RIA-653549

INDIGO – DataCloud Kick-off Meeting

The Project

Davide.Salomoni@cnaf.infn.it

INFN-CNAF

Bologna, 22-24/4/2015

A brief history of the project

- 28th April '14: first technical ideas presented by INFN to some WLCG agencies at CERN
- June '14: first phone meetings of the "DataCloud" Consortium with ~10-15 partners
- 27th June '14: first face-to-face meeting in Rome, Consortium has ~20 partners
- 1st August '14: second face-to-face meeting at CERN, the "INDIGO-DataCloud" Consortium has now 26 partners
- 2nd September '14: the INDIGO-DataCloud proposal (duration 30 months, budget 11.1 M€) is submitted to the EINFRA-1-2014 call
- 14th January '15: Communication by the EC that INDIGO was favorably evaluated, invitation to the preparation of the Grant Agreement
- 1st April '15: official starting date of INDIGO
- 15th April '15: the EC signed the Grant Agreement
- 22-24 April '15: INDIGO-DataCloud kick-off meeting in Bologna
- ... and many, many other meetings, phone calls, travels in between

INDIGO - DataCloud

The context

- Since the very beginning we identified **key issues with both Grid and Cloud technologies** that prevented scientific communities an easy and optimal exploitation of data and compute resources. (see Giacinto's presentation for details)
- We therefore decided to propose the development of a **software platform** centered around two of the EINFRA-1-2014 pillars:
  - **Large scale virtualization of data/compute center resources**. This became the focus of INDIGO WP4.
  - **Development and adoption of a standards-based computing platform (with an open software stack)**. This became the focus of INDIGO WP5.

EINFRA-1-2014 items addressed by INDIGO

INDIGO - DataCloud

- **Item 4:**
  - Large scale virtualization of data/compute centre resources to achieve on-demand compute capacities, improve flexibility for data analysis and avoid unnecessary costly large data transfers.
- **Item 5:**
  - Development and adoption of a standards-based computing platform (with open software stack) that can be deployed on different hardware and e-infrastructures (such as clouds providing infrastructure-as-a-service (IaaS), HPC, grid infrastructures...) to abstract application development and execution from available (possibly remote) computing systems. This platform should be capable of federating multiple commercial and/or public cloud resources or services and deliver Platform-as-a-Service (PaaS) adapted to the scientific community with a short learning curve. Adequate coordination and interoperability with existing e-infrastructures (including GÉANT, EGI, PRACE and others) is recommended.

Some starting observations (Apr 2014)

- **Grid**:
  - Long, steep learning curve.
  - SSO based on X.509 certificates.
  - Difficult to use for real-time analysis, visualization, provisioning of complex virtual environments.
  - Storage management normally at the file/block level, not as distributed objects.
  - Data preservation?
  - Migration to / interworking with Cloud computing/storage?
- **Cloud**:
  - Simple-minded scheduling policies in IaaS infrastructures.
  - Complexity in the exploitation of distributed Cloud resources due to multiple IaaS and/or PaaS solutions.
  - Data preservation? Access? Use?

The next steps for distributed computing

1. Ease of access and use for small and big collaborations alike.
2. Software and economic sustainability.
3. Robustness (no single points of failure).
4. Modular, scalable architecture.
5. Open source software, vendor independence, hybrid infrastructures.

INDIGO - DataCloud

Our objective

- The development of an **easy-to-use, Cloud-based open source software platform**, without restrictions on the e-Infrastructure(s) to be accessed (public or commercial, GRID/Cloud/HPC) or its underlying software, **targeted at scientific communities**, addressing current technology gaps linked to specific use cases.
- Wherever possible, we will **exploit existing solutions**, learning, re-using and extending them according to user requirements, and having in mind the expected evolution of technology.

INDIGO - DataCloud

The Work Packages (WP)

- They are divided among Network Activity (NA), Service Activity (SA) and Joint Research Activity (JRA).
  - **WP1 (NA)**: admin and financial management, project quality assurance, global oversight.
  - **WP2 (NA)**: this is where Research Communities express requirements, provide feedback, review deployed services. This WP also includes dissemination and communication activities.
  - **WP3 (SA)**: software management, deployment of pilot services.
  - **WP4 (JRA)**: resource virtualization.
  - **WP5 (JRA)**: PaaS framework.
  - **WP6 (JRA)**: APIs and portals, data-driven workflows.

The INDIGO Consortium

- "The consortium is of exceptional quality, and complementary, and with good relevant experience and skills." (from the EC Evaluation Summary Report) – see Luciano's presentation.
- 26 members from 11 European countries, including 4 big industrial partners.
- The expertise of the INDIGO Consortium builds on the body of work of several research groups with **significant and unrivalled experience** in middleware development for production-class distributed e-infrastructures. The quality of their achievements is demonstrated by the **results of previous projects and initiatives carried out in Europe across the past ten years of collaborations on scientific computing** based on different consolidated and emerging paradigms (HPC, Grid and Cloud).

INDIGO - DataCloud

Work Program

- 30 months, starting April 1st, 2015
- 44 deliverables, review at M15
- **In the first six months**:
  - M2: quality plan (CSIC)
  - M3: communication plan (INFN), HR allocation (INFN), initial req's from Research Communities (EGI.eu), communication plan (INFN), dissemination plan (RBI)
  - M4: risk analysis (INFN), WP3 plan (LIP)
  - M5: WP5/PaaS state of the art (AGH)
  - M6: general architecture (INFN), WP4 software design and work plan (DESY), WP5 design architecture and work plan (AGH), WP6 design architecture (PSNC)

Governance

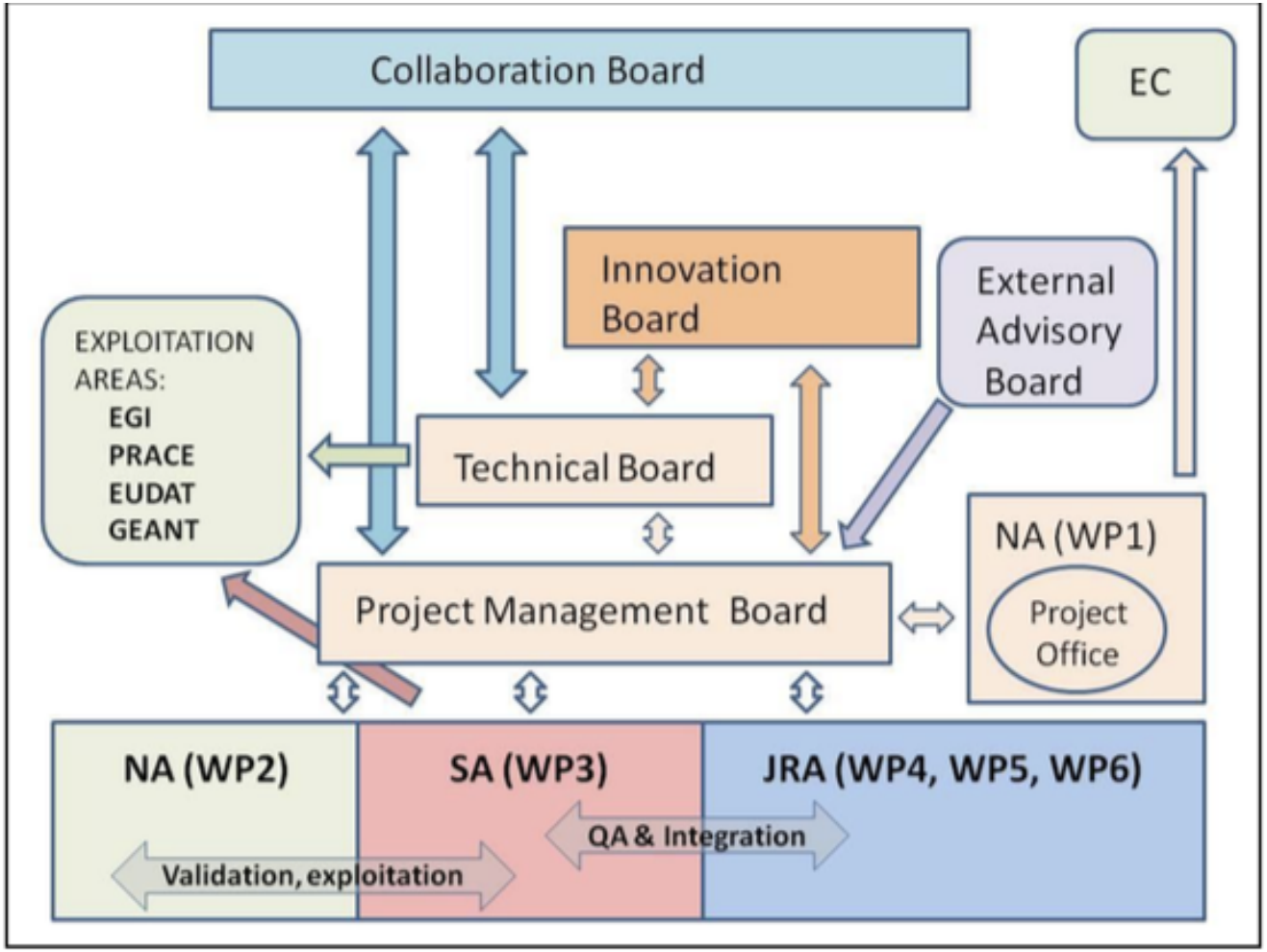

Figure 16 INDIGO Governance structure

INDIGO - DataCloud

A key point: *agility*

- We are working in a fast-moving technology sector.
- **We need to be agile** in all areas:
  - In **capturing user requirements** and in understanding their technological implications
  - In **building and sharing know-how** across the partners
  - In **prototyping solutions and test-beds**, letting early adopters validate them as early as possible
  - In **deploying our software** to diverse e-infrastructures
  - In **communicating preliminary results** through software releases, conferences, workshops, publications, deliverables
  - In **talking to user communities and projects** outside the INDIGO Consortium, agreeing on common actions and synergy
- There are clearly **high expectations** from INDIGO. We have the potential to make this a very successful project. **Let's do it.**

INDIGO - DataCloud

Thank you!