INDIGO - DataCloud

RIA-653549

The INDIGO-DataCloud Software Platform

Lightning Talk - GARR Workshop 2016

Better Software for Better Science.

Davide Salomoni, INFN-CNAF
INDIGO-DataCloud Project Coordinator

Giacinto Donvito, INFN-Bari
INDIGO-DataCloud Technical Director

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

INDIGO-DataCloud

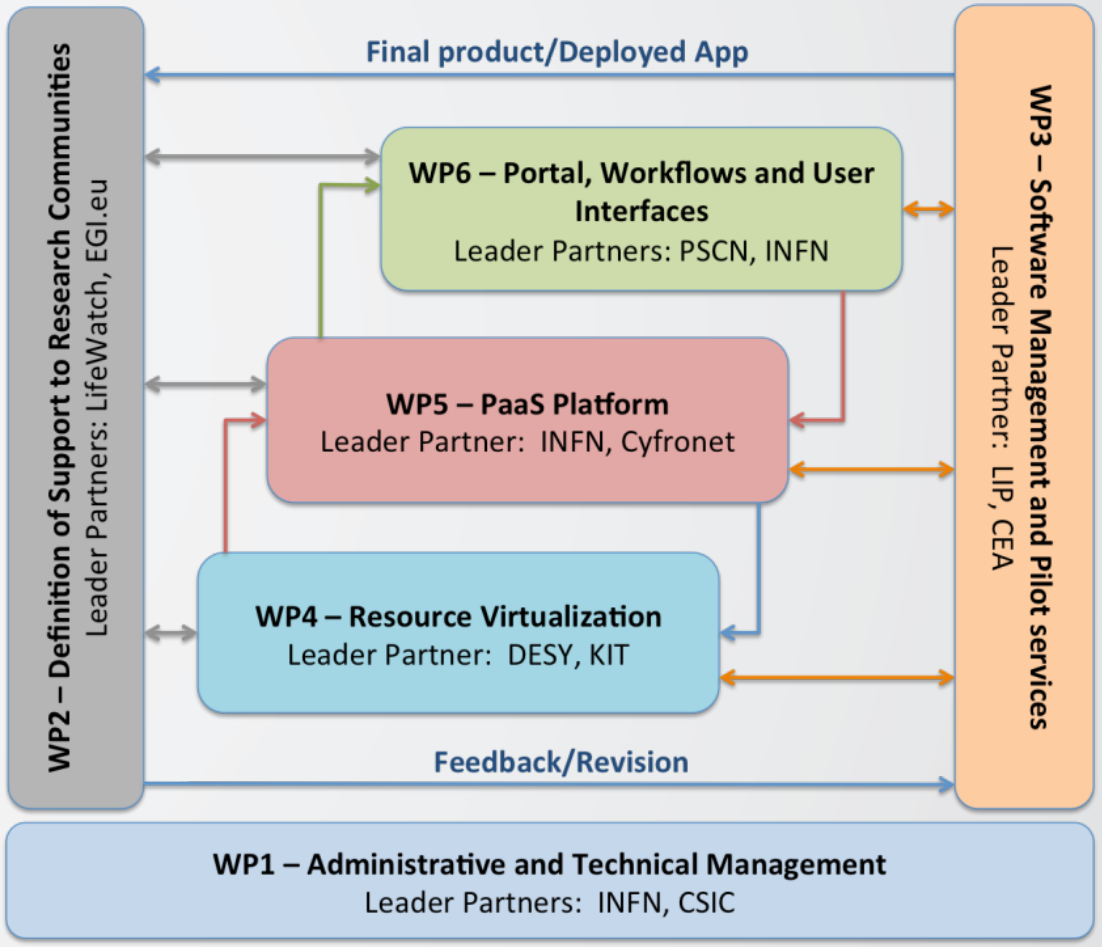
- **An H2020 project** approved in January 2015 in the EINFRA-1-2014 call
  - 11.1M€, 30 months (**from April 2015 to September 2017**)
- **Who**: **26 European partners** in 11 European countries
  - Coordination by the Italian National Institute for Nuclear Physics (INFN)
  - Including developers of distributed software, industrial partners, research institutes, universities, e-infrastructures
- **What**: **develop an open source Cloud platform** for computing and data ("DataCloud") tailored to science.
- **For**: **multi-disciplinary scientific communities**
  - E.g. structural biology, earth science, physics, bioinformatics, cultural heritage, astrophysics, life science, climatology
- **Where**: deployable on **hybrid (public or private) Cloud infrastructures**
  - INDIGO = **IN**tegrating **D**istributed data **I**nfrastructures for **G**lobal Expl**O**itation
- **Why**: answer to the technological **needs of scientists** seeking to easily exploit distributed Cloud/Grid compute and data resources.

INDIGO Addresses Cloud Gaps

- **INDIGO focuses on use cases presented by its scientific communities** to address current Cloud data / computing gaps, with regard to:
  - Redundancy / reliability
  - Scalability (elasticity)
  - Resource utilization
  - Multi-tenancy issues
  - Lock-in
  - Moving to the Cloud
  - Data challenges: streaming, multimedia, big data
  - Performance
- **Reusing existing open source components** wherever possible and **contributing to upstream projects** (such as OpenStack, OpenNebula, Galaxy, etc.) for sustainability.

See for example this EC Expert Group Report on Cloud Computing:
http://cordis.europa.eu/fp7/ict/ssai/docs/future-cc-2may-finalreport-experts.pdf

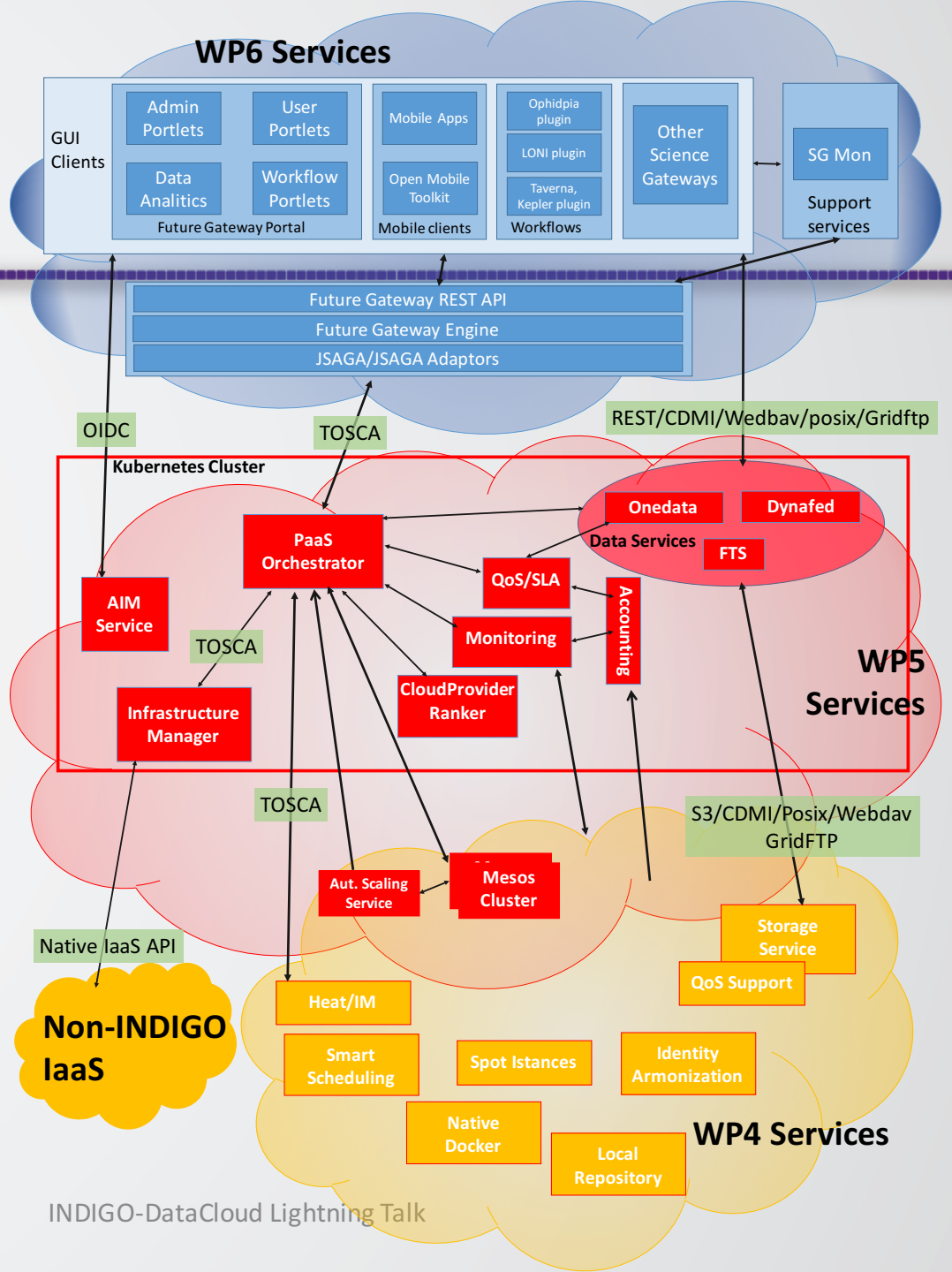
The INDIGO Work Package Structure



The INDIGO Solutions

- **The INDIGO architecture can be seen as providing:**
  - **Site-level solutions** → improved IaaS frameworks
  - **Data solutions** → unified data access to e-Infra, Cloud and personal storage
  - **Automated solutions** → automating provisioning of complex services
  - **User-level solutions** → scientific gateways, comprehensive AAI, workflows
- All of them integrate in a consistent global framework. Frequently a given solution spans multiple INDIGO WPs.
- See http://arxiv.org/abs/1603.09536 for a general description of the INDIGO architecture, or refer to the detailed deliverables published on https://www.indigo-datacloud.eu.

The INDIGO Architecture



Conclusions

- INDIGO is **working together with many scientific communities and industrial partners** to build efficient and sustainable ways for the exploitation of Cloud computing and data resources.
- **The first public INDIGO software release is due by July 2016**; the second release is due by March 2017.
- The INDIGO Consortium is **open to collaborations with other projects and external partners and actively participates in both national and international initiatives** (among them, the European Open Science Cloud and the European Data Infrastructure).
- **At the national level in particular**, we are keen to let Italian communities test and contribute to the software once the first public release is out, either directly or through our test-beds.

Thank you for your attention!

https://www.indigo-datacloud.eu

Better Software for Better Science.