RIA-653549

The INDIGO-DataCloud Project

Davide Salomoni
INDIGO-DataCloud Project Coordinator
1st Asterics/Obelics Workshop, Dec 2016
davide.salomoni@cnaf.infn.it

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

From the Paper “Advances in Cloud”

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

To reach the full promises of CLOUD computing, major aspects have not yet been developed and realised and in some cases not even researched. Prominent among these are **open interoperation across (proprietary) CLOUD solutions at IaaS, PaaS and SaaS levels**. A second issue is **managing multitenancy** at large scale and in heterogeneous environments. A third is **dynamic and seamless elasticity** from in- house CLOUD to public CLOUDs for unusual (scale, complexity) and/or infrequent requirements. A fourth is **data management in a CLOUD environment**: bandwidth may not permit shipping data to the CLOUD environment and there are many associated legal problems concerning security and privacy. All these challenges are opportunities towards a more powerful CLOUD ecosystem.

[…] **A major opportunity for Europe involves finding a SaaS interoperable solution across multiple CLOUD platforms.**

Another lies in migrating legacy applications without losing the benefits of the CLOUD, i.e. exploiting the main characteristics, such as elasticity etc.

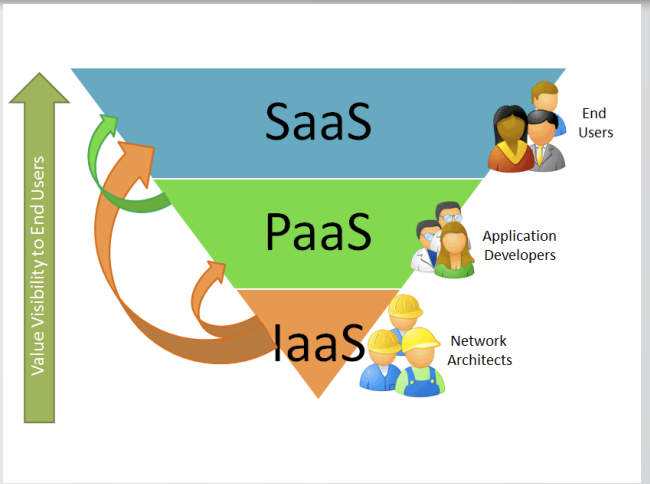
Highlighting the main points

INDIGO - DataCloud

- What is missing:
  - Open **interoperation** / federation across (proprietary) CLOUD solutions at
    - IaaS,
    - PaaS,
    - and SaaS levels
  - Managing **multitenancy**
    - At large scale…
    - … and in heterogeneous environments
  - Dynamic and seamless **elasticity**
    - For both private and public cloud…
    - … and for complex or infrequent requirements
  - **Data management** in a Cloud environment
    - Due to technical…
    - … as well as to legal problems

Filling these gaps should lead to:

- **Interoperable PaaS/SaaS solutions exploiting both public and private Cloud infrastructures**
- **Migration of legacy applications to the Cloud**

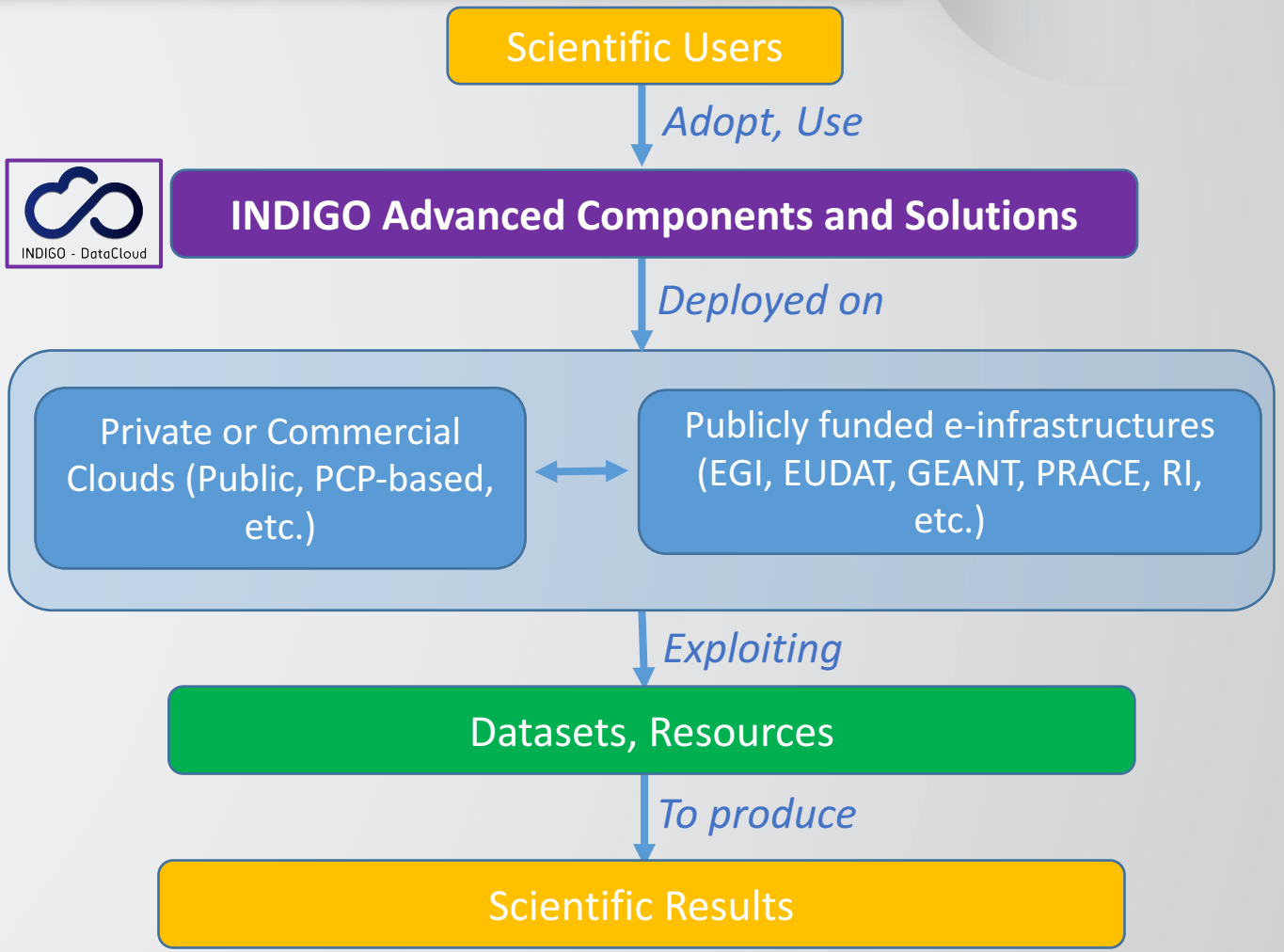


INDIGO-DataCloud

- **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-DataCloud Positioning

- INDIGO aims to:
  1. **Develop open, interoperable solutions for scientific data.**
  2. **Support open science** organizing the European data space.
  3. **Enable collaborations** across diverse scientific communities worldwide.
- INDIGO offers its architecture, analysis, expertise and software components as a **concrete step toward the definition and implementation of a [European] Open Science Cloud and Data Infrastructure**.



Putting Users First

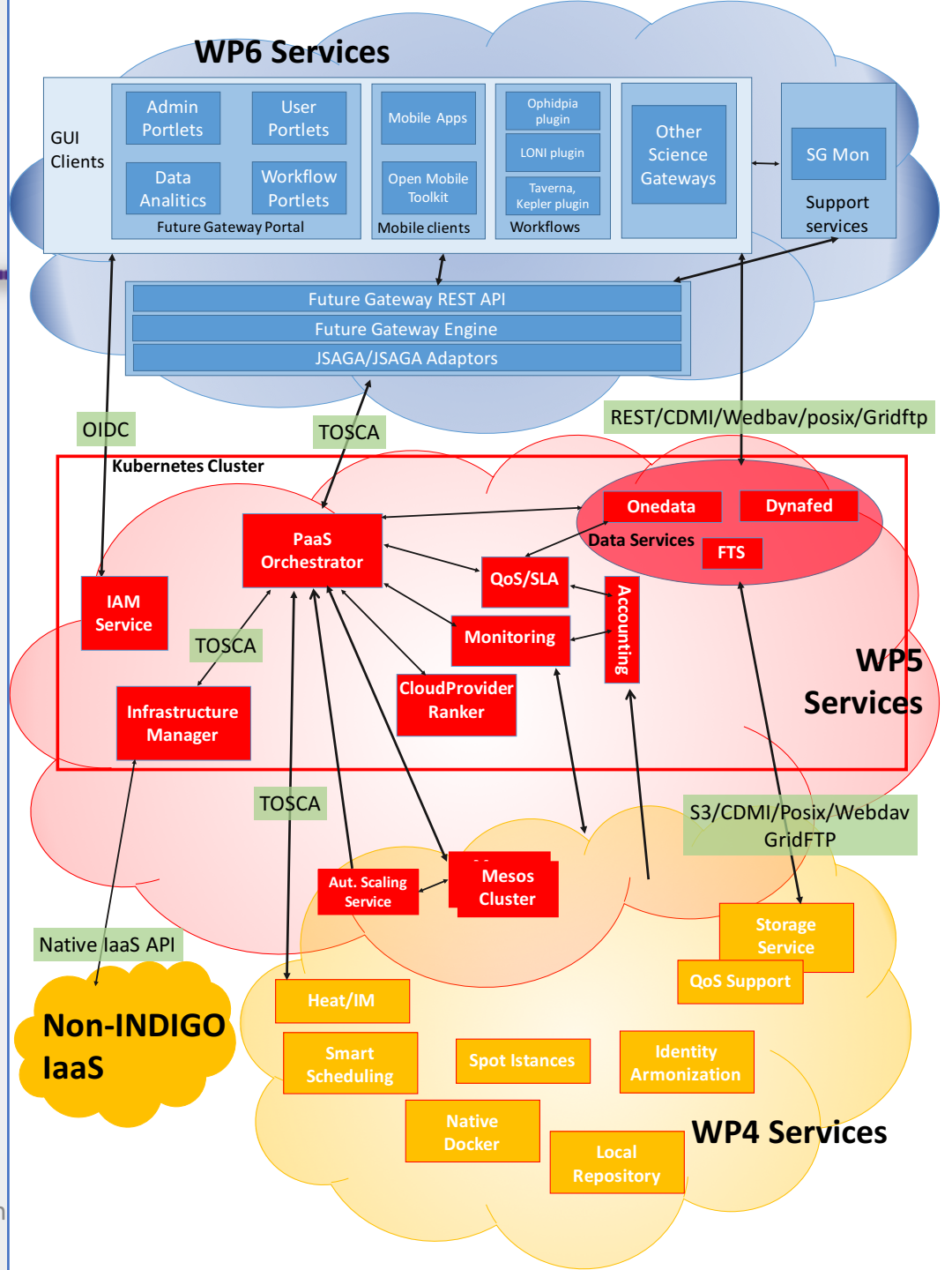
- **Requirements come from research communities**
  - "The proposal is oriented to support the use of different e-infrastructures by a wide-range of scientific communities, and aims to address a wide range of challenging requirements posed by leading-edge research activities" (From the DoW)
- **We gathered use cases** from many scientific communities.
  - LifeWatch, EuroBioImaging, INSTRUCT, LBT, CTA, WeNMR, ENES, eCulture Science Gateway, ELIXIR, EMSO, DARIAH, WLCG.
- We grouped ~100 distinct requirements **into 3 categories: Computational requirements, Storage requirements, Requirements on infrastructures**, and associated each one with a ranking (mandatory / convenient / optional).

From Deliverable D2.1

#REQ	Description
CO#1	Deployment of Interface SaaS
CO#2	Deployment of Customized computing back-ends as batch queues
CO#3	Deployment of user-specific software
CO#4	Automatic elasticity of computing batch queues
CO#5	Terminal access to the resources.
CO#6	Privileged access
CO#7	Execution of workflows
CO#8	Provenance information
CO#9	Cloud bursting
CO#10	Data-aware scheduling
CO#11	Provisioning of efficient Big Data Analysis solutions exploiting server-side and declarative approaches
CO#12	Execution across multiple centres.
CO#13	On-line processing of data
CO#14	Special hw configuration - MPI, multicore, GPGPU
SO#1	Shared storage accessible like a POSIX filesystem
SO#2	Persistent data storage
SO#3	Long-term availability of results
SO#4	Local user storage
SO#5	Availability of reference data
SO#6	Interoperability with application domain specific software and services (e.g. IS-ENES/ESGF)
SO#7	Metadata management / Database as a Service
SO#8	Share data capabilities
SO#9	Data replication
SO#10	Distributed storage
SO#11	Dropbox-like storage
PL#1	Global-level AAI
PL#2	On-line access to data
PL#3	Network configuration
PL#4	Monitoring and operation

Translating requirements into concrete solutions:
From the architecture…

This is the **INDIGO-DataCloud General Architecture***

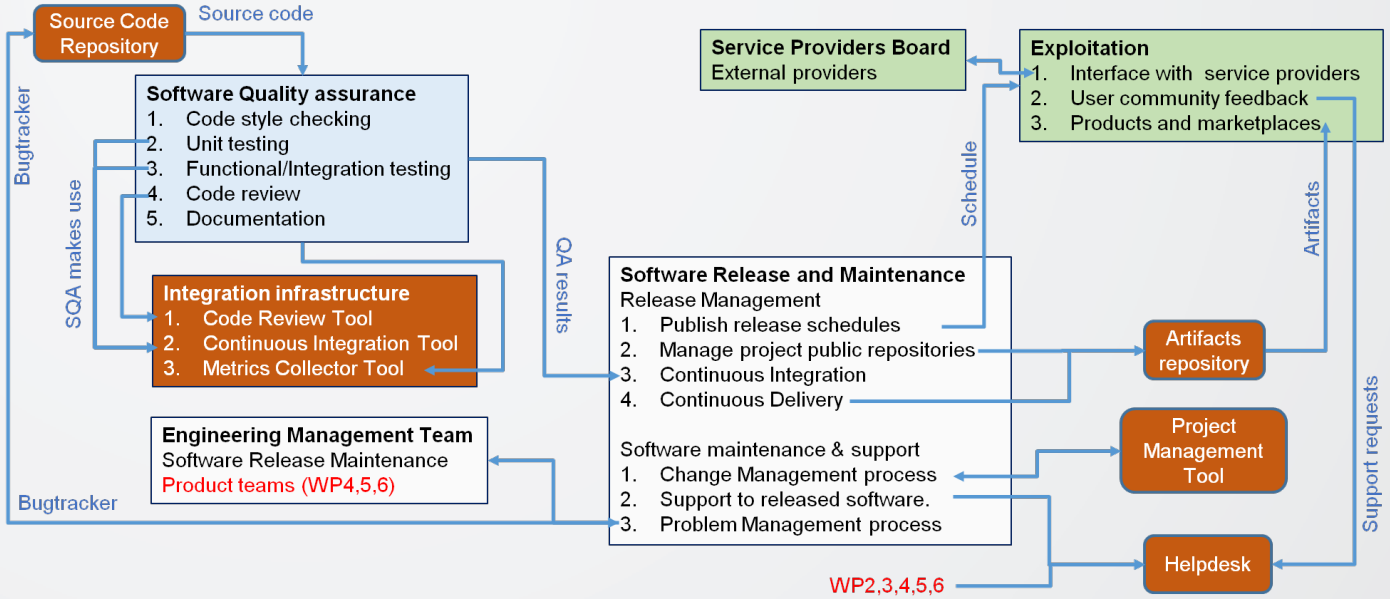


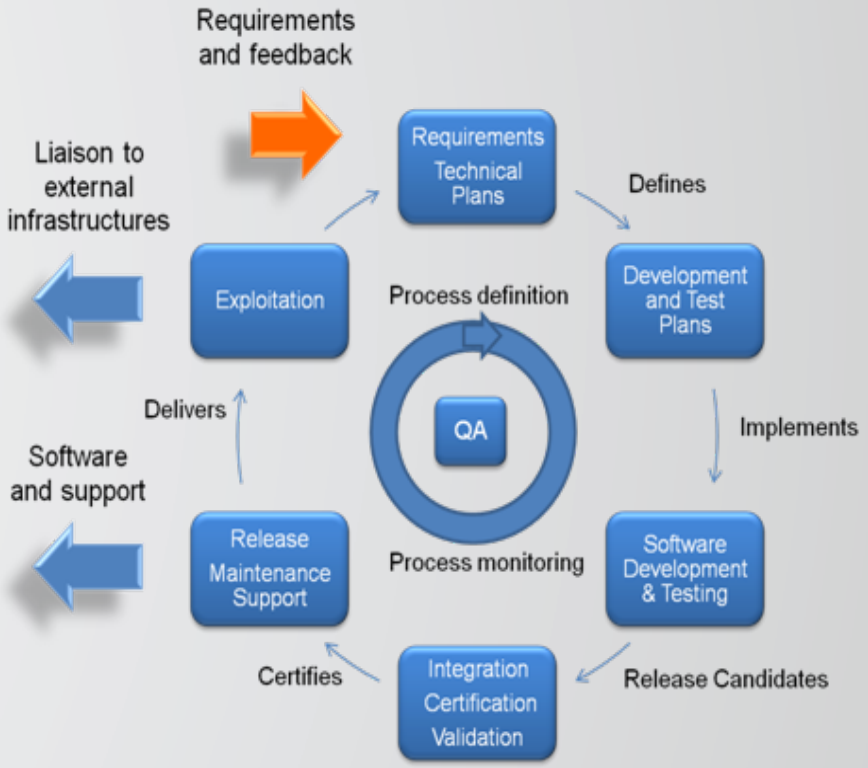
*: see details in http://arxiv.org/abs/1603.09536 or in https://www.indigo-datacloud.eu/documents-deliverables

… to the implementation…

INDIGO - DataCloud

This is our **software improvement cycle** and the **integration / release / software quality processes**





See https://www.indigo-datacloud.eu/communication-kit

… to INDIGO Releases…

Releasing software components implementing the INDIGO architecture and **providing concrete solutions** to the requirements of scientific communities **is the primary goal of the project.**

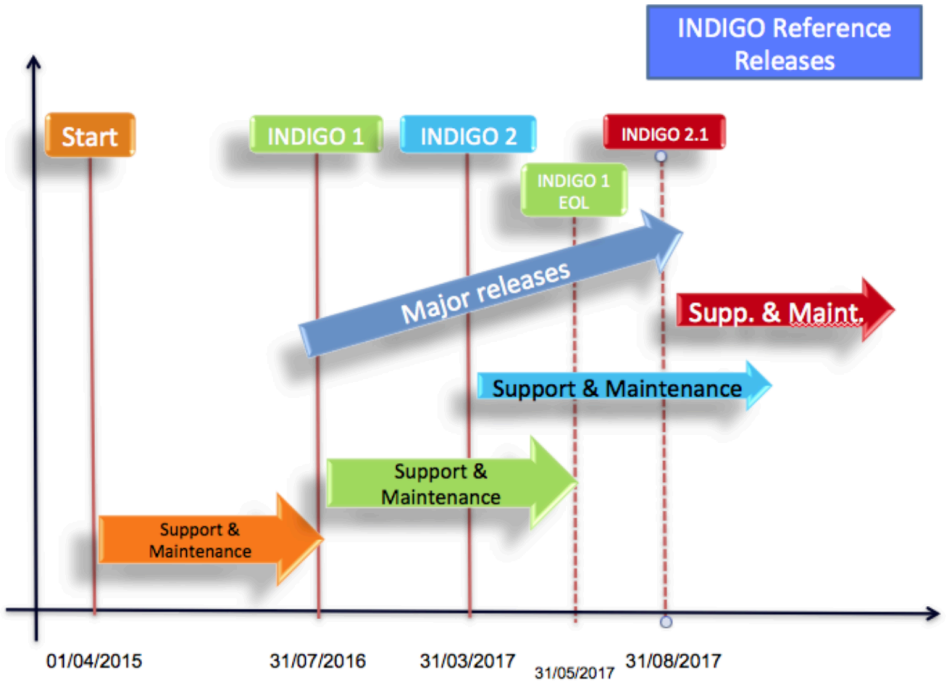


Figure 8: INDIGO-DataCloud Release Timeline

INDIGO - DataCloud

Better Software for Better Science

INDIGO MidnightBlue Service Catalogue

An unmatched open modular suite of software components for Data and Cloud computing is now available for resource providers and researchers from all disciplines, all around Europe

www.indigo-datacloud.eu

… and results.

Excerpt from an INDIGO Report detailing **how scientific communities are implementing their own requirements into applications using INDIGO-DataCloud components.**

From Deliverable D2.10

4 INTEGRATING INDIGO SERVICES INTO SOLUTIONS IN CASE STUDIES 9
4.1 INTEGRATING INDIGO SERVICE COMPONENTS IN CASE STUDIES OF RESEARCH COMMUNITIES 11
4.1.1 Case Study P0_1: Monitoring and Modelling Algae Bloom in a Water Reservoir 11
4.1.2 Case Study P0_2: TRUFA (Transcriptomes User-Friendly Analysis) 13
4.1.3 Case Study P1: Medical Imaging Biobanks 15
4.1.4 Case Study P2: Molecular Dynamics Simulations 17
4.1.5 Case Study P3_1: Astronomical Data Archives 19
4.1.6 Case Study P3_2: Archive System for the Cherenkov Telescope Array (CTA) 20
4.1.7 Case Study P4_1: HADDOCK Portal 22
4.1.8 Case Study P4_2: DisVis 25
4.1.9 Case Study P4_3: PowerFit 26
4.1.10 Case Study P5: Climate models inter comparison data analysis 26
4.1.11 Case Study P6: eCulture Science Gateway 31
4.1.12 Case Study P7: EGI FedCloud Community Requirements 34
4.1.13 Case Study P8: ELIXIR-ITA: Galaxy as a Cloud Service 36
4.1.14 Case Study P9: MOIST—Multidisciplinary Oceanic Information System 38
4.1.15 Case Study P10: Data repository platform for DARIAH 40
5 REQUIREMENTS ADDRESSED IN THE FIRST INDIGO RELEASE 42
6 EXTENDED LIST OF ADDITIONAL REQUIREMENTS 49
6.1 INITIAL LIST OF COMMUNITIES' REQUIREMENTS TO BE INCLUDED IN INDIGO FUTURE RELEASES 49
6.2 EXTENDED LIST OF COMMUNITIES' REQUIREMENTS ON AAI 49
6.3 EXTENDED LIST OF REQUIREMENTS: NEW CASE STUDIES 50
6.3.1 Cast Study P11: Seamless instantiation of on-demand, opportunistic LHC/CMS computing centers 50

Pick a paper copy here, or download it from https://www.indigo-datacloud.eu/communication-kit

About **40 software products**
More than **200 software packages**
40 ready-to-use docker containers
Released under the **Apache 2.0 license**
Available on the **INDIGO website & on github**

INDIGO - DataCloud
Better Software for Better Science

INDIGO MidnightBlue Service Catalogue

Updates and new releases of the INDIGO services are expected to come in the forthcoming months.
The first scientific applications and use cases adopting this first INDIGO release are expected starting from September 2016.

An unmatched open modular suite of software components for Data and Cloud computing is now available for resource providers and researchers from all disciplines, all around Europe

INDIGO MidnightBlue

Four main "solution blocks":

- Data Center Solutions
- Data / Storage Solutions
- Automated Solutions
- User-Oriented Solutions

And "common solutions":

- Authentication and Authorization

December 2016

High level application portals and mobile applications

Data Center and Storage solutions

Services and solutions allowing data and compute resource centers to increase efficiency for customers


Fairshare Scheduler for OpenStack


Cloud Provider Ranker


Global Data Access


Infrastructure Manager


Storage Quality of Service and Data Lifecycle Support


Partition Director Service for Batch and Cloud resources


Extended OpenStack and OpenNebula Functionalities


OCCI Support for OpenStack and OpenNebula

User-oriented solutions

Services and solutions to improve the QoS user experience


Userspace Container Support


Data Mining and Analytics for e Science Server


Indigo Plug-ins for scientific workflow systems


Future Gateway (Programmabile Scientific Portal)

Automated solutions

A rich set of high-level automated functionalities for Data Centers and customers


PaaS Orchestrator


QoS/SLA Management Service


Core PaaS

Authentication and Authorization service


User Authentication

Index of Services

INDIGO - DataCloud

Common Solutions	
Identity and Access Management	14
Data Center Solutions	
Fairshare Scheduler for OpenStack	15
Partition Director Service for Batch and Cloud resources	16
Cloud Provider Ranker	17
Infrastructure Manager	23
OCCI support for OpenStack and OpenNebula	24
Extended OpenStack and OpenNebula Functionalities	25
Data Solutions	
Global Data Access	18
Storage Quality of Service and Data Lifecycle support	27
Automated Solutions	
PaaS Orchestrator	20
Core PaaS	26
QoS/SLA Management Service	30

User-oriented Solutions	
Userspace Container Support	21
Data Mining and Analytics for eScience Server	22
Future Gateway (Programmable Scientific Portal)	28
INDIGO Plug-ins for scientific workflow systems	29
INDIGO Mobile Toolkit	31

An excerpt of these services follows…

1. Identity and Access Management

Short Service Name

IAM

Solution Type

Common Solution

Installation Area

The AAI common glue, implemented across the entire INDIGO architecture

The IAM service provides a layer where identities, enrolment, group membership, attributes and policies to access distributed resources and services can be managed in a homogeneous and interoperable way. It supports the federated authentication mechanisms behind the INDIGO AAI.

The IAM service provides user identity and policy information to services so that consistent authorization decisions can be enforced across distributed services.

Identity and Access Management is provided through multiple methods (SAML, OpenID Connect and X.509) by leveraging on the credentials provided by the existing Identity Federations (i.e. IDEM, EDUGAIN, etc). Distributed authorization policies and Token Translation Service will guarantee selected access to the resources as well as data protection and privacy.

2. Fairshare Scheduler for OpenStack

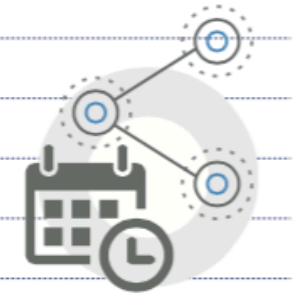
Short Service Name

Synergy

Solution Type

Data Center Solution

Installation Area



Synergy, the NDIGO Fairshare Sceduler for Openstack, is an extensible all-purpose management service for integration with OpenStack Infrastructures. It is implemented by a collection of independent pluggable tasks and executed periodically (e.g. cron jobs) or interactively (e.g. RESTful API). Synergy can be used to allocate a set of dynamic OpenStack resources to be shared among different projects. Moreover, Synergy offers a queuing mechanisms for requests until relevant resources are available. It can oversee the instantation of both virtual machines and containers managed via the nova-docker service.

3. Partition Director Service for Batch and Cloud resources

Short Service Name

Dynpart

Solution Type

Data Center Solution

Installation area

Dynpart, the Partition Directore Service for Batch and cloud resources, facilitates the management of a hybrid data center that provides both batch-system based services and cloud-based services. Physical computing resources, in fact, can act as member of batch system cluster or as compute node in a cloud environment.

Dynpart can easily manage such mutual exclusive approach of physical resources making the data center dynamic and flexible.

6. PaaS Orchestrator

Short Service Name

Orchestrator

Solution Type

Automated Solution

Installation Area

PaaS Orhcestrator iis the core component of the PaaS layer. It collects high-level deployment requests from the software layer, and coordinates the resource or service deployment over dynamic Mesos clusters or directly over IaaS platforms.

«**OneData allows us** to implement a first prototype of distributed archive for the Cherenkov Telescope Array (CTA) project. The distributed architecture of the CTA Archive will allow to lower costs with respect to a single huge data centre including easy manageability and maintenance».

INAF

Eva Sciacca,
Researcher at INAF – Astrophysics Observatory of Catania, Italy

5. Global Data Access

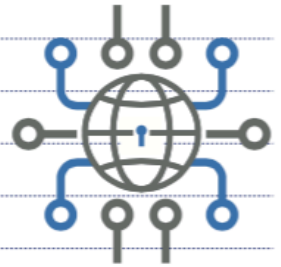
Short Service Name

OneData

Solution Type

Data Solution

Installation Area



Global Data Access is the global data management system providing easy access to distributed storage resources and supports a wide range of use case, from data management to data-intensive scientific computations

INDIGO communities

Physics & Astrophysics

«**We tried Onedata** for the Large Binocular Telescope (LBT) use case and configured this service for a simulated distributed archive. We tried to install and configure all Onedata components (Onezone, Oneprovider and Oneclient) on dedicated virtual machines, deploying docker images. The main goal was to use Onedata to store and distribute data to different sites according to defined data policies».

«Onedata provides an intuitive configuration interface and a very flexible framework to store data using global distributed storage providers».

Andrea Bignamini,
Researcher at the Astronomical Observatory of Trieste, Italy

12. Core PaaS

Short Service Names

Kubernetes, Monitoring, Accounting

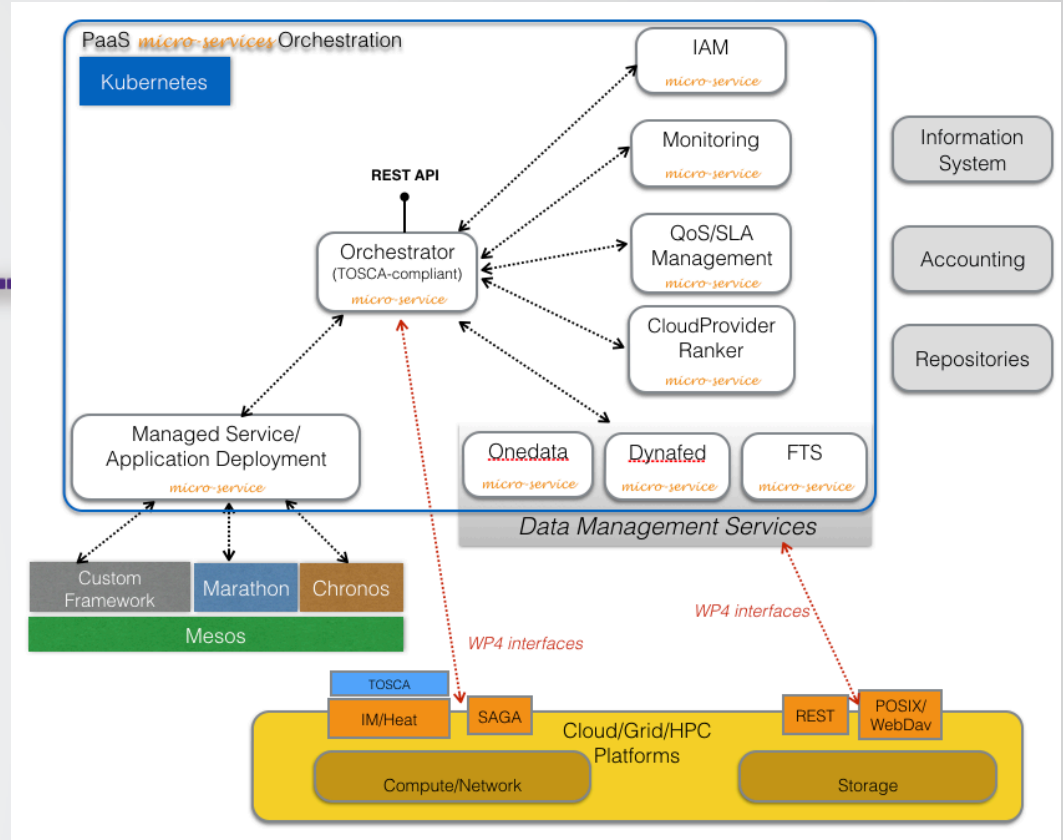
Solution Type

Automated Solution

Installation Area

Core PaaS provides the basic functionalities and tools to steer the performance of all the PaaS services available in the infrastructures. In particular:

- availability and scalability of the core services
- monitoring of the Computational and Storage resources and of the PaaS µServices
- accounting of the resource usage in terms of computing and storage



The INDIGO PaaS core is built upon a set of services (exposing REST interfaces) that are:

- Deployed
- Scaled
- Managed
- Upgraded
- Monitored
- Self-healed

through **Kubernetes** (http://kubernetes.io), an open source system for managing containerized applications across multiple hosts in a cluster.

13. Storage Quality of Service and Data Lifecycle support

Short Service Name

CDMI

Solution Type

Data Solution

Installation Area

This solution implements the INDIGO-DataCloud CDMI Server, a set of functionalities aimed at Improving QoS capabilities of storage resources for better support of high-level storage requirements, such as flexible allocation of disk or tape storage space and support for data life cycle.

- This is an enhancement also with respect to what is currently available in public clouds, such as Amazon Glacier and Google Cloud Storage.

CDMI provides the official reference implementation of the SNIA Cloud Data Management Interface (CDMI), an ISO standard, and also a Spring Boot application port of the SNIA CDMI-Server. The CDMI server has been extended to support Quality-of-Service (QoS) and Data Life-cycle (DLC) operations for multiple storage back-ends like dCache, Ceph, GPFS, Gemss+TSM, StoRM and HPSS.

14. Future Gateway (Programmable Scientific Portal)

Short Service Names

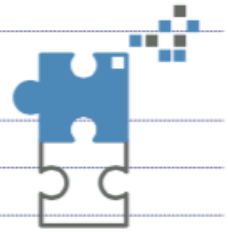
fgAPIServer, fgAPIServerDaemon

PortalSetup, fgTools

Solution Type

User-oriented Solution

Installation Area



Future Gateway is a programmable interface of a RESTful API Server, compliant with CSGF APIs specifications, able to provide an easy acces to the PaaS layer by leveraging on recent Web technologies

The FutureGateway consists of a set of software components able to build, or assist existing web portals or other community oriented interfaces to become Science Gateways. FutureGateway allows the access to distributed computing resources such as grid, cloud and HPC. FutureGateway comes from a four years experiences gained with a similar component. In particular the following key features have been identified:

- Provide a more flexible way accessing the distributed computing services.
- Leave to the FutureGateway adopters the choice of the backward portal technology.
- Provide the most simple way to develop ScienceGateway applications.

The FutureGateway is composed by a set of tools:

- FutureGateway API Server
- FutureGateway API Server Daemon
- FutureGateway jSAGA Adaptors
- FutureGateway Portal Setup
- fgTools

15. INDIGO Plug-ins for scientific workflow systems

Short Service Names

indigoclient, indigoKepler

Solution Type

User-oriented Solution

Installation Area

Workflow management systems provide an infrastructure for the set-up, performance and monitoring of a defined sequence of tasks, arranged as a workflow by using resources and services made available by providers and other communities.

The INDIGO Plug-ins for scientific workflow systems are aimed at interacting with the already available INDIGO solutions, facilitating the deployment of complex scientific workflow.

This plugin could be used in order to allow to exploit the INDIGO APIs and submit requests for execution of applications directly from the workflow manager to the INDIGO APIs.

- Indigo-DC client API is Java based library that can perform calls to Future Gateway API. This library provide basic means for accessing resources provided by Indigo-DC project. It can be used by Java based applications as Workflow manager like Kepler.

Technical Support

Most complex software contains bugs, and we are not an exception. One of the features of free and open source software is the ability to report bugs, helping to fix or improve the software you use. The INDIGO-DataCloud project uses the GGUS (Global Grid User Support) tool as its user support system. It provides sophisticated search functionality, report generation, interfaces to bug tracking systems used by different middleware components, and automatic ticket reminder including escalation indication. Please use the INDIGO-DataCloud Catch-All GGUS Support Unit or directly contact us through the indigo-su@lists.indigo-datacloud.eu mailing-list.

Share the INDIGO Experience

Developers, researchers and IT enthusiasts: feel free to write to info@indigo-datacloud.eu to ask for more information on how to deploy your PaaS-based solution for your work. For automatic notifications, you can register to the INDIGO-DataCloud RSS release feed or subscribe to the INDIGO-DataCloud Announce Mailing list. You can also socialize with us via Twitter, Facebook and LinkedIn. Finally, you can also subscribe to **INDIGO Newsletters** and receive communications about the project, such as new releases, community events and other events where to meet the INDIGO team, tutorials, workshops, webinars, guides, and more.

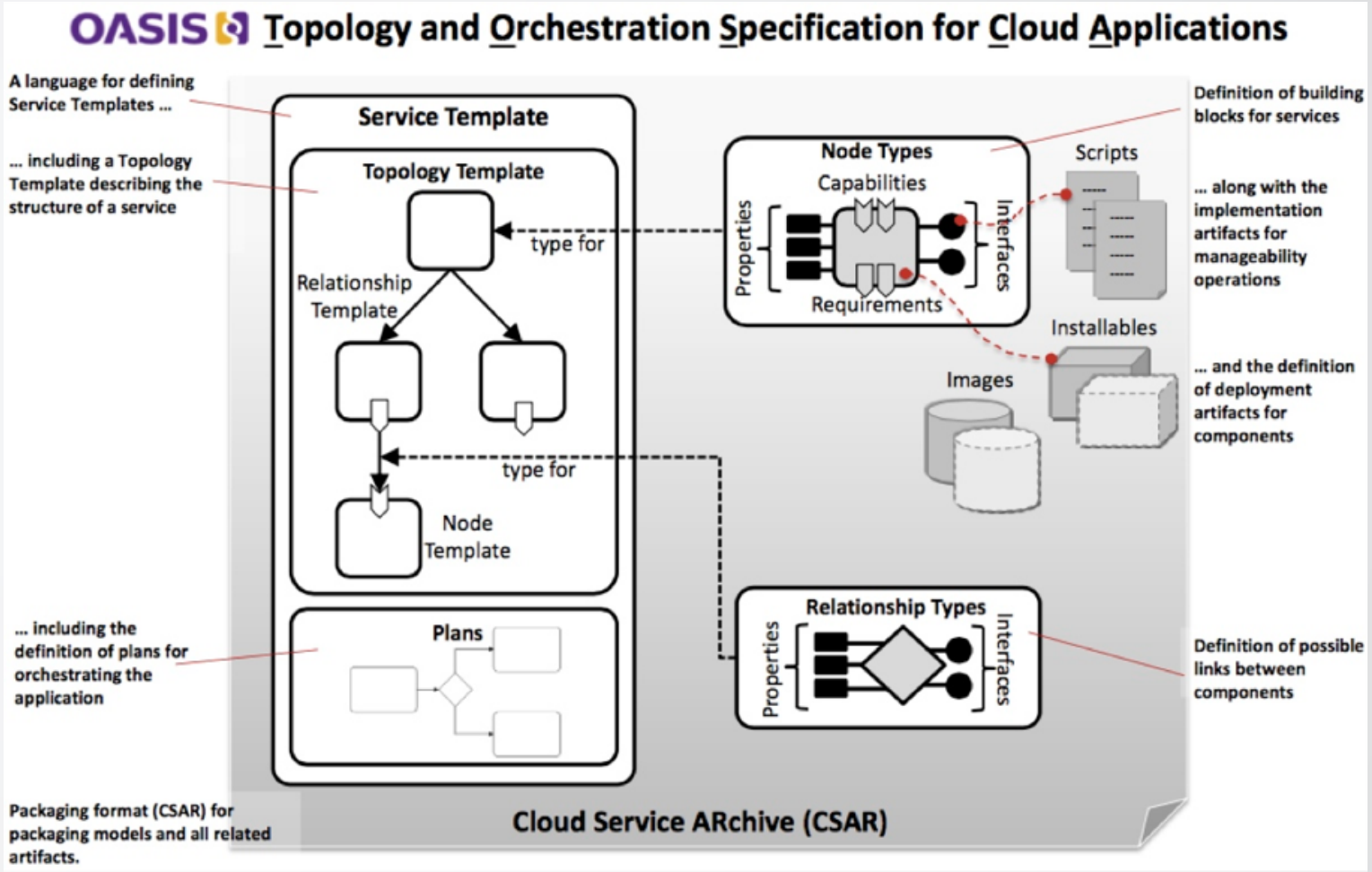
- INDIGO-Datacloud Website: https://www.indigo-datacloud.eu
- Technical Support: https://www.indigo-datacloud.eu/indigo-support-and-technical-services
- Twitter: https://twitter.com/indigodatacloud
- Facebook: https://www.facebook.com/indigodatacloud
- Linkedin: https://it.linkedin.com/in/indigodatacloud

INDIGO - DataCloud

Technology Detour: TOSCA

- **T**opology and **O**rchestration **S**pecification for **C**loud **A**pplications
- Standardizes the language to describe:
  - *The structure* of an IT service (its **topology** model)
  - *How to orchestrate operational behavior* (plans such as build, deploy, patch, shutdown, etc.)
- It is a declarative model that spans applications, virtual and physical infrastructures.
- INDIGO supports TOSCA declarations at the IaaS, PaaS and SaaS levels to automatize the definitions & instantiation of services.

TOSCA in a nutshell

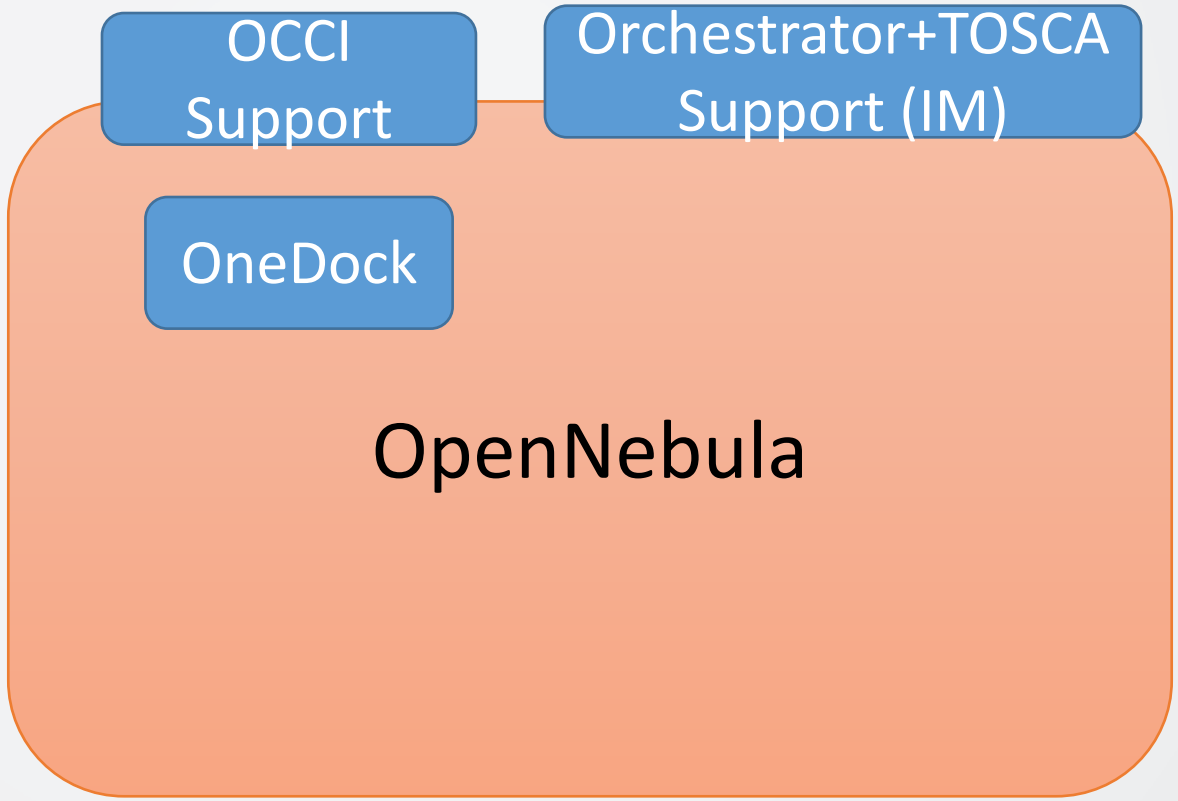


TOSCA topology template example

```
tosca_definitions_version: tosca_simple_yaml_1_0

description: Template for deploying a single server with predefined properties.

topology_template:
  inputs:
    cpus:
      type: integer
      description: Number of CPUs for the server.
      constraints:
        - valid_values: [ 1, 2, 4, 8 ]

  node_templates:
    my_server:
      type: tosca.nodes.Compute
      capabilities:
        # Host container properties
        host:
          properties:
            # Compute properties
            num_cpus: { get_input: cpus }
            mem_size: 2048  MB
            disk_size: 10 GB

  outputs:
    server_ip:
      description: The private IP address of the provisioned server.
      value: { get_attribute: [ my_server, private_address ] }
```

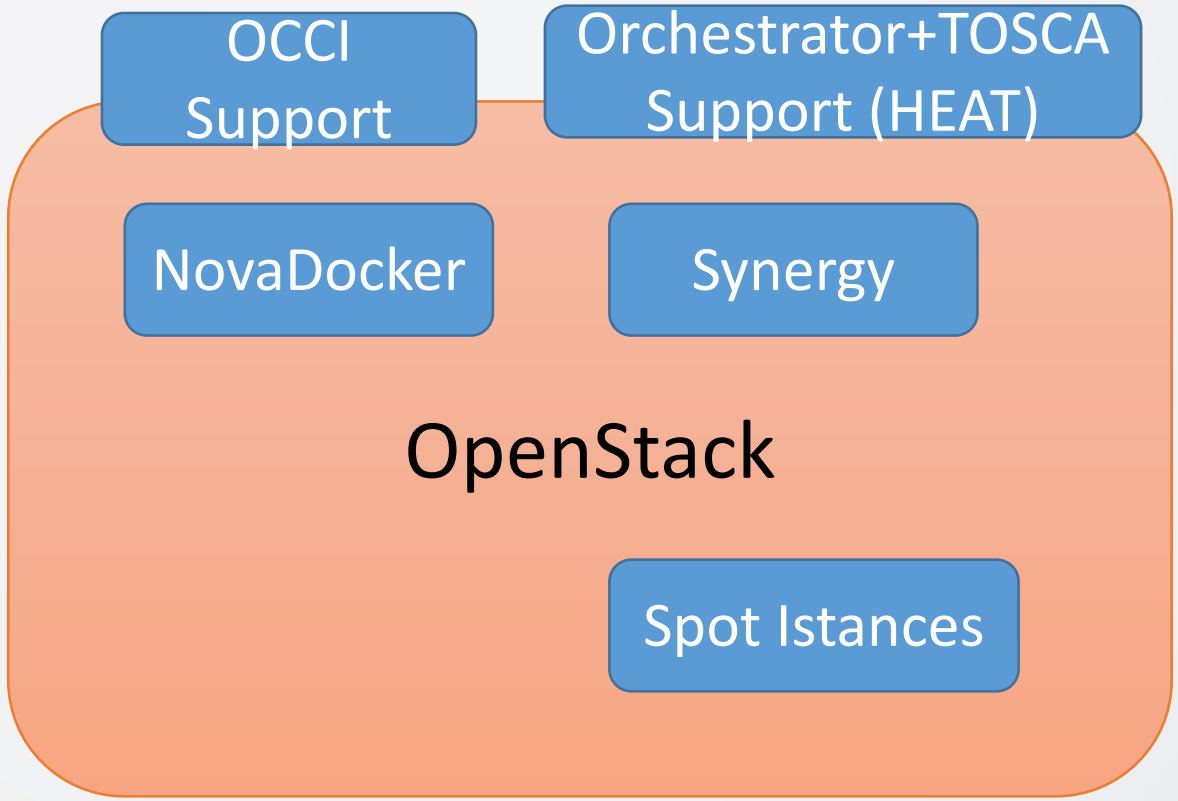
INDIGO - DataCloud

Enhanced Resource Virtualization -> *Computing (OpenNebula)*



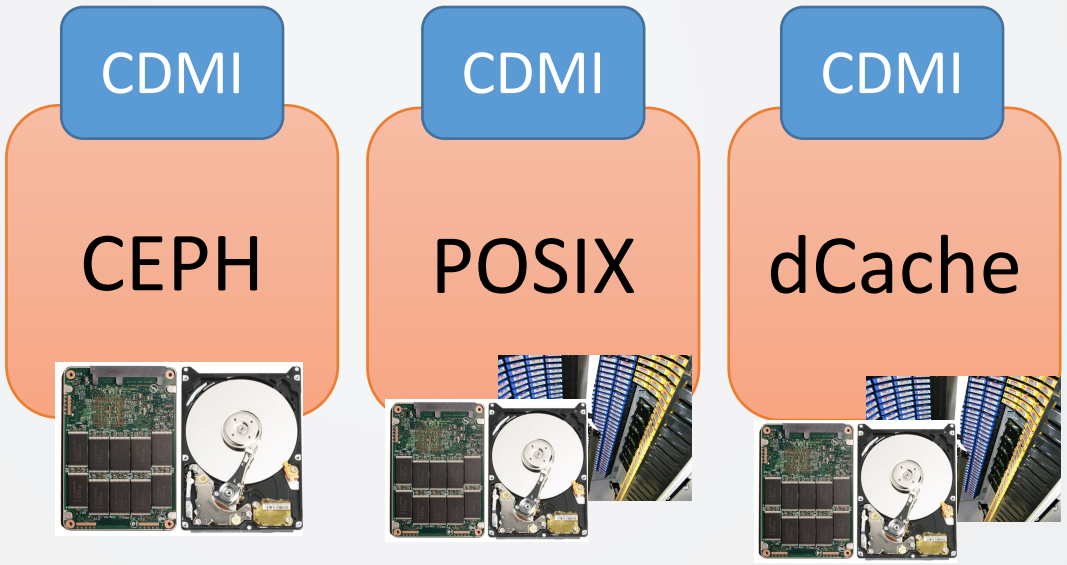
1. **IM**: Provides
   a) Advanced **IaaS Orchestrator** capabilities
   b) TOSCA Support
2. **OCCI**:
   a) Enhanced Network capabilities
   b) Docker support
3. **OneDock**:
   a) Support for **native Docker** (on bare-metal)

Enhanced Resource Virtualization -> *Computing (OpenStack)*



1. **TOSCA on HEAT**
2. **OCCI**:
   a) **Enhanced Network** capabilities
   b) **Docker** support
3. **Nova Docker**:
   a) Support for **native Docker** (on bare-metal)
4. **Synergy**:
   a) **Fair-share** on cloud resource usage
5. **Spot-instances**

Enhanced Resource Virtualization -> *Storage (QoS)*

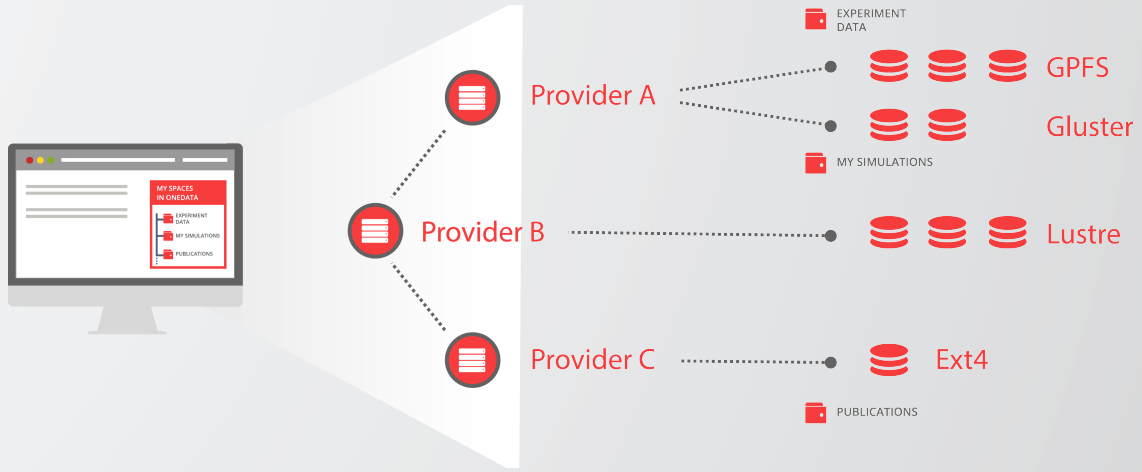


1. **CDMI service** provides the capability to manage the QoS of storage
2. Independent from the technology used
3. CDMI is not used to access files at the site level
   a) The files still could be accessed/stored using the original protocols
      a) WebDAV
      b) Posix
      c) S3
      d) GridFTP

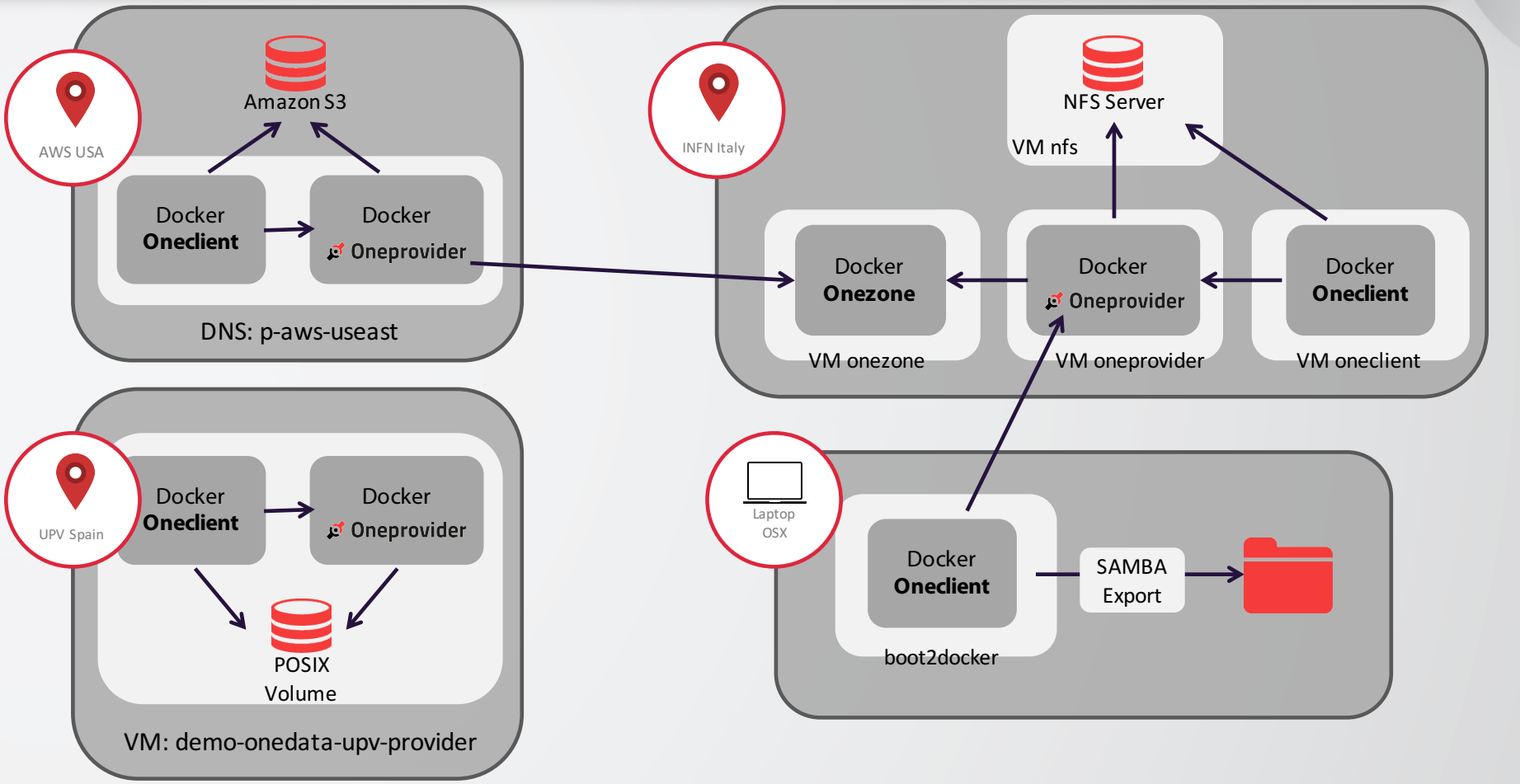
Data Services: Onedata

- **Unified data access** over heterogenous infrastructures
- **High-performance** data **access** and **migration**
- **POSIX interface** for accessing large data sets without pre-staging them
- Flexible security framework based on **access tokens** and **ACL's**
- **Metadata editing** and querying using **key-value** pairs, **JSON** and **RDF**

- Support for **POSIX**, **Ceph**, OpenStack **SWIFT** and Amazon **S3**
- **Graphical User Interface** for easy data management
- Comprehensive **REST API** for integrating with other services



Data Federation through INDIGO Onedata



Sample use cases:

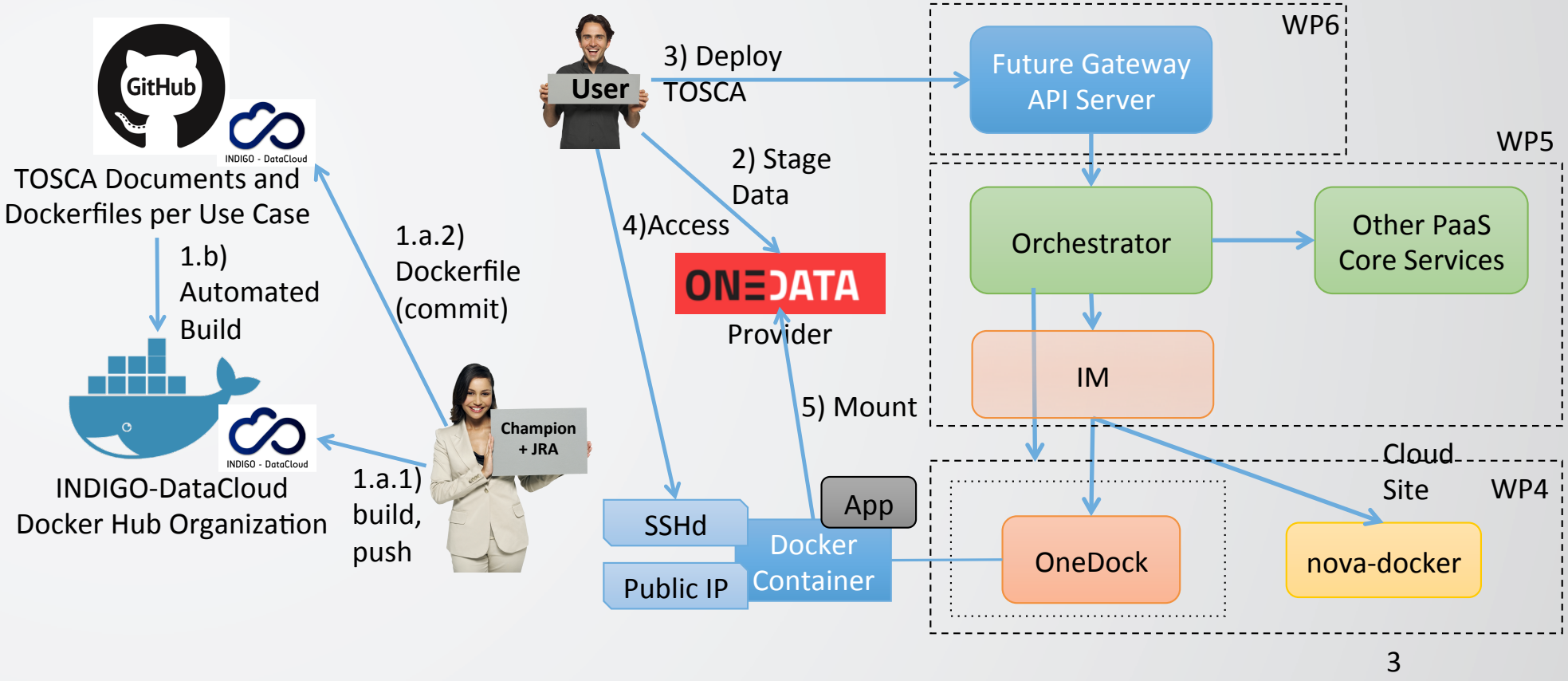
INDIGO - DataCloud

1. Interactive usage of a Docker container with *ssh*
2. A web portal that uses a batch system to run applications
3. Virtual infrastructures for Medical Imaging Biobanks
4. An application to CMS
5. Running Docker containers without Docker

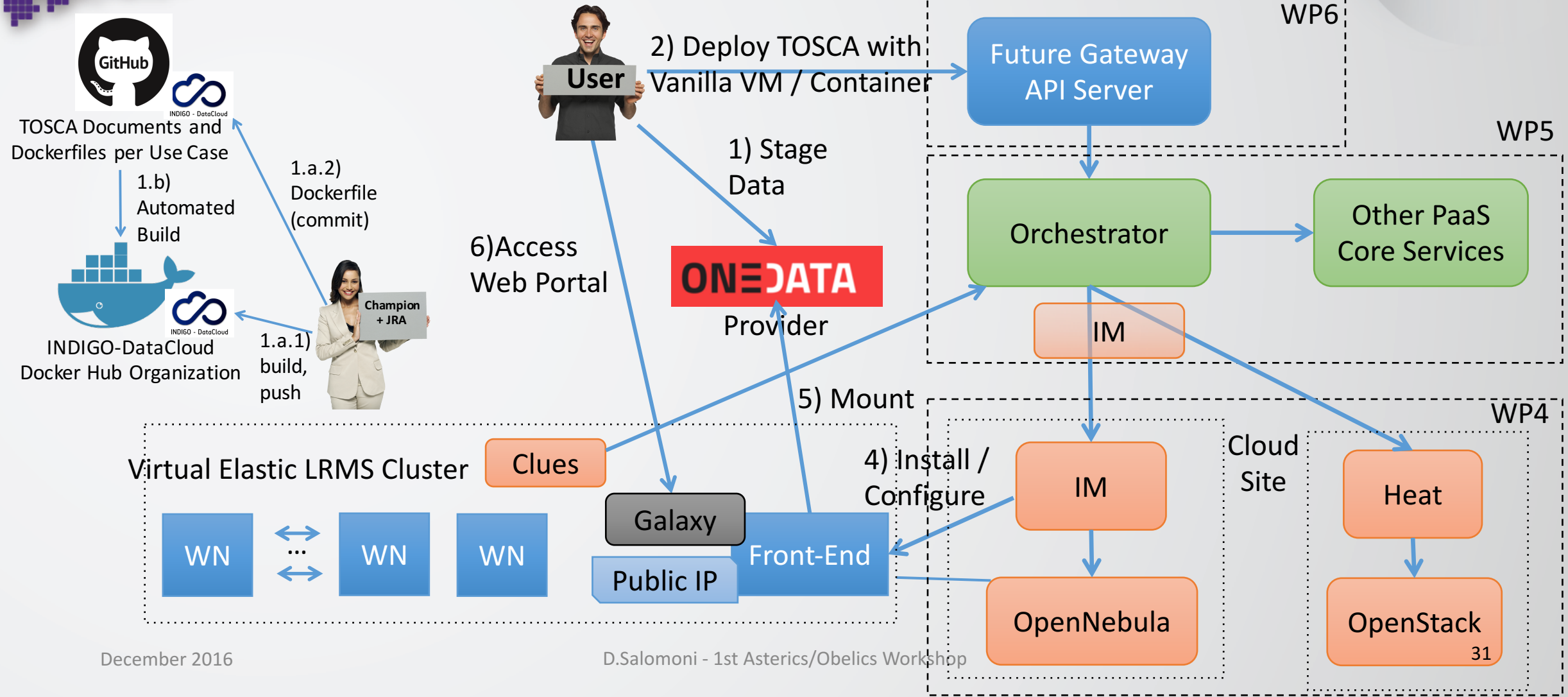
UC #1: Interactive usage of a Docker container with *ssh*

- A Docker container is instantiated automatically after a simple request on the web portal from an end-user.
  - This will exploit a **TOSCA Template** through the INDIGO orchestrator.
- The container has a public IP address and the user (or the portal) can directly get access to it.
- Users can mount a local **or remote** Posix filesystem through INDIGO **Onedata**.
- The application in the Docker container is able to simply read the files provided via web browser by the end user and to write Posix files that are available to users via web browsers.
- The same Docker container could be used to execute a large list of applications in a batch-like behaviour.

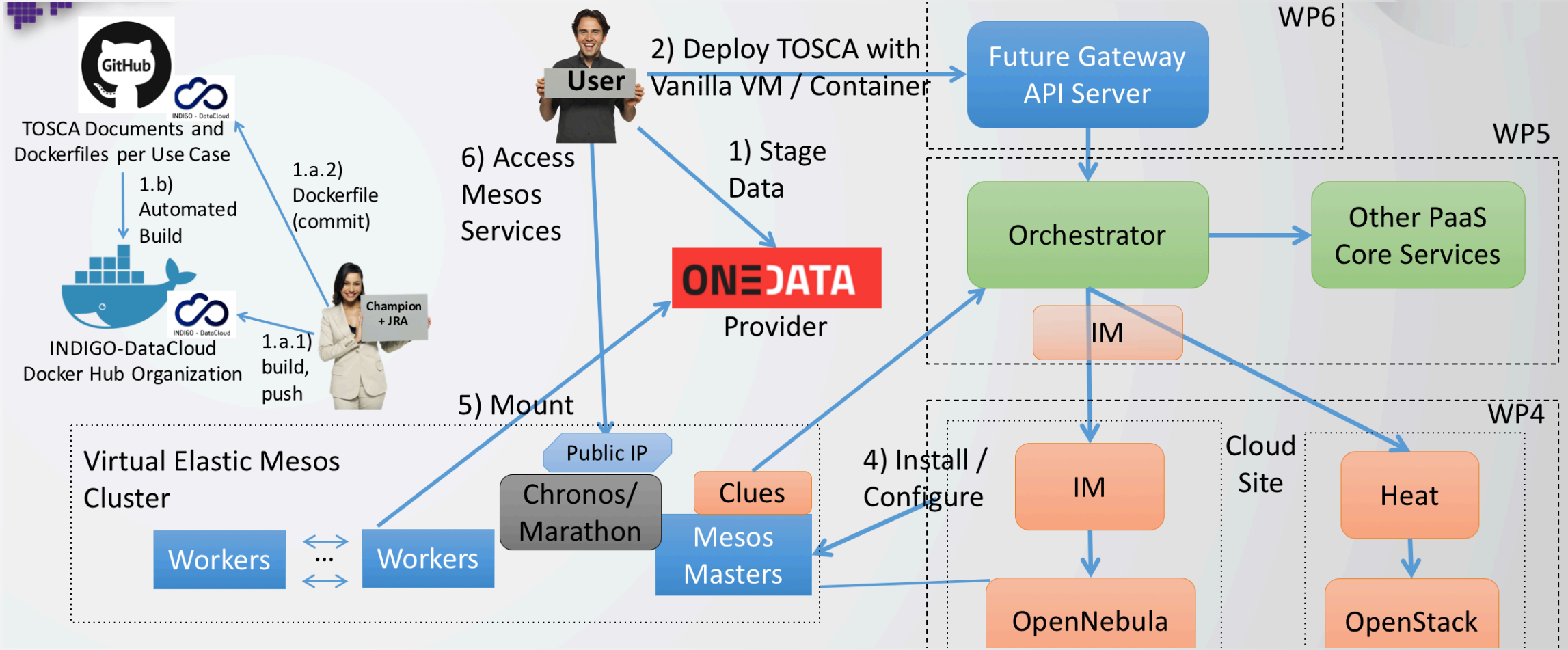
UC #1: Interactive usage of a Docker container with *ssh*



UC #2: A web portal that uses a batch system to run applications



UC#2, a variant: a Mesos-based scalable cluster

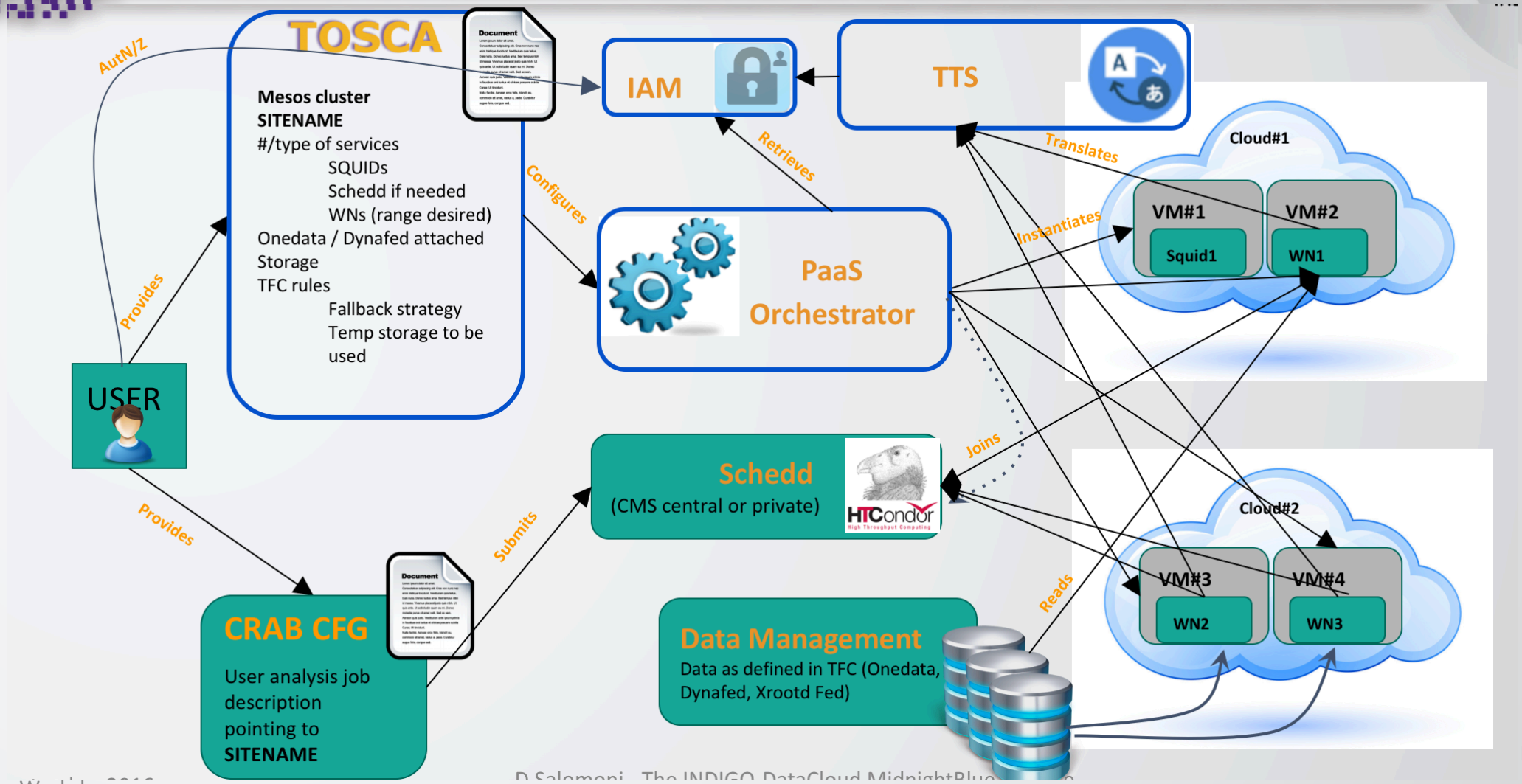


UC#3: An application to LHC/CMS

INDIGO - DataCloud

- The **goal** is to develop a solution for generating automatically an on-demand, container-based cluster for CMS in order to allow:
  - The effective use of **opportunistic resources**, such as general purposes campus facilities.
  - The **dynamic extension** of an already existing dedicated facility.
- By simplifying and automating the process of creating, managing and accessing a pool of computing resources the project aims to **improve**:
  - **Sites management:**
    - A simple solution for dynamic/elastic T2 extensions on “opportunistic”/stable resources
    - A friendly procedure to dynamically instantiate a spot “Tier3-like resource center”
  - **Users experience:**
    - Generation of an ephemeral on-demand T3 seen by the Experiment computing infrastructure as a personal WLCG-type facility, in order to serve a group of collaborators. The system must allow the use of standard/regular CMS Tools such as CRAB.
  - **Experiment-Collaboration resources:**
    - A comprehensive approach to opportunistic computing. A solution to access and orchestrate e.g. multiple campus centers, harvesting all the free CPU cycles without major deployment efforts.

UC#3: An application to LHC/CMS



UC#3: An application to CMS, four pillars:

- **Cluster Management:**
  - Mesos clusters as a solution in order to execute docker for all the services required by a regular CMS site (Worker Nodes, HTCondor Schedd and squids).
  - Marathon guarantees us the dynamic scaling up and down of resources, a key point.
- **AuthN/Z & Credential Management:**
  - The INDIGO Identity Access Management (IAM) service is responsible for AuthN/Z to the cluster generation.
  - The Token Translation Service (TTS) enables the conversion of IAM tokens into an X.509 certificate
    - NOTE: This allows Mesos slaves (running HTCondor_startd daemon) to join the CMS central queue (HTCondor_schedd) as a regular Grid WN
- **Data Management:**
  - Dynafed is the approach currently followed by the project. A further possibility we will investigate is Oneclient (from Onedata) as a tool allowing to mount remote Posix filesystems.
- **Automation:**
  - TOSCA templates, meant to be managed by INDIGO PaaS Orchestrator, allow the automation of the overall setup.
    - The aim is to produce a single YAML file describing the setup of all required services and deps.

UC#4: Running Docker containers without Docker ☺

- Observation: adoption of Docker is very slow in HPC centers
- Thus the typical situation is that Docker is not installed and one cannot run containers without some support from system software and/or friendly sysadmins.
- In general, Docker adoption will likely be slow in any computing farm or interactive linux system shared by many users.
  - It will take time for sysadmins to overcome the concerns of their security teams.
  - It is yet another service to maintain…
  - …. you name it.

UC#4: INDIGO udocker

- **A tool to execute content of docker containers in user space** when docker is not available
  - enables download of docker containers from dockerhub
  - enables execution of docker containers by non-privileged users
- **It can be used to execute the content of docker containers in Linux batch systems and interactive clusters managed by others**
- **A wrapper around other tools to mimic docker capabilities**
  - current version uses **proot** to provide a chroot like environment without privileges (it runs on CentOS 6, CentOS 7, Fedora, Ubuntu)
- **More info and downloads at:**
  - https://www.gitbook.com/book/indigo-dc/udocker/details
  - https://indigo-dc.gitbooks.io/udocker/content/doc/user_manual.html

UC#4: INDIGO udocker

- Examples:

```
# download, but could also import or load a container exported/save by docker
$ udocker.py  pull   ubuntu:latest
$ udocker.py  create  --name=myubuntu   ubuntu:latest

# make the host homedir visible inside the container and execute something
$ udocker.py  run   -v $HOME    myubuntu   /bin/bash   <<EOF
cat   /etc/lsb-release
ls  -l  $HOME
EOF
```

udocker is NOT an alternative to docker: we need the container image built by docker.

It is a tool to handle and run containers with regular user privileges and/or when docker is not available for some reason: **it is very convenient to access clusters and Grid resources**

UC#4: INDIGO udocker

- Everything is stored in the user home dir or some other location
- Container layers are download to the user home
- Directory trees can be created/extracted from these container layers
- proot uses the debugger ptrace mechanism to change pathnames and execute transparently inside a directory tree
- No impact on read/write or execution, only impact on system calls using pathnames (ex. open, chdir, etc)

- Does not require installation of software in the host system:
  - udocker is a python script
  - proot is statically compiled

Conclusions

- The **first public INDIGO release** was issued at the beginning of August 2016.
- Its services are already available in several testbeds.
- Concrete use cases are currently being implemented by many scientific communities.
- A lot of important developments are being carried on in coordination with upstream developers, so that code maintenance is not only upon us.
- **Now looking** for early adopters / people willing to test and run INDIGO components with their applications or requirements. **If interested, please contact us**.
- We look forward to providing these components in a future **European Open Science Cloud through INFRADEV-4-2016 and EINFRA-12-2017 projects**.
  - And extending them through **EINFRA-21-2017** projects.

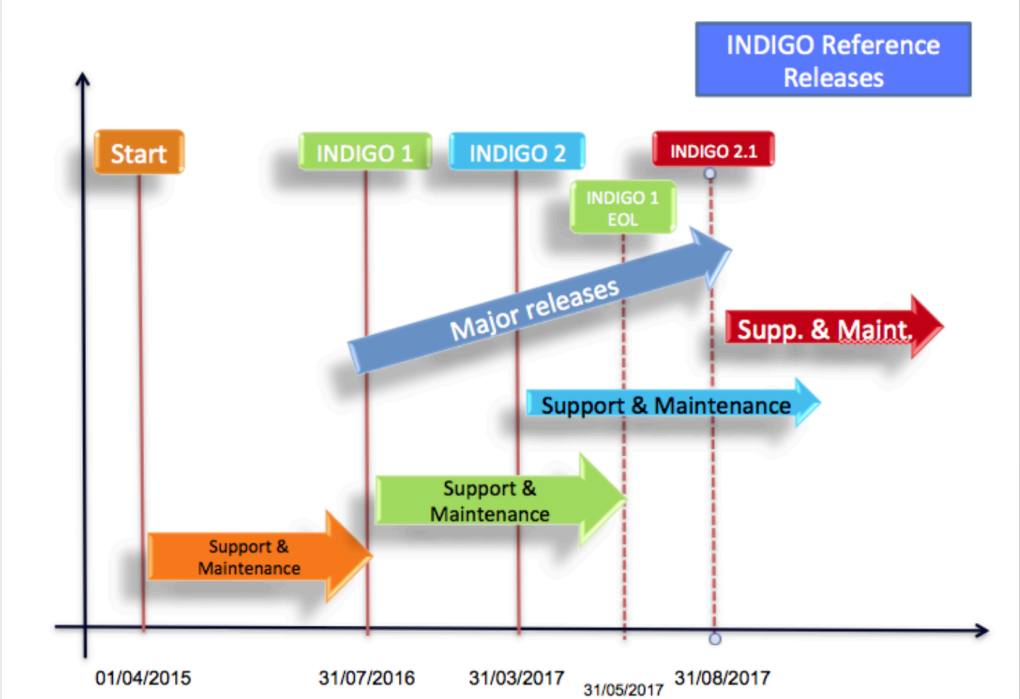


Figure 8: INDIGO-DataCloud Release Timeline

Thank you

INDIGO - DataCloud

https://www.indigo-datacloud.eu

Better Software for Better Science.