



INDIGO - DataCloud

# WP2 Summary Progress in Frascati AHM

Reported by Jesus Marco

marco@ifca.unican.es

See <https://project.indigo-datacloud.eu/projects/wp2/wiki/AHM> - Frascati - May 2016



INDIGO-DataCloud is co-funded by the  
Horizon 2020 Framework Programme

# Summary of WP2 sessions



- **Wednesday:**
  - Review Case Studies “in 2 slides” (all Research Communities) POSITIVE (see next)
    - Include new Use Cases from HEP community (thanks to Isabel Campos and Daniele Spiga)
  - Brief “hint of technical possibilities”
    - Deploying applications under the Cloud framework (thanks to Nacho Blanquer)
    - IAM linking to web based applications (thanks to Fernando Aguilar)
- **Thursday morning: ON DATA MANAGEMENT**
  - Starting an analysis of RDA outcomes to be able to improve the DATA requirements (in D2.1)
- **Thursday afternoon: TASKS AND DELIVERABLES**
  - D2.5 (Outreach and Dissemination Report, June) (thanks to Sara and Eva)
  - D2.6 (Exploitation Strategy and Sustainability, End of 2016)
  - D2.7 (Specifications of Data Ingestion and Use, Spring 2017)

# Where were we after AMS meeting?

- All Research Communities have defined their Champions
  - <https://www.indigo-datacloud.eu/champions>
- Variety of Profiles !!!
  - Share other activities inside INDIGO (in particular in JRA)
  - Advanced research activities (postdoc+ level)
  - Different Technical/Scientific background/interest
- All are committed with the project!
  - D2.4 requirements prepared with their input
  - OpenProject info: Case Studies, User Stories, Requirements
  - Detailed slides describing Case Studies
    - <https://project.indigo-datacloud.eu/projects/wp2/wiki/D24>
  - Dissemination info about Champions: <https://www.indigo-datacloud.eu/Champions>
  - TWO SLIDES ON IMPLEMENTATION USING INDIGO COMPONENTS here:
    - <https://project.indigo-datacloud.eu/projects/wp2/wiki/AMS-Meeting>

# List of Case Studies reviewed in “2 slides” to start linking to INDIGO solutions



- *Case Study P0\_1: Monitoring and Modelling Algae Bloom in a Water Reservoir (FERNANDO)*
- *Case Study P0\_2: TRUFA (Transcriptomes User-Friendly Analysis) (idem)*
- *Case Study P1: Medical Imaging Biobanks (IGNACIO)*
- *Case Study P2: Molecular Dynamics Simulations (GLEB / ANTONIO)*
- *Case Study P3\_1: Astronomical Data Archives (CRISTINA)*
- *Case Study P3\_2: Archive System for the Cherenkov Telescope Array (CTA) (STEFANO)*
- *Case Study P4\_1: HADDOCK Portal (ZEYNEP)*
- *Case Study P4\_2: DisVis (idem)*
- *Case Study P4\_3: PowerFit (idem)*
- *Case Study P5: Climate models inter comparison data analysis (SANDRO)*
- *Case Study P6: eCulture Science Gateway (LUCCA)*
- *Case Study P7: EGI FedCloud Community Requirements (YIN)*
- *Case Study P8: ELIXIR-ITA: Galaxy as a Cloud Service (MARCO)*
- *Case Study P9: MOIST—Multidisciplinary Oceanic Information System (STEPHEN)*
- *Case Study P10: Data repository platform for DARIAH (EVA, DAVOR)*

# D2.5 (M15, end of June)

## Actions by June to be included in D2.5 (Sara proposal) :

- Preparation of **a template message** to be sent to the champions in order to ask them to distribute it to their communities via newsletters, website, mailinglists and social media channels - deadline 20 May - Trust-IT + RBI
  - Collection of feedbacks from the champions on the messages sent / published - deadline 10 June - RBI
  - New press release** - Trust-IT - deadline end of May
  - Preparation of a new generic flier - Trust-IT - deadline 10 June
  - Preparation of a demo plan** - RBI - deadline 10 June
  - First online catalogue of INDIGO solutions published online - June
- Internal communication: Proposal to create a slack channel for Indigo - message to wp1 to organize this
- In addition, there are a couple of points more that can be added to the table of content of the deliverable:
1. Analysis of the current database of people authenticated on INDIGO developing tools - Davide will provide the DB
  2. report from the presentations from the Industry players that will be done tomorrow - stating that XX companies already engaged to take on INDIGO solutions is an important achievement so far.

**-Courses, training, plans for Krakow and Summer course in Catania**

# Related Champions' tasks

| Task  | When                                     | Who   | Extra   |
|---|--|---|---|
| Check D2.2, reconfirm communication channels, <b>and the events</b> , etc.  | May 12, next WP2                         | All champions   | Try to see if there is a mailing list, newsletter, etc. in your community Exploit Competence centers or similar |
| <b>Confirm that your application/Cases Study is ready for a demo (in mid September) and identify Indigo solutions that can be demonstrated to be discussed with JRA Reserve week 5-9<sup>th</sup> sept for pre-REVIEW ad-hoc meeting(s)</b> | <b>Today</b>                             | <b>1<sup>st</sup> level: ELIXIR, ENES, Powerfit Disvis, AMBER<br/>2<sup>nd</sup> level:<br/>EUROBIOIMAGING,<br/>LifeWatch, CTA, HADDOCK,<br/>EGI with AAI</b> | <b>IDENTIFY ADDED VALUE (for next WP2 meeting)</b>  |
| Identify how Indigo solutions released (list will be frozen by end of may) will be useful (i.e. continue the work on 2 slides and then link with the requirements in Open project)  | 26 <sup>th</sup> may (or so) WP2 meeting | All champions   | Look to “solutions” (see slides today from Davide) and services (check with Giacinto)                           |
| Test the proposed “communication channel” (slack or whatever)   | End of may (29 <sup>th</sup> )           | All   | Details and tool to be provided by Indigo management/Trust-IT   |
| Have a look to FAQ  | Next thursday                            | All   |   |

# Related Champions' tasks cont'd



| Task   | When  | Who                | Extra |
|--|---|--------------------|-------|
| Interest of a demo/submission for Krakow (deadline 3 <sup>rd</sup> June)           | Next week: interest, End of may proposal  | All                |       |
| Application calls for Catania Summer School (Hackathon like)(into Science Gateway) | End of may  | All                |       |
|  |   |                    |       |
| Think about training online material/short courses                                 | End of may (volunteer found) Work in Catania (July to record video, collaboration with WP6) Before September (an example ready) | Volunteers welcome |       |
| Have a look to FAQ   | Next thursday   | All                |       |



## D2.6 (by M21, end of year)



- **D2.6, Month 21 (R,CO): Exploitation strategy and sustainability**
- 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

*Involvement of EGI.eu (Matthew), correlation with WP3*



# D2.6 (by M21, end of year)

- 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.  
**(COMMENT: IDENTIFY SYNERGIES-COMPETITION WIT OTHER PROJECTS, AND FIND A FORUM TO DISCUSS WITH THEM)**
- 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) **(Consider adequate Forums at EU wide level , exploit EGI-eu connections...)**
- Eventually, consider the creation of an INDIGO Open Source consortium and promote the involvement of European companies to support key developments **(EU Open Science Cloud framework)**
- **EXPLOIT OTHER STAKEHOLDERS (LIKE ESA)**
- **LINKS WITH WP3 AND RESOURCE PROVIDERS**

# Related Champions' tasks



INDIGO - DataCloud

| Task  | When                     | Who                           | Extra  |
|---|--------------------------|-------------------------------|--|
| After INDIGO solutions, confront them with other ongoing work                           | Before review (mid sept) | In collaboration with seniors | The discuss in Krakow ith others (so we see how we fit in the ecosystem) |
| Find out commercial partners (start from outcome of Frascati)                           | Mid June                 | In collaboration with seniors | Then, in Spetember, define the channels                                  |
| Identify explicitly international events and contacts (already in the planning of D2.2) | May 12 <sup>th</sup>     | All (see also D2.5 actions)   | Need to define stable collaboration channels                             |
| Identify the resource providers as requested by WP3                                     | TODAY                    | All                           | Completed, confirm for CTA   |

# D2.7 (M24 March 2017)

- **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.

## T2.2

- 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.

## D2.7 (M24 March 2017)

- 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.**

# Related Champions' tasks



INDIGO - DataCloud

| Task  | When          | Who  | Extra   |
|---|---------------|--|---|
| Review Data Management section in D2.1 and update it                                | End of June   | All champions  | First, background on Data Management, RDA, etc. |
| Complete the background on Data management techniques (formats, etc)                | End of May    | Champions should read/complement the info prepared by T2.2 |   |
| Identify ADVANCED Case Studies (where DM includes using linked data, ontologies...) | Mid September | Volunteers   |   |
| BIG DATA CASE STIUDIES  | Mid September | Volunteers   |   |
| INGESTION FROM MULTIPLE SOURCES- INTEGRITY TEST (Data Quality as a Service)         | ?             | ?  |   |
| LEGAL ISSUES (LICENCES)   | ?             | ?  | Consider RDA draft                              |

# What to discuss with other WP

- Need yet to do!
  - Lets hear next reports
  
- See also WP2-WP6 session today (but notice some champions leaving)
  - Title: Scientific data analysis use cases with Ophidia
  - The presentation will highlight practical use cases regarding scientific data analysis.
  - Workflow examples will be presented. Batch and interactive experiments/sessions will be presented.
  
- Title: Using FutureGatewayAPIs
  - The presentation will highlight how to use the FutureGateway APIs, for integrating the FG functionality within the domain specific portals.
  
- Title: FutureGateway Installation
  - Description: How to install the FG