2015

Proposed Agenda for WP2 meeting

- Presentation of WP2 (see next slides)
- Introduction of partners (idem)
- What next:

Revision of details from partners
Start discussion on requirements (Peter)
    Revise EGI experience
    Start from VENUSC input
    Wiki Template to be prepared by Jesus/Fernando
Continue with DMP (Massimiliano)
Use Cases (Alexandre)
Dissemination plan (Karolj)
Plan next steps (today and next two months) In particular: D2.1, D2.2
Agree bi-weekly teleconf and short meeting in Lisbon (18-22 may)
WRAP-UP (Jesus)

Introduction to INDIGO WP2

From the PROPOSAL, page 10:

Work package 2 (WP2, NA) represents the interest of Research Communities to assure that their requirements will be satisfied by the project outcomes, by providing feedback and participating in the revision of the services deployed.

WP2 will **keep the focus also on big data research use and management** through a dedicated task oriented to track the different needs at the data life-cycle, following the reference models used by the different Research Communities.

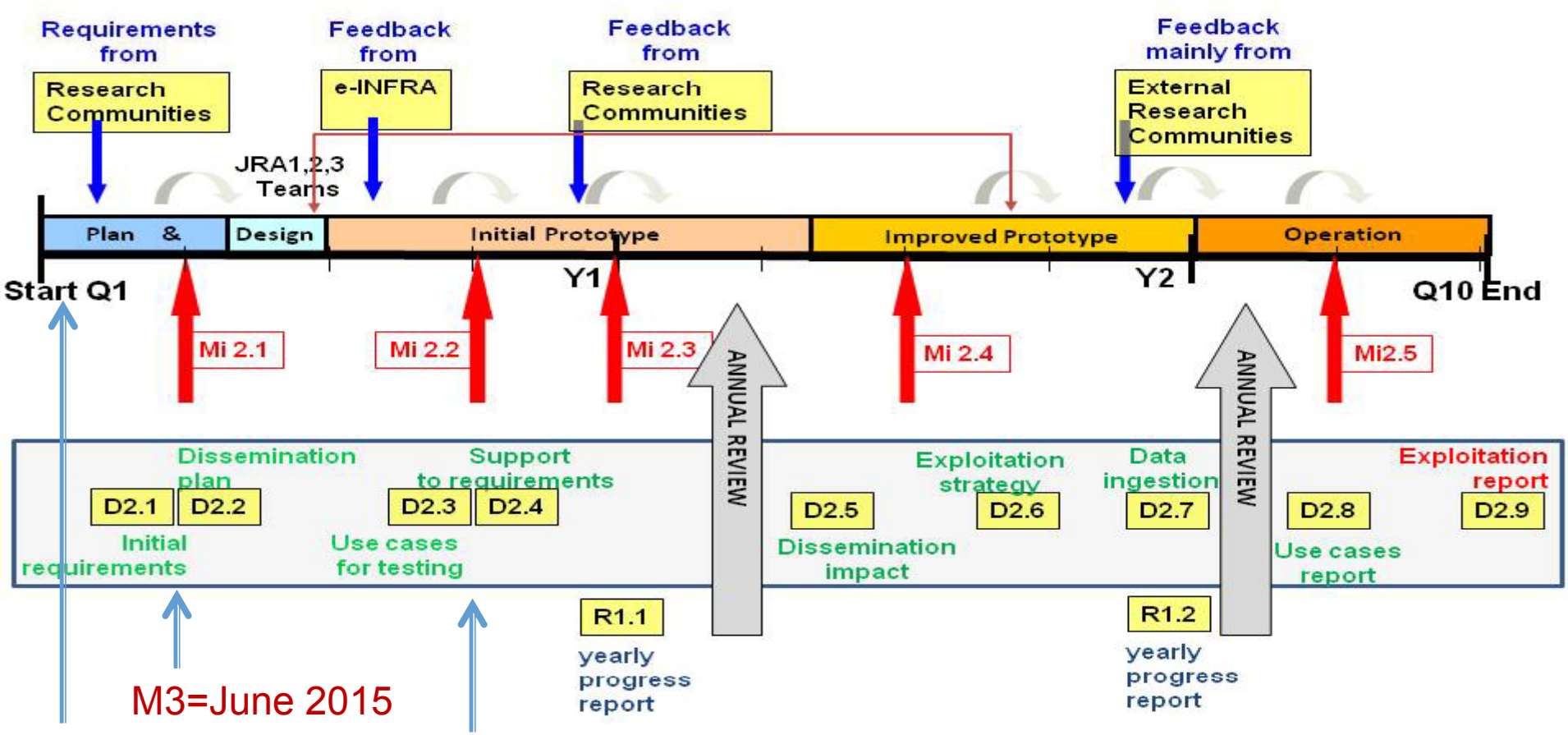
The proposed **Dissemination and Communication** activities include both strengthening **Research Community Forums** and **relations with e-infrastructure stakeholders and policy makers**.

A task in WP2 will then be devoted to **sustainability**, where the analysis of the relationships between the **different stakeholders in an open framework**, like the one proposed in INDIGO, will be done. Cooperation mechanisms between the participants and also with external users and providers, will be analyzed.

Introduction (continued): IMPACT

SIMPLIFIED IMPACT TABLE **SELECTED OBJECTIVES versus REQUESTS/ POTENTIAL IMPACT FOR COMMUNITIES** **O1: Development of the INDIGO Platform based on open software without restrictions on the e-Infrastructure**	**Life Sciences**	**Physical Sciences & Astronomy**	**Social Sciences & Humanities**	**Environmental Sciences**
Research Communities & Initiatives , including ESFRIs	**ELIXIR INSTRUCT/ WeNMR EuroBioImaging**	**CTA LBT WLCG**	**DARIAH DCH-RP**	**EMSO LIFEWATCH ENES**
Examples of Applications	**HADDOCK GROMACS AMBER GALAXY**	**MIDAS, IRAF, IDL, Geant4 ROOT/PROOF Geant4**	**Fedora Digital Libraries**	**Delft3D R-Studio TRUFA MATLAB**
Design and development of a Platform providing advanced users and community developers a powerful and modern environment for development work. This includes programming and scripting tools, and composition of custom applications and software deployment	RELEVANT	**CRITICAL**	RELEVANT	**CRITICAL**
Developing a framework to enable the transparent execution on remote e-infrastructures of existing popular applications like MATLAB / OCTAVE, ROOT, MATHEMATICA, or R-STUDIO.	RELEVANT	**CRITICAL**	MINOR	**CRITICAL**
Provide the services and tools needed to enable a secure composition of services from multiple providers in support of scientific applications.	**CRITICAL**	**CRITICAL**	RELEVANT	RELEVANT
Develop and implement a solution that is able to deploy in a transparent and powerful way both services and applications in a distributed and heterogeneous environment made by several different infrastructures (EGI Grid and Federated Cloud, IaaS Cloud, Helix Nebula, HPC clusters)	**CRITICAL**	RELEVANT	MINOR	RELEVANT
Develop the capability in the PaaS to provide unified data access despite geographical location of data, including APIs access, based on existing standards, or virtually mount like a POSIX device to worker node, cloud virtual machines, personal computer etc.	**CRITICAL**	RELEVANT	**CRITICAL**	RELEVANT

Simplified schedule



INDIGO - DataCloud

Milestones (where decisions are to be taken!)

- M2.1 [Month 3=June 2015]: **Initial requirements from Research Communities**
- M2.2 [Month 9=December 2015]: **First use cases for testing and validation purpose**
- M2.3 [Month 12=March 2016]: **Support to Requirements**
- M2.4[Month 18=September 2016]: **Exploitation strategy**
- M2.5 [Month 27=June 2016]: **Test use cases**

INDIGO - DataCloud

Task T2.1: Research Communities Requirements

- Lead partner: **EGI.eu** [18 PMs], deputy: **UPV** [12 PMs], Contributors: CMCC [9 PMs], ICCU [9 PMs], RBI [3 PMs], INGV [3 PMs], INAF [3 PMs], all [unfunded effort]
- Analyze the use cases proposed by the communities participating to the consortium**. Capture the requirements for efficiently running the applications and workflows on Cloud, Grid or HPC infrastructures**.
- Capture requirements generated by user communities not part of the project (such as the EGI Federated Cloud users), which are relevant for the outputs of the project
- Liaise with the INFRADEV-4 projects to enable synergies between the projects, and interoperability between the INDIGO outputs and the VRE to be deployed by the E-INFRA-9 projects.
- **Produce an integrated document with the requirements captured, prioritized and grouped by technical areas, for example: Cloud, HPC, Grid and Data Management**.

INDIGO - DataCloud

Task T2.2: Defining support to Research Data

- Lead partner: **INGV** [18 PMs], deputy: **CSIC** [10 PMs], Contributors: CNR [12 PMs], ICCU [9 PMs]
- To guarantee a smooth and widespread usability of INDIGO, an appropriate integration and combination approach has to take into account the different Reference Models used by the Research Communities and Research Infrastructures and the diversity and heterogeneities of data services and catalogues.
- This task follows the data research use and management of the Research Communities and Research Infrastructures and points out the different needs at the data life-cycle level.
- In particular this task shall **undertake a survey on the research communities to collect and analyze the individual Data Management Plans (DMP) and data-life-cycle documentatio**n with the aim to ensure that the full data cycle and components will be supported in INDIGO, and with the aim to provide adequate specifications for the compliance with INDIGO.

INDIGO - DataCloud

Task T2.2: Defining support to Research Data

Accordingly, the following activities are foreseen:

- Development of individual search activities to acquire and analyze the available DMP of the research communities/infrastructures with special attention to **distributed/heterogeneous data services** and catalogues and to available **open data**
- Acquisition of procedure details/parameters (i.e., DMP, Collection, Authenticity & Provenance, Data Preservation) to elaborate the specifications for data ingestion and use in INDIGO
- Definition of the specifications of INDIGO ingestion integrity test.

Task T2.3: Application Test and Validation

INDIGO - DataCloud

Lead partner: **U.Utrecht** [21 PM], deputy: **CNR** [12 PM], Contributors: CIRMMP [15 PM], EGI.eu [6 PM],CSIC [10 PM],UPV[3 PM],RBI[3 PM],INAF [12 PM], ICCU [3 PM]

The main objective of T2.3 is to ensure that all the middleware and other solutions developed in WP4-WP6 are meeting the needs and requirements of the various user communities. It is therefore crucial to properly test and validate them and demonstrate their applicability on real use cases.

Subtasks:

- Definition of use cases. We have already identified a number of use cases that will be **implemented from the start of the project**:
- Creation of VMs for each use case. For each one, a virtual machine (or container) will be provided, meeting all requirements for running on the INDIGO testing infrastructure.
- Implementation of automatic probes (e.g. NAGIOS-like) for performing all tests on a regular basis. These will allow validation and future monitoring of the INDIGO infrastructure, and will be made in coordination with WP3.
- Creation of generic use case examples that can be used for dissemination and training purposes

INDIGO - DataCloud

Task T2.3: Application Test and Validation

- A local, self-contained version of the **HADDOCK** portal in a VM (to be used for both multicore and cluster-like implementations). HADDOCK is a typical high-throughput, highly distributed application, which has already been ported to the grid and is widely used (>4400 users worldwide).
- A multi-threading molecular dynamics use case based on the **GROMACS** and/or AMBER software for testing VM with a large number of cores (possibly with connections to PRACE).
- An **MPI-based molecular dynamics** use case to run on a virtualized cloud cluster.
- An approach for the characterization of **internal dynamics in multi-domain proteins** integrating different types of experimental data
- A **Climate model intercomparison** analysis, based on big data analytics workflows of climate data operators (including data reduction, re-gridding, intercomparison as well as statistical, outlier and ensemble analysis) on multi-terabyte climate datasets from large data collections (e.g. CMIP5)
- **An astronomical pipeline** to reduce proprietary or public data from LBT telescope (acquire data from LBT archive and running pipeline on a virtualized cloud cluster) and CTA simulations.
- An instantiator of self-contained **GALAXY** servers running on VMs. GALAXY is a workflow manager well known to the bioinformatics community. It is in particular widely adopted for setting up complex Next Generation Sequencing data analysis pipelines.

INDIGO - DataCloud

Task 2.4: Dissemination towards Research Communities

- Lead partner: **RBI** [24 PMs], deputy: **EGI.eu** [4 PMs], Contributors: CMCC [3 PMs], INGV[3 PMs], U.Utrecht[3 PMs],CIRMMP[6 PMs], all [unfunded effort]
- Dissemination activities target **both project partners and external researchers**, as well as any scientific bodies interested in intermediate and final results of the project. The information will be classified according to the target audience (internal, external) and according to the state of work (progress of ongoing activities, preliminary results, intermediate results, final results).
- An important feature of these dissemination activities, i.e. the promotion of the project solutions, lies its foundation on a **strong engagement of the Research Communities partners**, that include relevant ESFRIs, EIROS, and other large research initiatives in Europe. These partners will contribute with the identification of the best forums and dissemination channels, and will help also by providing specific examples of the usefulness of INDIGO solutions in their communities.

INDIGO - DataCloud

Task 2.4: Dissemination towards Research Communities

- The coordination of these activities, and the preparation of specific support, will be handled by RBI, with wide experience on these tasks, and by EGI.eu, that through its experience in large dissemination events and training activities, will contribute to a wider dissemination to many research areas.

- Following the ideas presented in section 2.2 for the initial dissemination, a **structured plan will be implemented covering the whole project duration**. The introduction of the cloud service technology platform is also an opportunity to integrate existing advanced technologies, including multimedia, hypermedia and other new models that will make the access to this platform attractive and easier, and so will also contribute to the sustainability of the project results.

- Different outreach actions will be made to assure the dissemination towards the general and specific research communities: EGI.eu Community Forum, EGU and AGU for Environmental and Earth Sciences, CHEP for High Energy Physics, INSTRUCT biennial meetings and relevant large life science conferences for Biomedical sciences, and also industry-specific meeting like the conferences targeting industry for structural biology/life sciences - pharmas, biotechs (see for example http://www.psdi2014.org).

INDIGO - DataCloud

Task 2.4: Dissemination towards Research Communities

The following is an initial list of activities planned:

- Promote the presentation of project results through **contributions at scientific conferences, publications in journals, participation in workshops and scientific events. Colocate as possible promotional stands** and/or material for the general and scientific public.
- Organize **specific workshops, tutorials and hackathon events to disseminate the project results at relevant** technical conferences (EGI TF, RDA meetings, EUDAT meetings, CloudOpen Europe, ESOCC, etc.) as well as a number of solicited workshops at relevant organizations.
- Attract new potential customers through **short on-line training courses on 'application enabling'. Promote** them in Universities and Research Infrastructures.
- Setup the required **e-infrastructure components (in collaboration with task T2.3 and WP3) that will facilitate both training and demos at workshops, and on-line courses.**
- Setup a section of the **project website/portal to present, from the perspective of final research users and of** developers and technologists, the project facts, current progress and results, as well as provide links to the corresponding documentation, and also to repositories. Ensure that the project outcomes are available, well documented, public and searchable.
- Collaborate technically to the design of **press releases, updates of the project story-factsheet, and dedicated technology factsheets of the solutions, including produced pipelines/workflows, that can be used to attract** new customers.
- Present results at identified **technology industry-related events in order to raise awareness of the** availability of the project results, and the channels to support their exploitation.

INDIGO - DataCloud

Task 2.5 Sustainability: exploitation strategy

- **Lead partner: CSIC [10 PMs], deputy: EGI.eu [6 PMs], Contributors: CIRMMP [9 PMs], all [unfunded effort]**
- Task T2.5 aims to provide an exploitation strategy for INDIGO products, assuring the sustainability of the initiative.
- The task will take into account the needs, limitations and objectives of the parts involved: research communities, technical developers and e-infrastructure providers, including commercial ones. The sustainability of the outcomes is only assured if the research communities use them.
- To expand the user base of INDIGO products among research communities, external partners (i.e. SMEs willing to collaborate), will be explored. This will be done considering the organizational possibilities and constrains of the research sector.

INDIGO - DataCloud

Task 2.5 Sustainability: exploitation strategy

Subtasks proposed are:

- Analyze the **organizational ecosystem around INDIGO** outcomes to expand the user base, finding ways to overcome possible constrains.
- **Identify, key components developed and supported within INDIGO**, and the way they are integrated into applications providing added value with focus on the perspective of Research Infrastructures and Communities.
- **Identify potential involvement of industrial partners**, starting with those already contributing to INDIGO JRA, and including also selected units within large corporations and SMEs mainly in Europe. *For example, a pool of Research Infrastructures interested in supporting INDIGO solutions through a commercial partner or several partners.*
- Define potential agreements with research consortiums and/or commercial companies (arising from the previous activity), considering revenue streams in any form: funding and/or potential business models, if any.
- Extend the study to **a global scope in Europe, and to a wider international range** (US, Asia, Latin America, Australia).
- Eventually, **consider the creation of an INDIGO Open Source consortium** and promote the involvement of European companies to support key developments.

INDIGO - DataCloud

Deliverables: next 3 months

D2.1, Month 3 (R; PU): Initial requirements from Research Communities [EGI.eu]

- This report summarizes the findings of T2.1 and T2.2 along the first three months of the project, **providing input to JRA activities**. The report will be an integrated document including a general description of the research communities involved and of the use cases and workflows proposed and will express requirements captured, prioritized and grouped by technical areas (**Cloud, HPC, Grid, Data Management**, etc.).

- In particular, the analysis of Data Management Plans (DMPs) and data lifecycle documentation aiming to identify both synergies and gaps among the different communities will be provided.

D2.2, Month 3 (R,DEC,PU): Consolidation of dissemination plan, including project website and promotion [RBI]

- This report includes the complete dissemination plan for the project, prepared in T2.4, consolidating the initial plan discussed in the proposal, confirming the strategy and indicators, defining a calendar, and actions involving the different partners. The report also describes the first actions, including the launching of the project website and promotional material, and participation in general events.

INDIGO - DataCloud

Deliverables: first year

D2.3, Month 8 (R;PU): Specifications of use cases for testing and validation purposes [U Utrecht]

- This deliverable will provide details of the use cases in T2.3 that have been originally defined and newly identified from the feedback of user communities (task 2.1) to test and monitor developments in WP3 (and WP6).

D2.4, Month 9 (R,PU): Confirmation of support to initial requirements and extended list [EGI.eu]

- This report will confirm from the point of view of the research communities that the initial requirements analysed in T2.1 and T2.2 have been taken into account in the design proposed by JRA for INDIGO solutions.
- Additionally, and based on the evolving experience and potential interest of the communities, an extended list will be provided to be taken into account in future developments and releases. This feedback will be provided to JRA workpackages.

INDIGO - DataCloud

Deliverables: Second year

D2.5, Month 15 (R,DEC;PU): Report on dissemination effort and impact [RBI]

- Prepared in T2.4, after D2.2, it will cover the experience along the first half of the project, including detailed metrics on the impact based on direct indicators (website, questionnaires, participation in events, in particular training and specific workshops) and indirect ones (references to INDIGO solutions in different research communities websites, publications, etc.).

D2.6, Month 21 (R,CO): Exploitation strategy and sustainability [CSIC]

- The report from T2.5 will analyse, once the initial solutions and impact of the project are established, possible ways to define an exploitation strategy involving the different stakeholders. Identifying the main actors (like for example, ESFRIs, technology providers, commercial partners including SMEs and selected units within large companies) and their roles. Also understanding how other open source projects are exploited and become sustainable.

D2.7, Month 24 (R,PU): Specifications of data ingestion and use in INDIGO [INGV]

- This report from T2.2 will present the results of the different enquiring activities for each Research Community/Infrastructure about procedure details and parameters and it will include the specifications of INDIGO ingestion integrity test.

INDIGO - DataCloud

Final Deliverables

D2.8, Month 27 (R,PU): Test and validation suite and results [U Utrecht]

- This deliverable (in T2.3 and related to D2.3) will describe the additional use cases implemented during the project to test and monitor developments in WP3 and WP6, together with the monitoring tools and a report of their application

D2.9, Month 30 (R,PU): Exploitation analysis based on agreements made and usage statistics [CSIC]

- The report, within task 2.5 and following D2.6, will describe the formula, agreed with the different research communities, and with the consortium, to guarantee the support of INDIGO solutions. The potential exploitation impact will be based on usage statistics along the last six months of the project, after the last main software release. The analysis will be used to explore new projects or exploitation initiatives.

Partner	Where	Community/ ESFRI, project	Tasks/PM 2.1	2.2	2.3	2.4	2.5	Team
CSIC IFCA	Santander ES	Biodiversity/Ecosystems LIFEWATCH		**10**	10		**10**	**Jesus Marco**, Fernando Aguilar
UPV I3M	Valencia ES	Biobanks, Molecular Imaging EUROBIOIMAGING, BBMRI	**12**		3			**Ignacio Blanquer,** Erik Torres
CIRMMP	Florence, IT	Structural Biology INSTRUCT/WENMR			15	6	9	**Antonio Rosato** Andrea Giachetti
INAF	IT	Astrophysics *LBT,CTA*	3		12			*Riccardo Smareglia, Lucio Angelo Antonelli*
UTRECHT	Utrecht NL	Structural Biology INSTRUCT/WENMR			**21**	3		**Alexandre Bonvin,** *postdoc*
CMCC	Lecce IT	Climate ENES	9			3		**Sandro Fiore**, Giovanni Aloisio
ICCU	Rome IT	Libraries DCH-RP	9	9	3			**Sara Di Giorgio,** Davide Madonna
EGI.eu	Amsterdam NL	ALL	**18**		6	**4**	**6**	**Peter Solagna**
CNR	Milan IT	Genomics (NGS) ELIXIR		12	**12**			**Federico Zambelli** Giuseppe Profiti
INGV	Rome IT	Seafloor Observatories EMSO	3	**18**		3		**Massimiliano Rossi** Lucio Badiali
RBI	Zagreb CRO	DARIAH	3		3	**24**		***Karolj Skala***

INDIGO - DataCloud

Communication and support

- Mail list: **indigo-wp2[at]lists.cnaf.infn.it**
- Adobe Connect: http://connect.ct.infn.it/data-cloud-wp2/
- WIKI: pending
- Meetings:
  - **Bi-weekly**
  - SAMOSAS rules
- "Methodology" tools:
  - tbd with JRA
  - AGILE style

Cal EECS

Good Meetings: SAMOSAS

(Photo by K.S. Poddar. Used by permission under CC-BY-SA-2.0.)

- Start and stop meeting promptly
- Agenda created in advance; no agenda, no meeting
- Minutes recorded so everyone can recall results
- One speaker at a time; no interrupting talker
- Send material in advance, since reading is faster
- Action items at end of meeting, so know what each should do as a result of the meeting
- Set the date and time of the next meeting

Minutes and action items record results of meeting, start next meeting by reviewing action items

INDIGO - DataCloud

P0 IFCA / CSIC

- **CSIC=National Research Council in Spain (>120 centers)**
- **IFCA=Institute de Física de Cantabria (at Santander)**
  - **High Energy Physics + Astrophysics + Complex systems**
  - **Advanced Computing and e-Science, Grid computing**
  - **Data Center: EGI, FedCloud, WLCG/CMS Tier-2, Supercomputing node**
- Relevant to INDIGO proposal:
  - IFCA/CSIC is participating in WP1, WP2, WP3, WP4, WP5, in **EGI-Engage**

Initial WP2 Team at IFCA

Jesús Marco de Lucas
marco@ifca.unican.es
Senior physicist (HEP and computing)
EU projects: CrossGrid, EGEE, DORII, EGI-Inspire, COOPEUS.
LifeWatch core-ICT board

Fernando Aguilar
aguilarf@ifca.unican.es
Senior engineer in informatics at IFCA.
Data management in ROEM+ (Life+) (whole data cycle & models)
Projects: COOPEUS, LifeWatch
Experience in connecting users and e-infrastructure

P0 IFCA/CSIC Recent experience

- **Collaborating to setup the core-ICT platform components for LifeWatch:**
  - IaaS resources (into FedCloud)
  - PaaS platform (collaborative framework) and SaaS examples
  - Control Platform

- Participating in different LifeWatch and biodiversity-related projects, including requirement gathering, data management, resource providing and support:
  - for GBIF-Spain (mainly IaaS)
  - for VLIZ /Be in marine biodiversity project
  - for large data processing to analyze snow coverage with UGr

NOW STARTING THE LIFEWATCH COMPETENCE CENTER IN EGI-ENGAGE

- Collaboration with US Research Infrastructures (COOPEUS FP7 project), NEON, DataOne Working to discover synergies in similar initiatives including:
  - DMPs (Data Management Planning)
  - Data Protocols
  - Data integration tools/Portals (GEOSS)
  - Modeling initiatives (for Global Carbon cycle)

INDIGO - DataCloud

P1 UPV (Valencia)

- Contact person
  - Ignacio Blanquer

Ignacio Blanquer is associate professor of the Computer Systems Department at the UPV. He was awarded with the extraordinary phd. prize in 2003. His research focuses on biomedical informatics and Parallel Computing, participating in 55 European, national and regional research projects. He is author and co-author of 32 publications in indexed journals and book chapters and over 82 communications conference proceedings.
He currently leads the Grid and High-Performance Computing (GRyCAP) research group. He is the coordinator of EUBrazilCloudConnect project, funded by the EC and the Brazilian CNPq and co-coordinator of the national project Virtual elastic clusters on Hybrid Cloud Infrastructures.

- Other UPV people in WP2
  - Erik Torres

Erik Torres is a Researcher in the I3M. He received his PhD. in computer science from Valencia Technical University in 2010. He has been involved in Grid Technologies and Job Execution Management for over 7 years ago, participating in 9 National and European Research Projects. In EUBrazilOpenBio, he was responsible for the adoption of Ecologic Niche Modelling and was the deputy coordinator of the WP on the specification and validation of requirements. He is currently the User Scenario Workpackage Leader in EUBrazilCloudConnect. He has authored and co-authored 5 publications in indexed journals and books chapters and more than 10 communications in conference proceedings.

P1 Previous Experiences in Gathering Requirements

INDIGO - DataCloud

- VENUS-C
  - We initiated a process of analyzing coarse-grain requirements related to representative use cases from the user community
    - 7 “Official” use scenarios
    - 20 additional ones.
  - D4.1. Defines the process and the template
    - http://www.venus-c.eu/deliverables_year1/VENUS-C_D4.1.pdf
  - D4.2 Also includes the evaluation of the fulfilment of requirements
    - http://151.1.219.218/ed6bcd7b-3378-48dc-a045-96fbaca1592c.pdf

- EUBrazilCloudConnect
  - We target 3 scientific use cases with different goals and requirements, including massive data processing, HPC and elastic scalability.
  - D5.1 defines the process and the templates, describing the requirements that arise from the use cases:
    - https://dl.dropboxusercontent.com/u/22347738/EUBrazilCC/EUBraCC-D5_1.pdf
    - 39 use case requirements & 12 infrastructure-level requirements.
  - D5.2 reports the progress made in the 1st year of the project:
    - https://dl.dropboxusercontent.com/u/22347738/EUBrazilCC/EUBraCC-D5_2.pdf
  - D5.3 and D5.4 (due to Jan-16) will provide the final report on implementation and validation.

INDIGO - DataCloud

P2 CIRMMP, Florence

- Prof. **Antonio Rosato**, Associate Professor of Chemistry at the University of Florence, Italy. He is an expert in structural biology and bioinformatics. His research work focuses on the development of software tools to assist the analysis of NMR data and on the bioinformatics analysis of genomes and proteins. He has been the PI of the Italian CIRMMP unit within the WeNMR e-Infrastructure project, a EC-funded project that developed a web-based, grid-enabled computational electronic infrastructure for structural biology (www.wenmr.eu). He has coauthored about 90 publications (http://www.researcherid.com/rid/D-8548-2011).

Recent relevant grants:

1.MoBrain competence center of EGI Engage - https://wiki.egi.eu/wiki/EGI-Engage:WP6

2.WeNMR - Collaborative Project & Coordination - Support Action - A worldwide e-Infrastructure for NMR and structural biology - n. 261572 – www.wenmr.eu

3.BioMedBridges – Combination of CP & CSA - Building data bridges from biology to medicine in Europe - n. 284209 – www.biomedbridges.eu

INDIGO - DataCloud

P2 CIRMMP, Florence

- Dr. **Andrea Giachetti** is an expert in the development of scientific software tools and web interfaces. He received an International Ph.D. in Structural Biology jointly by the Universities of Florence, Frankfurt and Utrecht. He has a significant expertise on the management of authentication and secure access over infrastructures for distributed computing. He has coauthored about 15 publications.
- Mr. **Enrico Morelli** is the system administrator and responsible for ICT at CIRMMP. He has a long-term experience in developing databases and web-based interfaces.
- Contact:

Antonio Rosato and Andrea Giachetti, together with Dr. Dario Carotenuto, are the cofounders of the company Cloud Centurion srls. The company is operative in the fields of distributed and cloud computing as well as of cloud storage, with a main focus on data security and data sharing over multiple cloud services.

INDIGO - DataCloud

P3 INAF (Astrophysics)

INAF is the main Italian research institution in the field of Astronomy and Astrophysics. INAF comprises 19 observatories and institutes spread across the country and employs more than 1200 people, about 700 of which are research astronomers. Its participating structures are: INAF-Osservatorio Astronomico di Trieste (OATs) and INAFOsservatorio Astronomico di Roma (OAR). INAF promotes, realizes and coordinates, also within programs of the European Union and international bodies, research activities in many astronomical fields. INAF scientific research activity is divided into five themes: Galaxies and Cosmology; Stars, stellar populations and interstellar matter; The Sun and the solar system; Relativistic astrophysics and astroparticles; and Advanced technologies and instrumentation. INAF also coordinates a national Virtual Observatory initiative (VObs.it). It participates in the Euro-VO initiative and plays an active role in the projects funded by the EU Framework Programmers like VO-TECH, EuroVO-DCA, EuroVO-AIDA and more.

INAF – OATrieste hosts the Italian Center for the Astronomical Archive (IA2) which manages also data for the whole LBT (Large Binocular Telescope) corporation and it is the main point of INAF activity on the **Virtual Observatory in Italy.**

INAF- OA Roma is deeply involved in the CTA activities and also manages the Italian support for LBT observation and science.

INDIGO - DataCloud

P3 INAF

Dr. Riccardo Smareglia *got several contracts at ESO and ESA to develop software for a number of astrophysical archives. Since 1988 at the* ***Astronomical Observatory of Trieste****; permanent staff since 1991.* ***He designs and implements the archive at TNG****. He participates to the design and development of a number of other astronomical archives making them fully compatible with VO standards. Participation in several EU FP7 funded projects such as Euro-VO-DCA, Euro-VO-AIDA, Euro-VI-ICE. Since 2004 responsible of of the creation and management of the* ***Italian Centre for Astronomical Archives (IA2) who manage data from the main ground base INAF telescopes*** *. In 2009 and 2012 he promotes and chairs the "Vo-Day … inTout" and "VO publish" initiative funded through the EU FP7 Euro-VO-AIDA project. Since 2013 he is the person in charge of the ICT INAF organization.*

Dr. Lucio Angelo Antonelli is a senior INAF Research Astronomer, *author/co-author of about 450 papers, proceedings of conferences, circulars and technical reports. Project Scientist of the ASI Science Data Center (2005– present), INAF representative in the MAGIC collaboration (2012-present),* ***member of the Executive Board of the MAGIC collaboration (2014). He joined the CTA Project and in particular the Data Management activity since 2006.***

Since 2009 he is responsible for the CTA Archive Work Package. Since 2009 he is also responsible for the Data Handling and Archiving of the ASTRI Project. ASTRI is the INAF Flagship Project for the realization of and end-toend

prototype of the CTA Small Size Telescope and a mini-array of SSTs

Relevant projects and activities

- IA213 - Italian Center for Astronomical Archive: IA2 is an ambitious Italian Astrophysical research infrastructure

P3 INAF, Relevant projects

IA213 - Italian Center for Astronomical Archive

IA2 is an ambitious Italian Astrophysical research infrastructure project that aims at coordinating different national initiatives to improve the quality of astrophysical data services. It aims at coordinating these developments and facilitating access to this data for research purposes. The target of the proposal aims at ensuring harmonization with other projects related to archiving of data of astrophysical interest, with particular reference to projects involving the Italian astronomical community (TNG, LBT, VST, GSC-II, DPOSS, …), to the Italian Solar and Solar System Physics community (SOLAR, SOLRA, ARTHEMIS which form SOLARNET) and to the national and international coordination efforts fostering the idea of a multiwavelength Virtual Astronomical Observatory within the IVOA (International Virtual Observatory Alliance) community, and the use of the archived data through the Italian Astronomical Grid. IA2 also provide to host several archives/catalogs from Italian research groups and publish their data by Virtual Observatory services.

IVOA14: INAF is involved in several Virtual Observatory (VObs) activities from standard development, software implementation and knowledge dissemination. All these activities were developed within EU-Funded projects (Vo-Tech, VO-DCA, VO-AIDA, Cosadie) from IA2 team. IA2 team had also developed and manage the VODance software: here users just have to provide a database connection to local center that points to their available data and fill out a metadata description form without having to export their data. Data Access Layer services are created on the fly and published to VObs

P4 U. Utrecht

Prof. Dr. A.M.J.J. Bonvin (PI)

Contact Address

Bijvoet Center for Biomolecular Research

Science Faculty/Chemistry, Utrecht University

Padualaan 8, 3584 CH Utrecht, the Netherlands

Web: http://www.uu.nl/staff/AMJJBonvin / http://bonvinlab.org

Short CV

Alexandre Bonvin studied Chemistry at Lausanne University, Switzerland and obtained his PhD at Utrecht University in the Netherlands (1993). After two post-doc periods at Yale University and the ETHZ, he joined Utrecht University in 1998 where he was appointed full professor of computational structural biology in 2009. In 2006, he received a prestigious VICI grant from the Dutch Research Council. He is participating to several EU projects, has coordinated the WeNMR e-Infrastructure project (http://www.wenmr.eu) and is leading the MoBrain Competence Center under EGI-Engage. His work has resulted in over 170 peer-reviewed publications (Research ID A-5420-2009). His group is developing the widely used HADDOCK software for the modelling of biomolecular complexes (http://bonvinlab.org/software).

To be hired: post-doc (expected starting date September 1st)

INDIGO - DataCloud

P5 CMCC

Giovanni Aloisio:

- **Giovanni Aloisio** is full professor of Information Processing Systems at the Dept. of Innovation Engineering of the University of Salento, Lecce, Italy, where is leading the HPC laboratory. Former director of the Scientific Computing and Operations (SCO) Division at CMCC, he is now the Director of the Supercomputing Center and member of the Strategic Council and of the Executive Committee. His expertise concerns high performance computing, grid & cloud computing and distributed data management. He was strongly involved in EU grid projects such as GridLab, EGEE, IS-ENES1. Presently, he is coordinating CMCC activities into several EU FP7 projects such as EUBrazilCC, IS-ENES2, CLIP-C and the G8 ExArch. He has been the responsible for **ENES** of the EU-FP7 EESI (European Exascale Software Initiative) project and for University of Salento (as PRACE Third Party) of the EU-FP7 EESI 2 project, also chairing in both cases the WCES (Weather, Climate and solid Earth Sciences) European Working Group. He is member of the ENES HPC Task Force and one of the key experts of the IESP project (International Exascale Software Project), whose main goal is the definition of the roadmap for a common, open source software infrastructure for scientific computing at exascale. He is the author of more than 100 papers in referred journals on high performance computing, grid computing and distributed data management.

Sandro Fiore:

- **Sandro Fiore** *Ph.D.*, is the Director of the Advanced Scientific Computing (ASC) Division of the Euro-Mediterranean Centre on Climate Change. His research activities focus on parallel, distributed, grid and cloud computing, in particular on distributed data management, data analytics/mining and high performance database management. He is Visiting Scientist at Lawrence Livermore National Laboratory (LLNL) working at PCMDI in the context of the Earth System Grid Federation (ESGF). Since 2004, he has been involved into several national and international projects like: EGEE (the 3 cycles), EGI-InSPIRE, IS-ENES1 and IS-ENES2, EUBRAZILCC, ExArch, ORIENTGATE, TESSA, OFIDIA, CLIP-C, INDIGO-DataCloud working on data management topics. Since 2010, he is the Principal Investigator of the Ophidia project, a research project on high performance data analytics and mining for eScience. He is author and co-author of more than 50 papers in refereed books/journals/proceedings on distributed and grid computing and holds a patent on data management topics. He is editor of the book “Grid and Cloud Database Management” (Springer, 2011). He is ACM Member.

Links

http://www.cmcc.it/divisions/asc
http://www.cmcc.it/supercomputingcenter
http://www.cmcc.it/projects/eubrazilcc-eubrazil-cloud-connect
http://www.cmcc.it/projects/is-enes2-infrastructure-for-the-european-network-for-earth-system modelling-phase-2
http://www.cmcc.it/projects/ofidia-operational-fire-danger-prevention-platform
http://www.cmcc.it/projects/clipc-climate-information-platform-for-copernicus

INDIGO - DataCloud

P6 ICCU (MiBACT)

- Contact person
  - Sara Di Giorgio

She's technical coordinator of the web-portal CulturaItalia and she participated, as representative of ICCU, in many EU projects, among the others: MINERVA, MICHAEL, ATHENA, ATHENA Plus, EUROPEANA family projects, DC-NET, DCH-RP. Actually she is involved in the EU projects ARIADNE and PARTHENOS that are investigating relationship between research infrastructures on digital cultural heritage and e-infrastructure. During her career S. DI Giorgio acquired specific competences in the field of Digital Humanities, mainly in Digital Libraries, metadata standards, thesauri, ontologies and knowledge-organization for cultural digital resources and long term preservation.

- Other ICCU people in WP2
  - Davide Madonna

He is the interoperability manager of CulturaItalia and he worked on the development of specific tools to improve the acquisition of metadata and digital collection between different informative systems.
He worked also for several European project like ATHENA, ATHENAPLUS, LINKED HERITAGE, PARTAGE PLUS, EUROPEANAPHOTOGRAPHY) as mapping and interoperability expert.

Unfortunately Davide will not participate to the kick-off meeting

P6 ICCU Previous Experiences in Gathering Requirements

- **DC-NET** (2009-2012)
  - The project analyzed how the co-operation between cultural institutions and e-Infrastructure providers ca generate innovative services, tools and data sets to support the research of multidisciplinary communities
    - **DC-NET Service priorities and best practices for digital cultural heritage** (PDF, 7516 kb)
- **INDICATE** (2010–2012)
  - The project coordinated policy and best practice regarding the use of e-Infrastructures for digital cultural heritage
    - Requirements for virtual exhibition using e-infrastructure **Handbook on virtual exhibitions and virtual performances** (PDF 10 mb)
    - A concrete analysis for the development of a vision for DCH shared by cultural institutions and e-infrastructure providers **Best practice for applying research pilots and use case studies to digital cultural heritage** (PDF 6,5 mb)
- **DCH-RP** (2012-2014)
  - ICCU conducted 2 Proofs of

    Concept for preserving data on an external e-infrastructure in order to find preservation solutions beyond the use of local storage.
  - We used Grid storage available on the European Grid Infrastrucure to store data and to use the **e-Culture Science Gateway** (eCSG) as the tool to copy data from the local store to the grid store and to access the data afterwards
    - D5.3 **Report on first Proof of Concept** (.PDF, 1,014kb)
    - D5.4 **Report on the second proof of concept** (.PDF,1,748 kb)

INDIGO - DataCloud

P7 EGI.eu

- Peter Solagna, *Senior Operations Manager*
  - Peter is responsible for the EGI infrastructure operations, the production infrastructure and the federation enabling services. Peter has been working in EGI for 5 years, and he has extensive experience in software development and IT service management applied to high performance and distributed computing infrastructures.
- Małgorzata Krakowian*, Senior Operations Manager*
  - Małgorzata has been working in operations since 2006, working at EGI.eu since 2012. Among other activities, Małgorzata has been working on operational integration of new technologies, and requirements gathering and pilots definition for the ENVRI and EGI-InSPIRE projects.
- Yin Chen, Senior Outreach Expert
  - Yin has joined EGI.eu outreach team in 2015. Previously she was mainly working in ENVRI project, she was leading ENVRI work package responsible for user communities requirements gathering and analysis. She led the definition of the ENVRI reference model standard.
- Sara Coelho, *Communications Manager*
  - Sara leads the EGI.eu communications activities, organizing the main EGI events and she is responsible for the communications strategy of the EGI-Engage project.

INDIGO - DataCloud

P7 EGI.eu
Requirements gathering in EGI

- During the EGI-InSPIRE project EGI.eu has established collaborations with the EMI and IGE projects, gathering requirements from the EGI communities, and worked together with the external technology providers to prioritize and deliver the developments in production for the users.
- The high level classification of the communities involved includes: high energy physics, life sciences, humanities and other disciplines.
- EGI.eu collaborated at different levels with many initiatives also outside EGI-InSPIRE, for example: DRIHM, VERCE, BioVEL, DCH-RP, EISCAT-3D, ENVRI project, CTA, WeNMR and more.
- By supporting the deployment of use cases in the fedcloud EGI.eu worked together with 26 communities to integrate 50 use cases

INDIGO - DataCloud

P8 CNR Bari, Milano, Bologna

- **Graziano Pesole**: is full professor of Molecular Biology in the University of Bari and Director of the Institute of Biomembranes and Bioenergetics of the National Research Council

- **Federico** has the role of **Technical Coordinator for the ELIXIR-Italy Joint Research Unit**. His interests are bioinformatics and in particular Next Generation Sequencing data and analysis tools development. He obtained a PhD in Cellular and Molecular Biology at University of Milan (Milano, Italy) in 2009 and is author of about fifteen bioinformatics papers with an h-index of 12 (Google Scholar) and more than 300 citations and co-author of books on Bioinformatics published by international (Wiley, Springer) editors.

- **Giuseppe Profiti** has the role of **Local Technical Coordinator for the University of Bologna node of ELIXIR-Italy**. His interests are algorithms and artificial intelligence, with applications on different domains. In June, he is going to discuss his PhD thesis in Computer Science at University of Bologna. He is also a research associate at the Health Sciences and Technologies center of the same university.

INDIGO - DataCloud

P9 INGV

- Contact person
  - Massimiliano Rossi

Massimiliano Rossi is a Computer Science Engineer at INGV with 12 years' experience in ICT; 7 years' experience and consultant in: - network, security and software architectures, - European project (FP), eLearning ;
5 years' experience in: Education and Training.
Current interests: Long Term Data Preservation, OAIS Standard, Network security, Support Vector Machine analysis, Cyber crime analysis.
Main Projects : SCIDIP-ES, WEIRD, EUQoS, DAIDALOS .

- Other INGV people in WP2
  - Lucio Badiali

Lucio Badiali earned a degree in Physics at "La Sapienza" Rome University, major in cybernetics and a PhD .INGV Data Protection Officer. INGV Member of the Italian Grid Infrastructure board, the italian branch of the EGI (European Grid Infrastructure).
Expert consultant for the Italian Ministry of Foreign Affairs.
Expert consultant for the UNO-PS where prepared the Contingency Plan for the Research Ministry of the Democratic Repubblic of Congo for the volcanic risks and designed the seismic and volcanic acquisition network for the Vocanic Observatory in Goma (RDC) for the Niyragongo volcano.

INDIGO - DataCloud

P9 -INGV

- Other INGV people in WP2
  - LauraBeranzoli

Laura Beranzoli is Technology Director with Degree in Physics with education in geophysics. She has about twenty years of experience in Seafloor Observatory Science. Her main expertise is in experiment design and deployment and seafloor geophysical time-series analysis. She has management experience in her field of interest as WP leader in EC projects COOPEUS, NEAREST, ESONET NoE and she is involved in running EC and national projects. She is in the coordination team of EMSO, research infrastructure of ESFRI.

INDIGO - DataCloud

P9 - INGV

- Other INGV people in WP2
  - Manuela Sbarra

Manuela Sbarra graduated in Computer Science from "La Sapienza" Rome University.
Currently works at INGV.
9 years' experience in the network-security, software architectures and web application development.
Main projetcs: Operational Earthquake Forecasting in Italy, SCIDIP-ES

- Other INGV people in WP2
  - Nicola Marcucci

Marcucci Nicola is a Computer Technician at INGV within the EMSO project with 7 years experience in ICT: data and database management, systems engineering, web development.
Main Projects: EMSO, ENVRI

P9 – INGV Previous Experiences in Gathering Requirements

INDIGO - DataCloud

SCIDIP-ES
SCIENCE DATA INFRASTRUCTURE FOR PRESERVATION-EARTH SCIENCE

http://www.scidip-es.eu/scidip-es/deliverables/

D15.1 User Surveys

We participated in the definition of users consultation to identify preservation policies and harmonised semantics.

D33.1 Data Access Policy report

Participation to independent search for gathering European data access policy

D12.1 Requirement specification and gap analysis report

Capturing the requirements of the SCIDIP-ES user community

MOIST-Adopted Scenario – Use-case

- Marine multidisciplinary datasets related to the GEOSTAR-type observatories, including seafloor data acquired by INGV from 1998 to the present.
- MOIST LIDO CTD campaign dataset is generated from the Iberian margin sea floor observatory in the Gulf of Cadiz. Data have a significant scientific value due to the complexity of the region, being on the junction of the Eurasian and African plates.

INDIGO - DataCloud

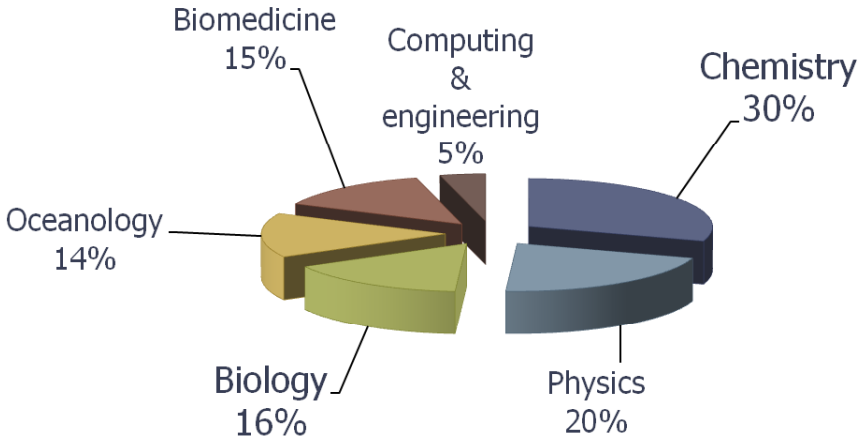
P10 RBI General Overview

- The largest scientific institute in Croatia
- Founded in 1950 by JASA
- The first research activities of the institute: theoretical, nuclear and molecular physics and electronics followed by cybernetics and computing
- Today consists of 12 divisions and 2 centres
- The fundamental and applied research areas are: biology, medicine, environmental science, computational science, electronics/engeenering, physics and chemistry

INDIGO - DataCloud

P10 RBI RBI Employees

- More than 800 employees, ~ 6% of the total no. of scientists in Croatia
- 20 000 m^2 area, 12 divisions, 82 laboratories



- Overview by branch

INDIGO - DataCloud

P10 RBI - Scientific Computing and Information Processing

- **Scientific computing, modelling and visualization**
  - High throughput and High Performance Computing
  - Methods of multidimensional and multispectral scientific visualization
  - Applications in marine sciences and health
- **Information processing and knowledge technologies**
  - Advanced measurement methods and reconfigurable embedded systems
  - Applications in bio-sciences
  - Data representation and transformation for information and signal processing: sparse representations, unsupervised factorization methods, nonlinear processing methods
  - Algorithms, methodologies and new architectures for machine learning, data mining, knowledge representation, string processing

INDIGO - DataCloud

P10 RBI

- Specifically, and explicitly pertinent to the aims of this project, is the Center's involvement in the project dissemination to large user communities. Having already participated in many eInfrastructure and eScience projects, especially on Galaxy/CloudMan projects, members of RBI are well versed regarding the general trajectory for this project, potential problems, and viable resolution of those. Working in over 45 FP projects RBI has large experience in project dissemination activities. Every year we organizing Research Infrastructure Dissemination Event (RIDE) and apply modern multimedia and hypermedia technologies for dissemination of information and knowledge.

- Participate in numerous ESFRI and ERIC pan European project initiatives and implementations.The RBI have a widespread collaborative network with relevant associations.

- ***Prof. dr. Karolj Skala is a full professor of University of Zagreb and a senior researcher at RBI*** *He received the Ph.D. degree in Electrical Engineering in 1983 at University of Zagreb. Since 2003, he is the head of Centre for Informatics and Computing of Ruđer Bošković Institute. He is the chairman of the international scientific symposium Distributed Computing and at MIPRO convention. He lead 15 EU FP 6/7 projects oriented to distributed computing (Cluster, Grid, Cloud). He initiated the development of a national grid infrastructure CRO GRID, which is now a part of CRO NGI. For the last three years he organizes RIDE (Research Infrastructure Dissemination Event) in which modern multimedia and hypermedia technologies are utilized to disseminate the most recent both European and national projects, initiatives, and research accomplishments based on the research infrastructures. He is a member of the MIPRO and ELMAR programme committees, and a member of Croatian Academy of Technical Science and the associate member of Hungarian Academy of Science.*

P10 RBI

Dr. Enis Afgan is a research scientist at the Ruđer Bošković Institute (RBI), Croatia. *He obtained his Ph.D. in 2009 from the Department of Computer and Information Sciences at the University of Alabama at Birmingham (UAB, USA). Since 2009 he has been a member of the Galaxy Project team and since 2012 a member of the Genomics Virtual Lab project. He is a recipient of the FP7 Marie-Curie Reintegration Grant, AIS-DC. He leads on the CloudMan project. His research interests focus around distributed computing, with the current emphasis on application- and user-level accessibility of cloud computing resources.*

Dr. Branka Medved Rogina is an Associate professor of University in Zagreb and Senior Research Associate *at Division of Electronics, Laboratory for Stochastic Signals and Processes Research at RBI. From 1983. she is working at the Institute Ruđer Bošković, today as the head of Laboratory for Stochastic Signals and Processes Research in the Division of Electronics. Since 1998. she teaches at the Faculty of Electrical Engineering and Computing in Zagreb. Her main research interests are in the area of generation, detection, measurement an analysis of high-speed pulse signals in electronic and optoelectronic sensors and high speed communication systemsand DataFlow Computing.She is a collaborating member of Croatian Academy of Engineering (HATZ) since 1998.*

Dr. Davor Davidović is a research assistant at RBI and a PhD student at the Faculty of Electrical Engineering *and Computing, University of Zagreb. He obtained his MSc in 2008 from the Faculty of Science, Department of Mathematics at the University of Zagreb. His research interests are in parallel and distributed computing, with the emphasis on the grid and cloud computing, GPU programming, eigenvalue problems, and numerical mathematics. He worked on three EU FP7 grid-oriented projects: Sci Bus, EGEE-III and SEE-GRID-SCI.*

Zorislav Šojat professor of General Linguistics and Socio-Humanistic. His main field of interest is Cybernetics, *and he applies Cybernetic principles in scientific research in the area of computer science and intelligence stimulation. From 1991 till 2000 he was Assistant Professor for APL, Lisp and Robot Languages on the Department of Information Science at the Faculty of Philosophy University of Zagreb. As cyberneticist, his particular emphasis is on the exploration of the concept of intelligence and human communication, and their application to computers.He studied Linguistics and Socio-Informatics. He has published hundreds of papers in the field of linguistics, ethics, informatics, philology, computing, economics, philosophy, as well as several books. He is co-author of several worldrenowned patents.*

INDIGO - DataCloud

Next

- Revision of details from partners
- Start discussion on requirements (Peter)
  - Revise EGI experience
  - Start from VENUSC input
  - Wiki Template to be prepared by Jesus/Fernando
- Continue with DMP (Massimiliano)
- Use Cases (Alexandre)
- Dissemination plan (Karolj)
- Plan next steps (today and next two months) In particular: D2.1, D2.2
- Agree bi-weekly teleconf and short meeting in Lisbon (18-22 may)

WRAP-UP (Jesus)

EGI.eu partner profile for WP2

Peter Solagna

Senior Operations Manager – EGI.eu

People

- Peter Solagna, *Senior Operations Manager*
  - Peter is responsible for the EGI infrastructure operations, the production infrastructure and the federation enabling services. Peter has been working in EGI for 5 years, and he has extensive experience in software development and IT service management applied to high performance and distributed computing infrastructures.
- Małgorzata Krakowian, *Senior Operations Manager*
  - Małgorzata has been working in operations since 2006, working at EGI.eu since 2012. Among other activities, Małgorzata has been working on operational integration of new technologies, and requirements gathering and pilots definition for the ENVRI and EGI-InSPIRE projects.
- Yin Chen, Senior Outreach Expert
  - Yin has joined EGI.eu outreach team in 2015. Previously she was mainly working in ENVRI project, she was leading ENVRI work package responsible for user communities requirements gathering and analysis. She led the definition of the ENVRI reference model standard.
- Sara Coelho, *Communications Manager*
  - Sara leads the EGI.eu communications activities, organizing the main EGI events and she is responsible for the communications strategy of the EGI-Engage project.

Requirements gathering in EGI

- During the EGI-InSPIRE project EGI.eu has established collaborations with the EMI and IGE projects, gathering requirements from the EGI communities, and worked together with the external technology providers to prioritize and deliver the developments in production for the users.
- The high level classification of the communities involved includes: high energy physics, life sciences, humanities and other disciplines.
- EGI.eu collaborated at different levels with many initiatives also outside EGI-InSPIRE, for example: DRIHM, VERCE, BioVEL, DCH-RP, EISCAT-3D, ENVRI project, CTA, WeNMR and more.
- By supporting the deployment of use cases in the fedcloud EGI.eu worked together with 26 communities to integrate 50 use cases

INGV
Istituto Nazionale di Geofisica e Vulcanologia

Profile and role in INDIGO

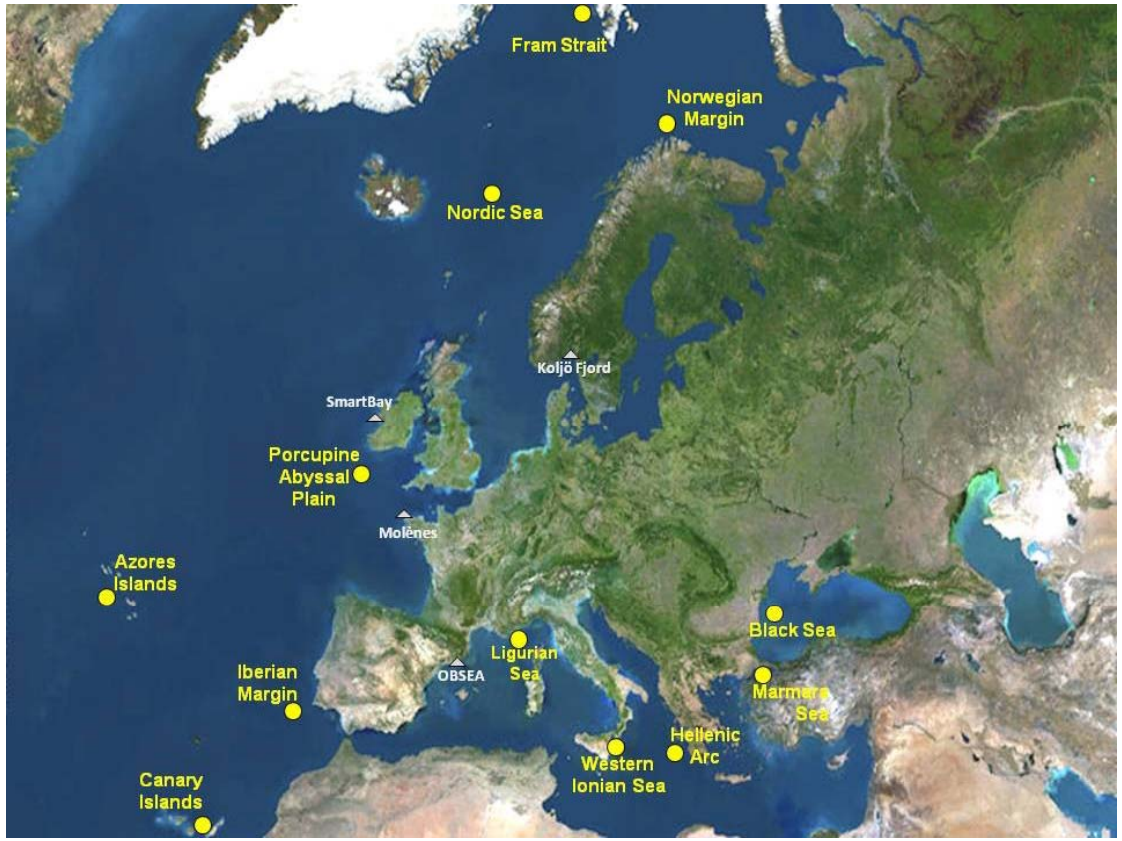
Team: Massimiliano Rossi, Lucio Badiali, Manuela Sbarra, Paolo Favali, Laura Beranzoli, Paola Materia

- INGV is one of the major European research institution in Earths Science. It has 11 sites and a staff of around 900 people 600 of which researchers, engineers. Disciplinary sector are: seismology, volcanology, geodesy, remote sensing, geochemistry, physics of the atmosphere, oceanography, marine geophysics and technological research in support to the disciplinary areas.

- INGV manages national and local sensor networks to monitor seismic, volcanic activity over the Italian territory and volcanic areas and off-shore, officially charged with 24h surveillance service by the Civil Protection Agency.

- INGV is leading two research infrastructures of ESFRI (European Strategy fForum on Reseacrh Infrasrtucture): EPOS and EMSO.
  In INDIGO INGV represents the research infrastructure EMSO.

EMSO is the European Multidisciplinary Seafloor and water column Observatory: a network of marine monitoring observatories around the European plate located in key sites for the comprehension of local and global scale, short-term and long-term deep sea processes.



The main research areas addressed by EMSO are:
Natural Hazards,

Marine Ecosystems

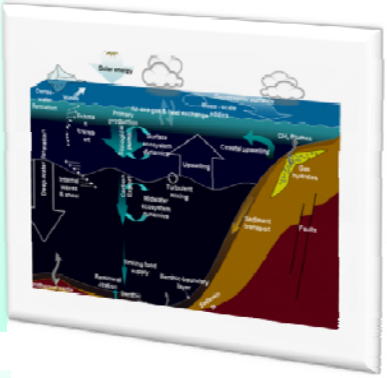
Climate Changes

EMSO Challange:
Sustained time-series of physical parameters over decades

BIG DATA

EMSO Key Scientific Objectives

Geosciences
- Seismicity
- Gas hydrate stability
- Seabed fluid flow
- Submarine landslides
- Submarine volcanism
- Geo-hazard early warning



Physical Oceanography
- Ocean warming
- Deep-ocean circulation
- Benthic and water column interactions
- Marine forecasting

Biogeochemistry
- Ocean acidification & Solubility pump
- Biological pump
- Hypoxia
- Deep-ocean biogeochemical fluxes
- Continental shelf pump

Marine Ecology
- Climate forcing of ecosystems
- Molecules to microbes
- Fisheries
- Marine noise
- Deep biosphere
- Chemosynthetic ecology

Task T2.2: Defining support to Research Data

Lead partner: **INGV [18 PMs], deputy: CSIC [10 PMs]**

Contributors: CNR [12 PMs], ICCU [9 PMs]

To guarantee a smooth and widespread usability of INDIGO, an appropriate integration and combination approach has to take into account the different Reference Models used by the Research Communities and Research Infrastructures and the diversity and heterogeneities of data services and catalogues.

This task follows the data research use and management of the Research Communities and Research Infrastructures and points out the different needs at the data life-cycle level.

In particular this task shall undertake a survey on the research communities to collect and analyze the individual Data Management Plans (DMP) and data-life-cycle documentation with the aim to ensure that the full data cycle and components will be supported in INDIGO, and with the aim to provide adequate specifications for the compliance with INDIGO.

Accordingly, the following activities are foreseen:

- **Development** of individual search activities to acquire and analyze the available DMP of the research communities/infrastructures with special attention to distributed/heterogeneous data services and catalogues, and to available open data;
- **Acquisition** of procedure details/parameters (i.e., DMP, Collection, Authenticity & Provenance, Data Preservation) to elaborate the specifications for data ingestion and use in INDIGO;
- **Definition** of the specifications of INDIGO ingestion integrity test.

Task T2.2: Defining support to Research Data

Development of individual search activities.
Distribution of questionnaire with few questions that each partner can circulate in its community for collecting information about the available Data Manag. Plan.
Target ⟹ Partnership and eventual external entities

Acquisition of procedure details/parameters.
Distribution of **survey** on procedure details/parameters to address the specifications for data ingestion and use in INDIGO.
Target ⟹ Partnership

Definition of the specifications of INDIGO ingestion integrity test.
Verify INDIGO does not alterate the ingested data