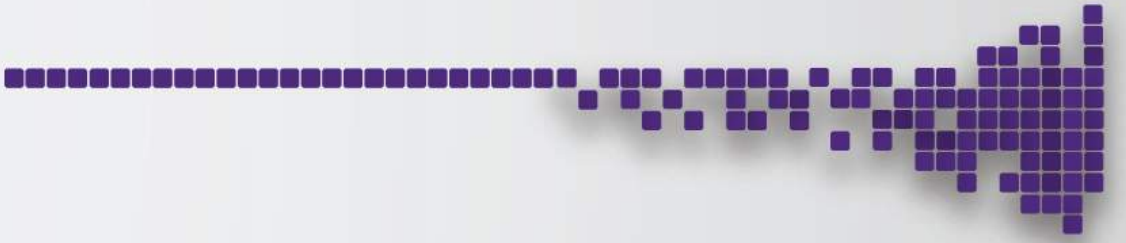
RIA-653549

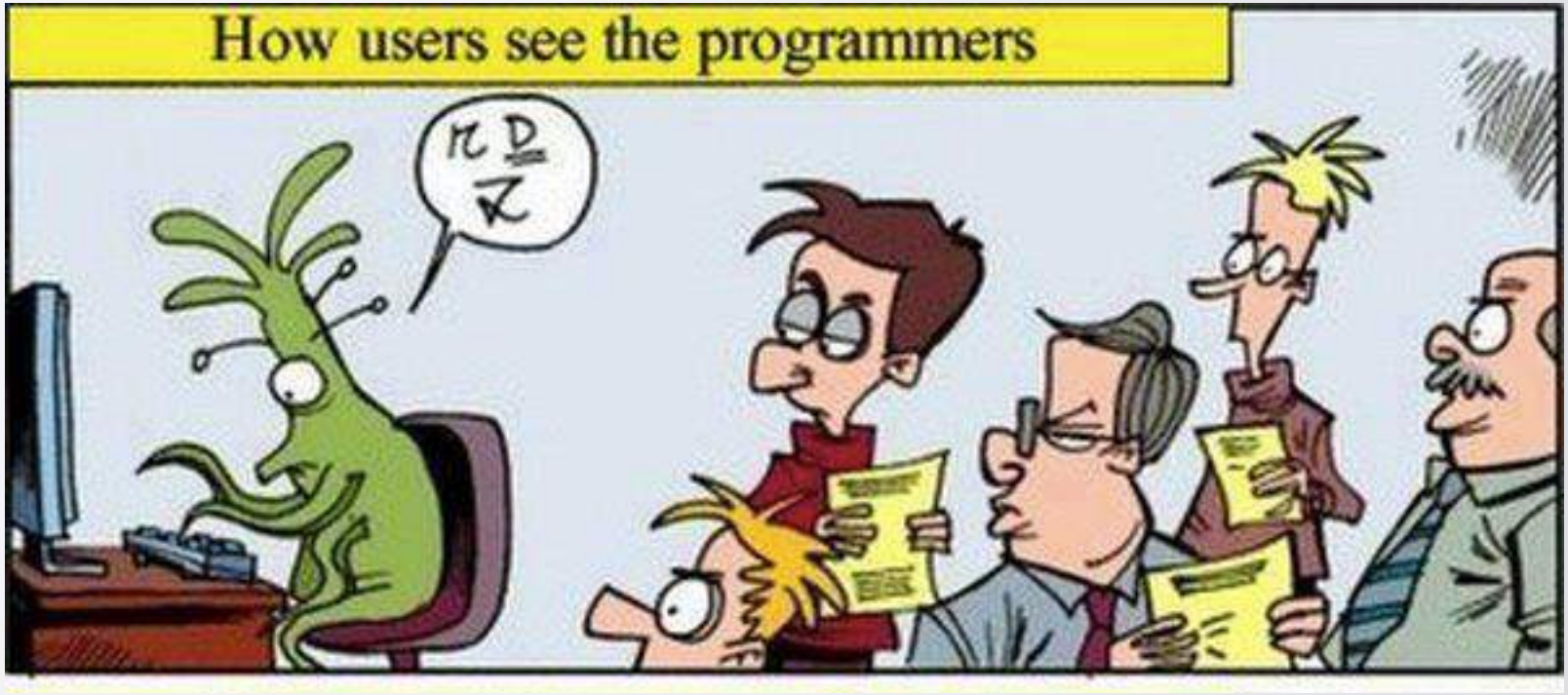
INDIGO - DataCloud a Platform enabling Cloud Federation Solutions



Doina Cristina Duma, INFN-CNAF
INDIGO-DataCloud Release Manager
cristina.aiftimiei@cnaf.infn.it

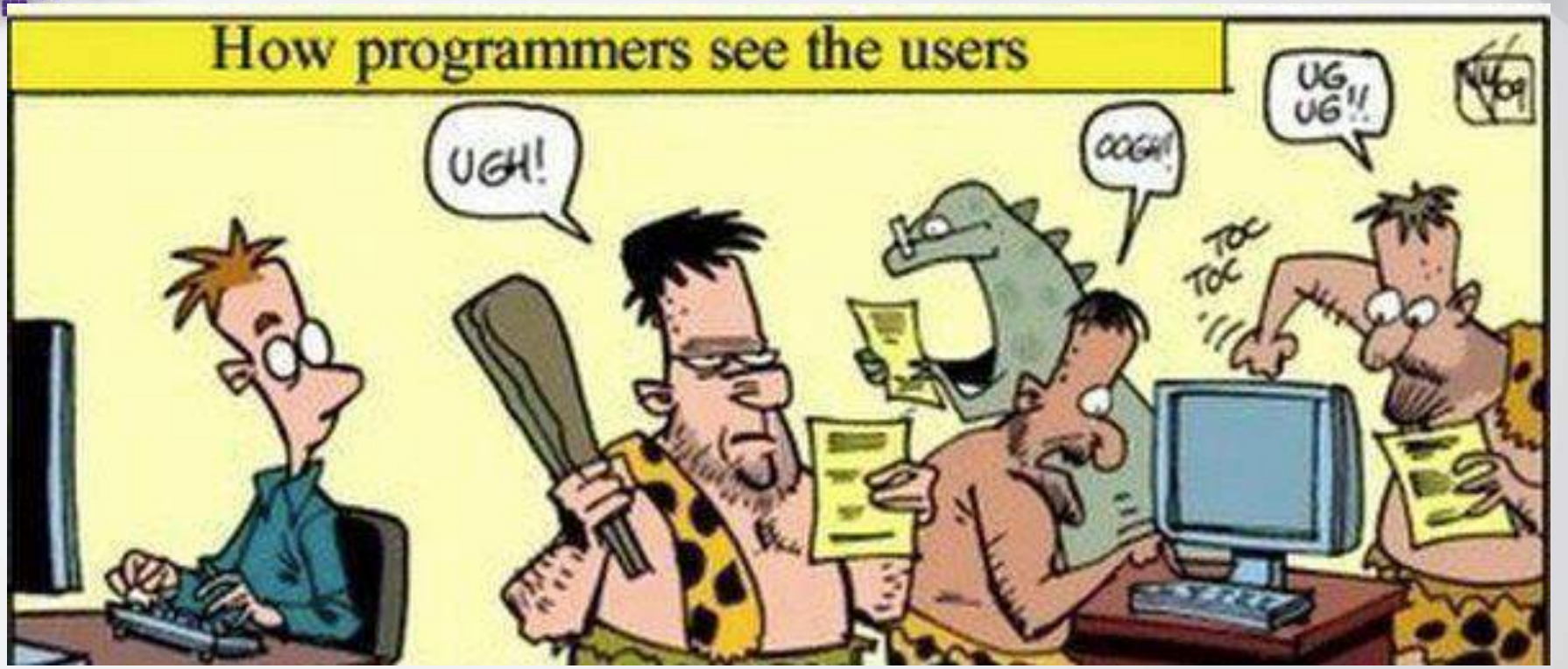
INDIGO-DataCloud is co-founded by the Horizon 2020Framework Programme

Something was ~~is~~ still missing in the Cloud world…



Source: http://goo.gl/wT8XEq

Something ~~is~~ was still missing in the Cloud world…



Source: http://goo.gl/wT8XEq

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 realized 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.

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.

The INDIGO-DataCloud overall objective

To **develop software components and solutions** to facilitate (or simply make possible) the **exploitation of distributed cloud and storage resources** through **public or private infrastructures**.

“Better Software for Better Science.”

INDIGO-DataCloud's Vision

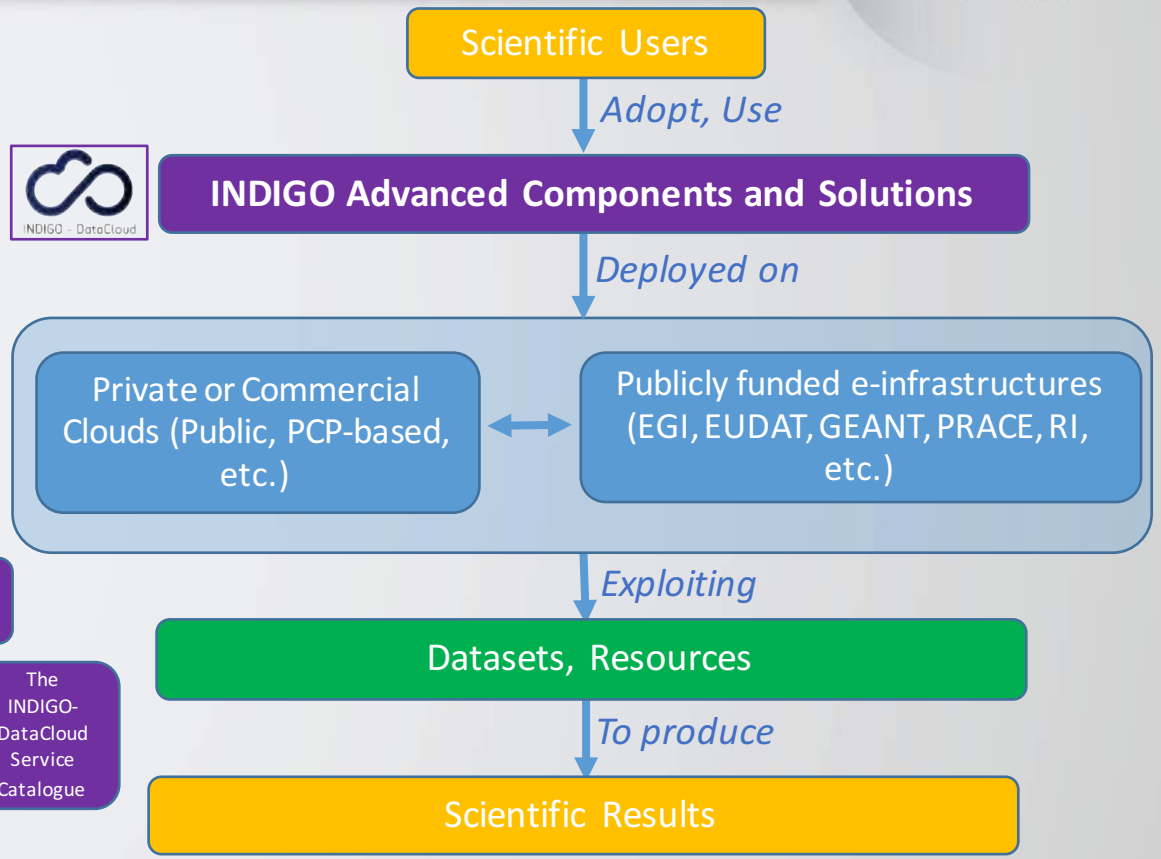
- INDIGO:
  1. **Develops open, interoperable solutions for scientific data.**
  2. **Supports open science** organizing the **European data space**.
  3. **Enables collaborations** across diverse scientific communities worldwide.
- INDIGO offers its
  - architecture,
  - analysis,
  - expertise
  -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

D1.8, General Architecture

D2.1 and D2.4, community requirements

INDIGO's 34 deliverables (so far)

The INDIGO-DataCloud Service Catalogue

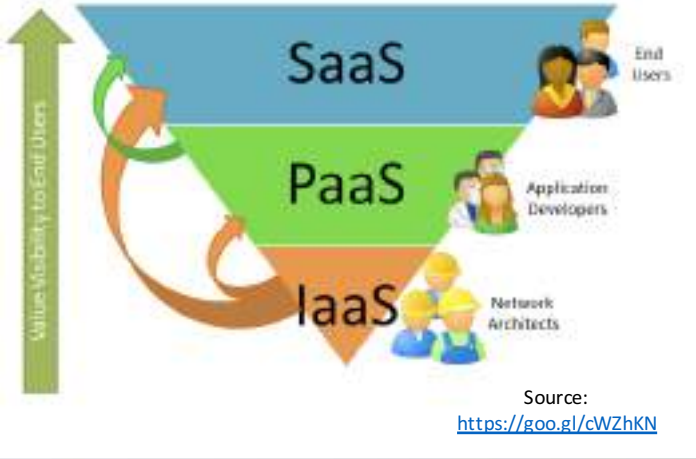


What did INDIGO originally want to address?

- Open **interoperation** / federation across (proprietary) Cloud infrastructures at the
  - IaaS,
  - PaaS,
  - and SaaS levels
- Managing **multitenancy**
  - At large scale…
  - … and in heterogeneous environments
- Handle dynamic and seamless **elasticity**
  - For both private and public clouds…
  - … for complex or infrequent requirements…
  - … through expressive and simple to use methods
- **Data management** in a Cloud environment
  - Tackling QoS, data replication, caching, transparent remote access

Addressing all of this should lead to:

- **Interoperable PaaS/SaaS services addressing both public and private Cloud infrastructures.**
- **Porting of legacy applications to the Cloud.**
- **Increased focus on user-oriented, high-value solutions.**



Source: https://goo.gl/cWZhKN

What INDIGO actually did

- INDIGO, driven by scientific communities, has been developing a **comprehensive open source Cloud architecture**, which provides **many new functionalities previously unavailable in open source and in some cases also in proprietary Cloud offerings.**
- These functionalities abstract from underlying IaaS technologies through the consistent use of both de jure and de facto standards. This allows **interoperability with hybrid (public/private) infrastructures.**
- After beta testing and demos shown as early as November 2015 (at the EGI Community Forum), **we released our first major software release (MidnightBlue) in August 2016, 9 software updates in the following months, and our second and final major release (ElectricIndigo) in April 2017.**

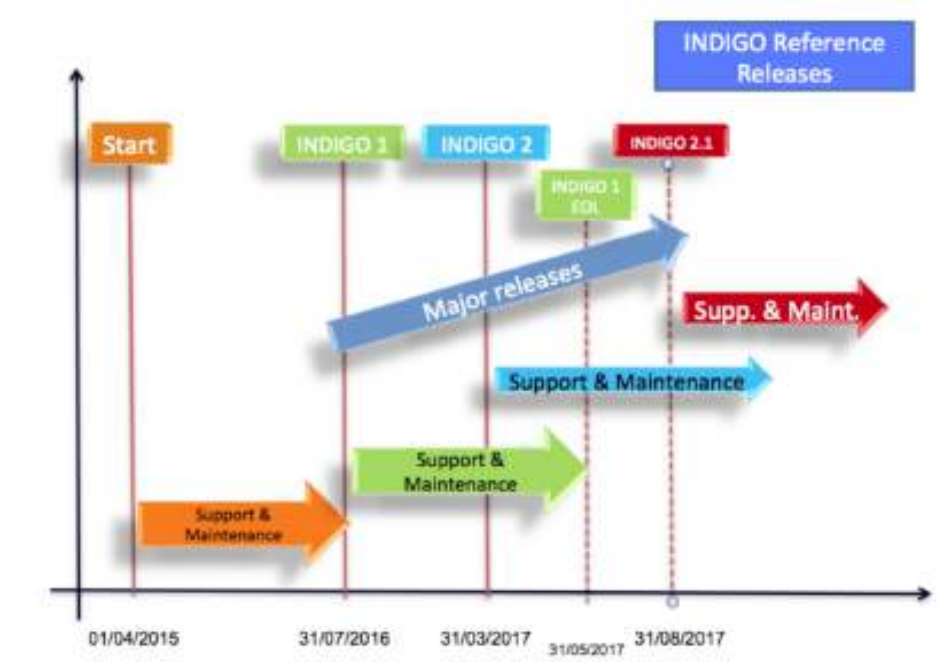


Figure 8: INDIGO-DataCloud Release Timeline

Users first: from here…

INDIGO - DataCloud

Use-Cases from
LifeWatch
EuroBioImaging
INSTRUCT
LBT
CTA
WeNMR
ENES
eCulture
ELIXIR
EMSO
Dariah
WLCG

100 distinct requirements

Users first: from here…

INDIGO - DataCloud

Use-Cases from
LifeWatch
EuroBioImaging
INSTRUCT
LBT
CTA
WeNMR
ENES
eCulture
ELIXIR
EMSO
Dariah
WLCG

100 distinct requirements

Converted to concrete activities in the Project DoW

Computational

Storage

Infrastructure

Users first: from here…

INDIGO - DataCloud

Use-Cases from
LifeWatch
EuroBioImaging
INSTRUCT
LBT
CTA
WeNMR
ENES
eCulture
ELIXIR
EMSO
Dariah
WLCG

100 distinct requirements

Converted to concrete activities in the Project DoW

Computational

- Software as a Service
- Execution of Workflows
- Cloud Bursting
- X-Site Execution *
- Improved Scheduling
- Access to GP-GPU's

Storage

- Distributed Storage, accessible via POSIX
- Persistent Data Storage *

Infrastructure

- Global Level AAI *
- Software Defined Networks

… to here … Community Case Studies

Report on how several scientific communities are implementing their own requirements into concrete applications using INDIGO-DataCloud components.

- Monitoring and Modelling Algae Bloom in a Water Reservoir
- TRUFA (Transcriptomes UserFriendly Analysis)
- Medical Imaging Biobanks
- Molecular Dynamics Simulations
- Astronomical Data Archives
- Archive System for the Cherenkov Telescope Array (CTA)
- HADDOCK Portal
- DisVis
- PowerFit
- Climate models inter comparison data analysis
- eCulture Science Gateway
- EGI FedCloud Community Requirements
- ELIXIR-ITA: Galaxy as a Cloud Service
- MOIST – Multidisciplinary Oceanic Information System
- Data Repository platform for DARIAH

https://www.indigo-datacloud.eu/documents-deliverables

... t

Repor
scient
are im
own re
concr
us
DataCl

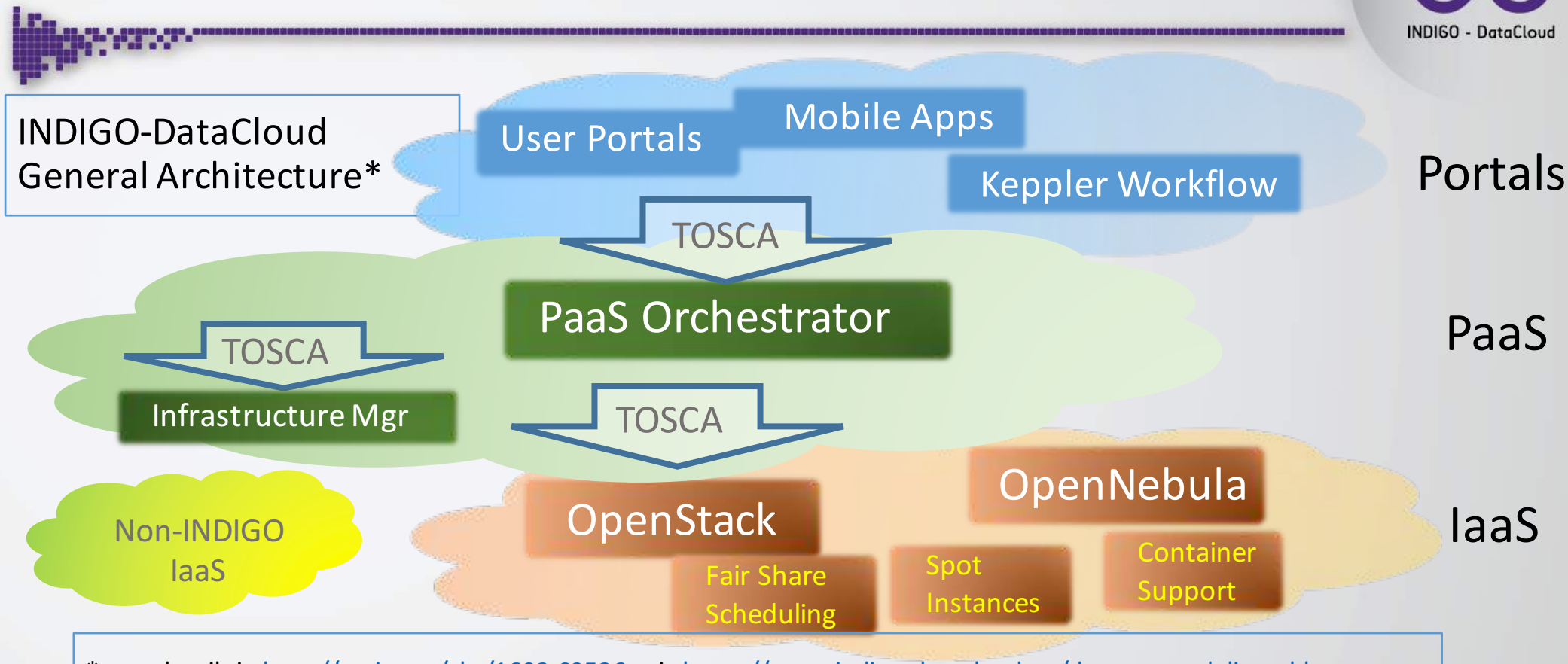
Using “Champion” approach :

Communities have to provide a scientist, becoming an expert in computing and INDIGO terminology.

- Data Repository platform for DARIAH

https://www.indigo-datacloud.eu/documents-deliverables

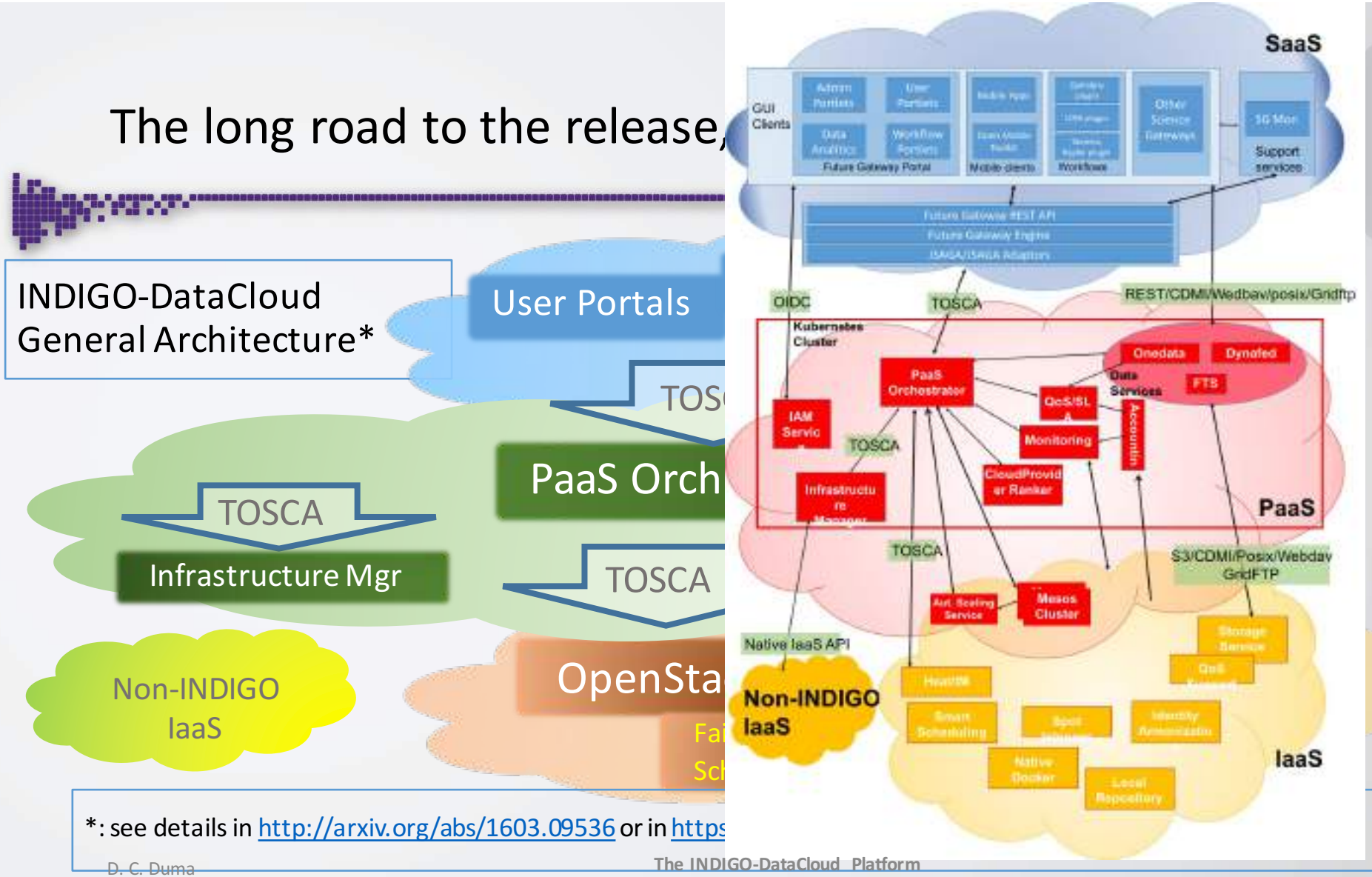
The long road to the release, from the architecture…



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

The long road to the release,

INDIGO-DataCloud General Architecture*



INDIGO - DataCloud

Portals

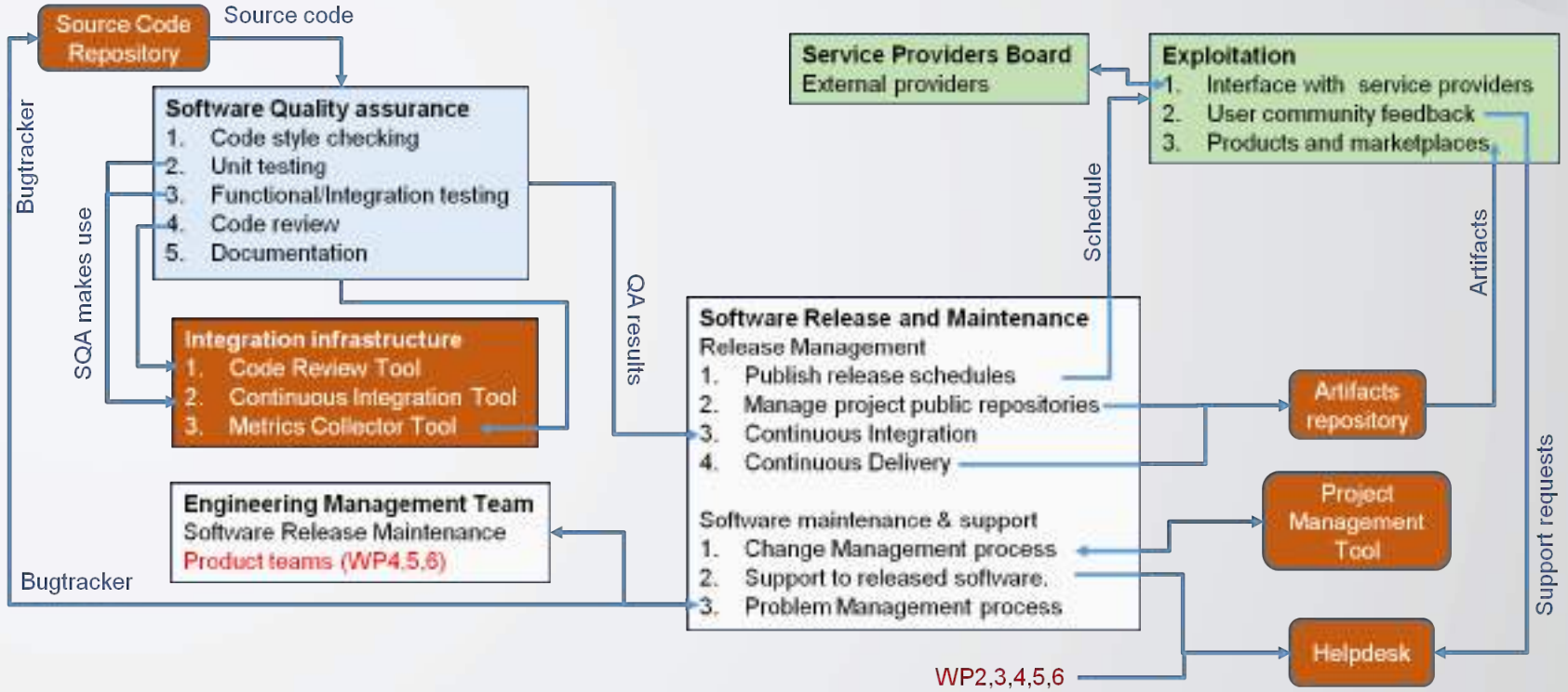
PaaS

IaaS

*: see details in http://arxiv.org/abs/1603.09536 or in https

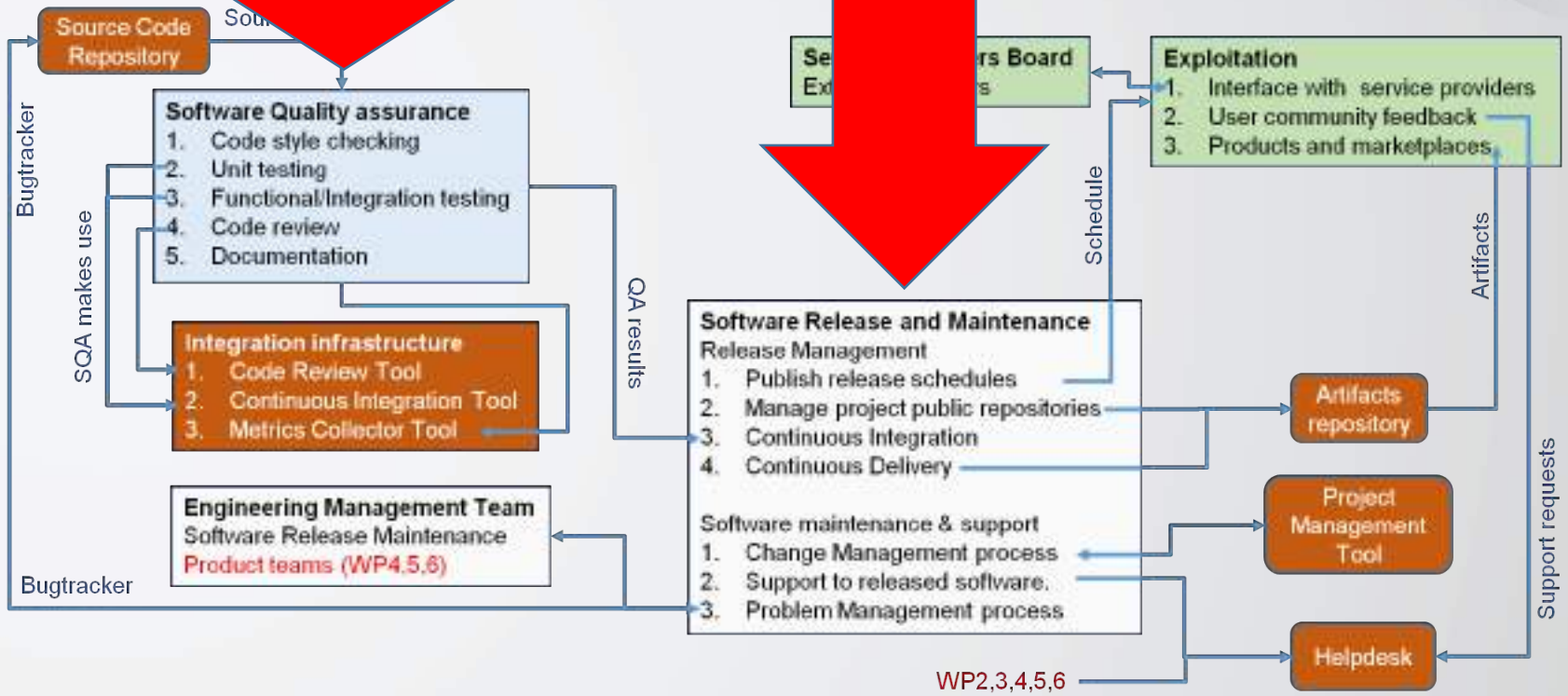
… to the implementation

software improvement cycle and the integration / release / software quality process



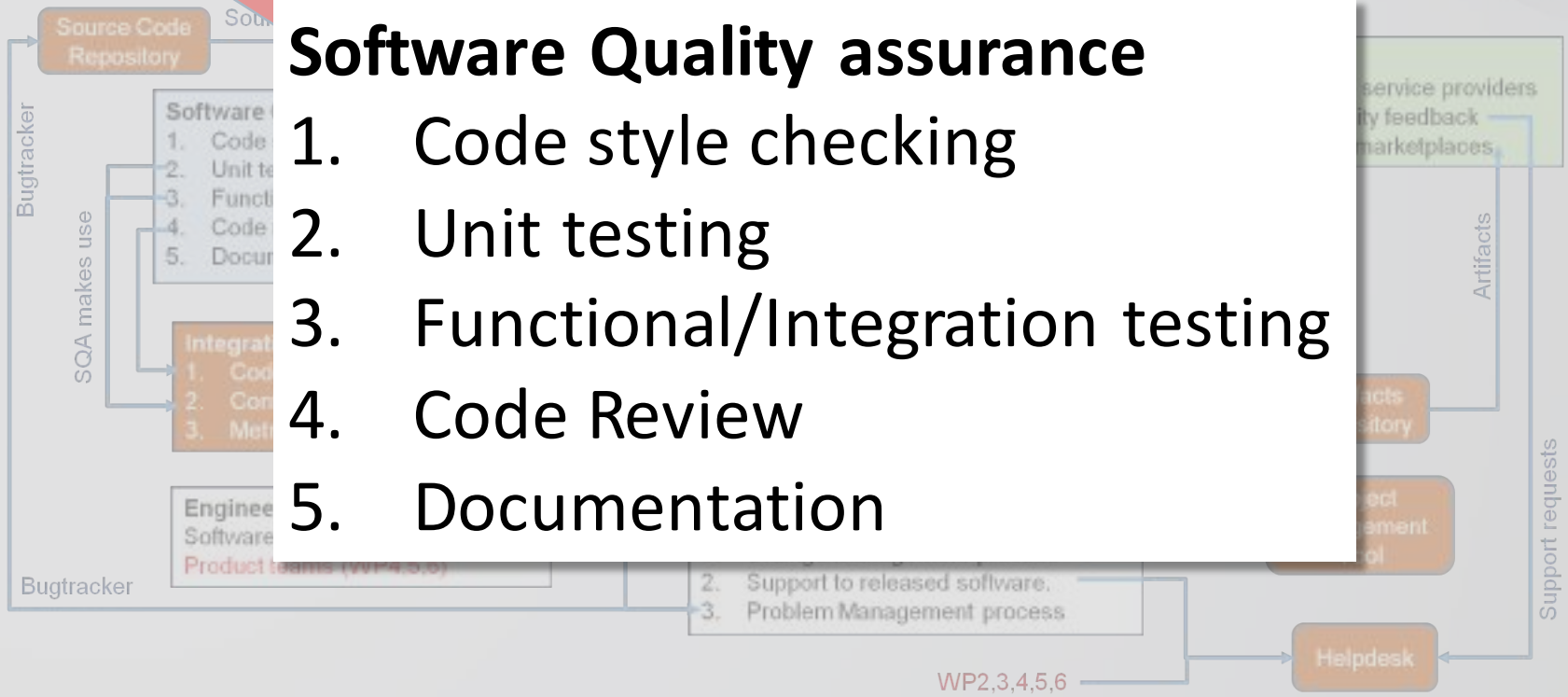
… to the implementation

software impro cycle and the integration / r oftware quality process



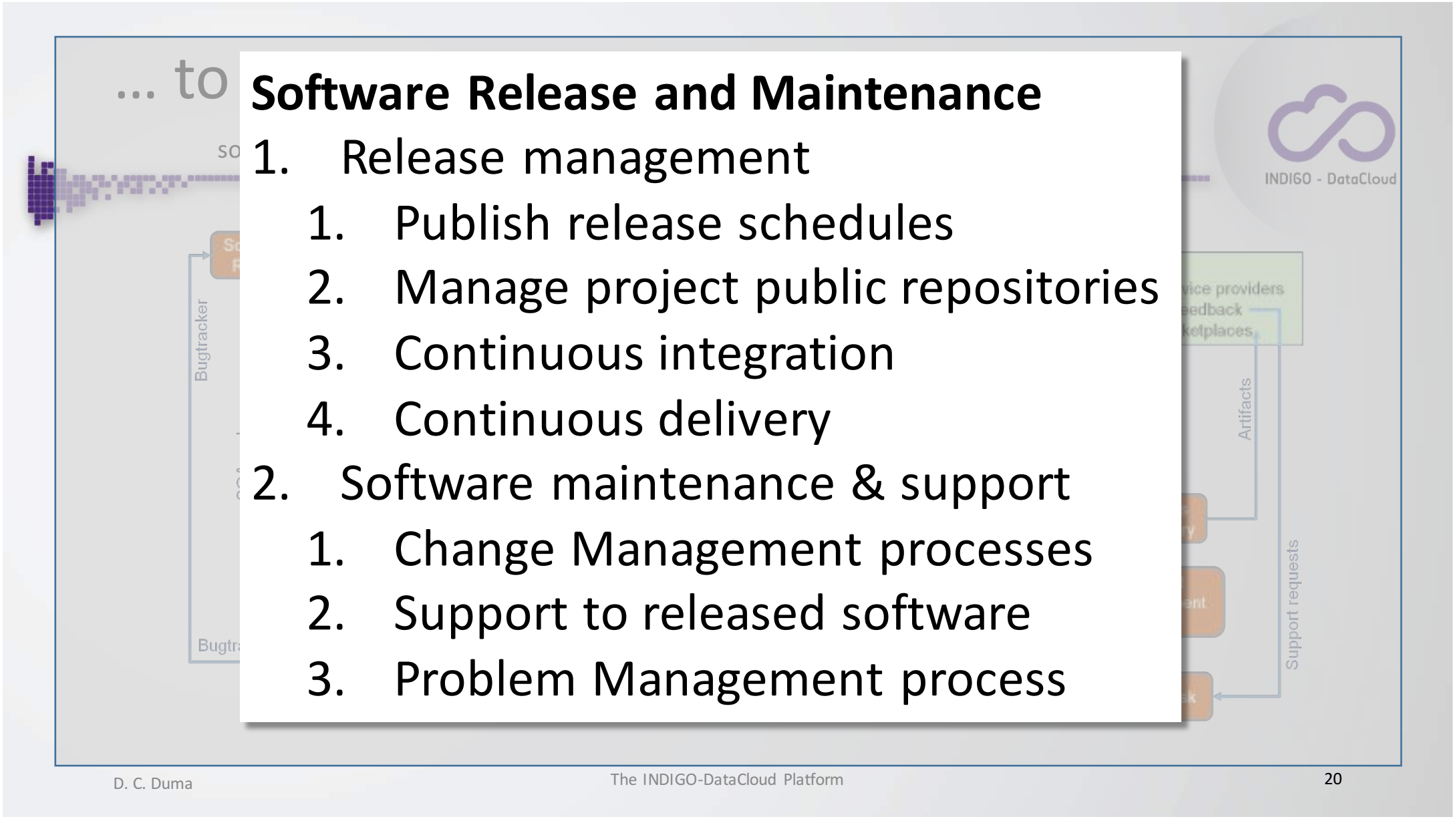
… to the implementation

software impro cycle and the integration / r oftware quality process



Software Quality assurance

1. Code style checking
2. Unit testing
3. Functional/Integration testing
4. Code Review
5. Documentation



Software Release and Maintenance

1. Release management
    1. Publish release schedules
    2. Manage project public repositories
    3. Continuous integration
    4. Continuous delivery
2. Software maintenance & support
    1. Change Management processes
    2. Support to released software
    3. Problem Management process

Timeline of the INDIGO releases

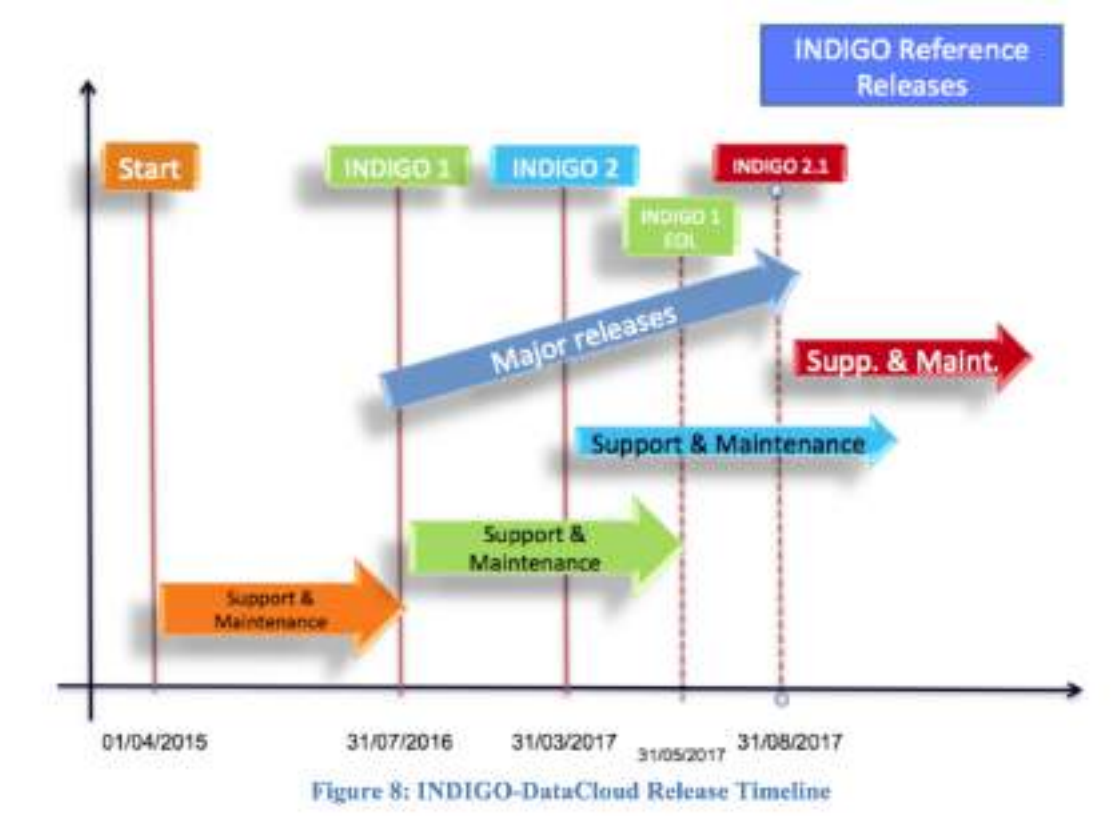


Figure 8: INDIGO-DataCloud Release Timeline

- **Release 1: MidnightBlue, August 2016**
- **Release 2: ElectricIndigo, April 2017**

IaaS Features (1)

- **Improved scheduling for allocation of resources** by popular open source Cloud platforms, i.e. OpenStack and OpenNebula.
  - Enhancements address both better scheduling algorithms and support for spot-instances. The latter are in particular needed to support allocation mechanisms similar to those available on public clouds such as Amazon and Google.
  - We also support dynamic partitioning of resources among “traditional batch systems” and Cloud infrastructures (for some LRMS).
- **Support for standards in IaaS resource orchestration engines** through the use of the TOSCA standard.
  - This overcomes the portability and usability problem that ways of orchestrating resources in Cloud computing frameworks widely differ among each other.
- **Improved IaaS orchestration capabilities** for popular open source Cloud platforms, i.e. OpenStack and OpenNebula.
  - Enhancements include the development of custom TOSCA templates to facilitate resource orchestration for end users, increased scalability of deployed resources and support of orchestration capabilities for OpenNebula.

IaaS Features (2)

- **Improved QoS capabilities of storage resources**.
  - 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.
- **Improved capabilities for networking support**.
  - Enhancements will include flexible networking support in OpenNebula and handling of network configurations through developments of the OCCI standard for both OpenNebula and OpenStack.
- **Improved and transparent support for Docker containers**.
  - Introduction of native container support in OpenNebula, development of standard interfaces using the OCCI protocol to drive container support in both OpenNebula and OpenStack.

PaaS Features (1)

- **Improved capabilities in the geographical exploitation of Cloud resources**.
  - End users need not know where resources are located, since the INDIGO PaaS layer is hiding the complexity of both scheduling and brokering.
- **Standard interface to access PaaS services**.
  - Currently, each PaaS solution available on the market is using a different set of APIs, languages, etc. INDIGO uses the TOSCA standard to hide these differences.
- **Support for data requirements in Cloud resource allocations**.
  - Resources can be allocated where data is stored.
- **Integrated use of resources coming from both public and private Cloud infrastructures**.
  - The INDIGO resource orchestrator is capable of addressing both types of Cloud infrastructures through TOSCA templates handled at either the PaaS or IaaS level.

PaaS Features (2)

- **Distributed data federations** supporting legacy applications as well as high level capabilities for distributed QoS and Data Lifecycle Management.
  - This includes for example remote Posix access to data.
- **Integrated IaaS and PaaS support in resource allocations**.
  - For example, storage provided at the IaaS layer is automatically made available to higher-level allocation resources performed at the PaaS layer.
- **Transparent client-side import/export of distributed Cloud data**.
  - This supports dropbox-like mechanisms for importing and exporting data from/to the Cloud. That data can then be easily ingested by Cloud applications through the INDIGO unified data tools.
- **Support for distributed data caching mechanisms and integration with existing storage infrastructures**.
  - INDIGO storage solutions are capable of providing efficient access to data and of transparently connecting to Posix filesystems already available in data centers.

PaaS Features (3)

- **Deployment, monitoring and automatic scalability of existing applications**.
  - For example, existing applications such as web front-ends or R-Studio servers can be automatically and dynamically deployed in highly-available and scalable configurations.
- **Integrated support for high-performance Big Data analytics**.
  - This includes custom frameworks such as Ophidia (providing a high performance workflow execution environment for Big Data Analytics on large volumes of scientific data) as well as general purpose engines for large-scale data processing such as Spark, all integrated to make use of the INDIGO PaaS features.
- **Support for dynamic and elastic clusters of resources**.
  - Resources and applications can be clustered through the INDIGO APIs. This includes for example batch systems on-demand (such as HTCondor or Torque) and extensible application platforms (such as Apache Mesos) capable of supporting both application execution and instantiation of long-running services.

AAI Features

- INDIGO provides an advanced **set of AAI features** that includes:
  - User authentication (supporting SAML, OIDC, X.509)
  - Identity harmonization (link heterogeneous AuthN mechanisms to a single VO identity)
  - Management of VO membership (i.e., groups and other attributes)
  - Management of registration and enrolment flows
  - Provisioning of VO structure and membership information to services
  - Management, distribution and enforcement of authorization policies

Front-end integration schemas

- We provide graphical user interfaces in the form of **scientific gateways and workflows** through the INDIGO FutureGateway (FG). The FG can directly access the INDIGO PaaS services and software stack and allows to define and set up on-demand infrastructures for the use cases presented by our scientific communities.
  - Setting up whole use case infrastructure: The administrator will be provided with ready to use receipts that he will be able to customize. The final users will be provided with the service end-points and will not be aware of the backend.
  - Use the INDIGO features from their own Portals: User communities, having their **own Scientific Gateway setup**, can exploit the FutureGateway REST API to deal with INDIGO whole software stack.
  - Use of the INDIGO tools and portals, including the FutureGateway, Scientific Workflows Systems, Big Data Analytics Frameworks (such as Ophidia), Mobile Applications or Kepler extensions. In this scenario the final users as well as domain administrators will use the GUI tools. The administrator will use it as described in first case. In addition domain specific users will be provided with specific portlets/workflows/apps that will allow graphical interaction with their applications run via INDIGO software stack.

INDIGO-DATACLOUD FIRST PUBLIC RELEASE IS OUT!

INDIGO MIDNIGHTBLUE

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

https://www.indigo-datacloud.eu/service-component

The INDIGO Software

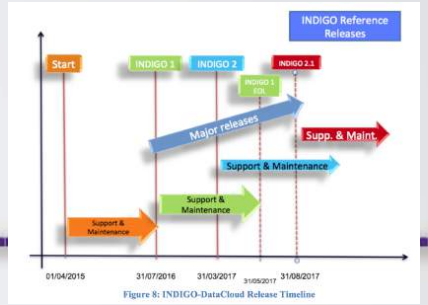


Figure 8: INDIGO-DataCloud Release Timeline

INDIGO - DataCloud

- Our second and final one is called **ElectricIndigo**
- **ElectricIndigo (https://www.indigo-datacloud.eu/service-component):**
  - 40 open source modular components, distributed via 170 software packages, 50 ready-to-use Docker containers
  - Supported operating systems: CentOS 7, Ubuntu 16.04
  - Supported cloud frameworks: OpenStack Newton, OpenNebula 5.x (plus connection to Amazon, Azure, Google)
  - Download it from the INDIGO-DataCloud Software Repository: http://repo.indigo-datacloud.eu/index.html

The ElectricIndigo Release

- The **ElectricIndigo modular software components** are organized around **5 areas**:

1. Application-level Interfaces to Cloud Providers and Automated Service Composition
   - For users porting their apps to the Cloud
2. Flexible Identity and Access Management
   - For users needing to handle AAI
3. Data Management and Data Analytics Solutions
   - For users managing distributed [big] data
4. Programmable Web Portals, Mobile Applications
   - For the creation of front ends
5. Enhanced and Scalable Services for Data Centers and Resource Providers
   - For providers wishing to optimize/enhance their service offerings

ElectricIndigo:
Application-level Interfaces for Cloud Providers and Automated Service Composition

- Easily **port applications to public and private Clouds** using open programmable interfaces, user-level containers, and standards-based languages to **automate definition, composition and instantiation of complex set-ups**.

- **Typical questions**: How can I run my application on Cloud provider X? What if I want to use Docker but my provider does not support it (e.g. also on HPC systems)? How do I automate the creation and management over public or private Clouds of dynamic clusters running multiple services?

ElectricIndigo:
Flexible Identity and Access Management

- **Manage access and policies to distributed resources** using multiple methods such as **OpenID-Connect, SAML, X.509** digital certificates, through **programmable interfaces and web front-ends**.

- **Typical questions**: How can I manage access to distributed resources by users, identified through diverse methods? (e.g. Google ID, digital certificates) How should I modify / write my apps to benefit from that?

ElectricIndigo:
Data Management and Data Analytics Solutions

- **Distribute and access data** through multiple providers via **virtual file systems and automated replication and caching**, exploiting scalable, **high-performance data mining and analytics**.

- **Typical questions**: How can I automatically replicate datasets to multiple sites? Can I transparently access my distributed datasets from my app? Can I cache the most accessed data, so that it's close to where users need it? How do I instantiate clusters and databases for big data analysis?

ElectricIndigo:
Programmable Web Portals, Mobile Applications

- **Create and interface web portals or mobile apps**, exploiting **distributed data as well as compute resources** located in public and private Cloud infrastructures.

- **Typical questions**: How can I easily provide my app with a pluggable, extensible web front-end capable of managing my workflows? Can this front-end interface with all the features provided by INDIGO? How can I write an INDIGO-enabled app for Android or iOS?

ElectricIndigo:
Enhanced and Scalable Services for Data Centers and Resource Providers

- **Increase the efficiency of existing Cloud infrastructures** based on OpenStack or OpenNebula through **advanced scheduling, flexible cloud / batch management, network orchestration** and interfacing of high-level Cloud services to existing storage systems.

- **Typical questions**: How can my cloud data centers provide flexible and fair scheduling policies for access to resources? How do I balance traditional vs. cloud resources in my data center? How do I connect novel INDIGO features to my existing systems? How can I manage storage Quality of Service?

INDIGO in practice

(some real apps already integrating INDIGO components)

- LifeWatch: algae bloom modeling
- RNA sequencing with TRUFA
- Deploying an elastic, complex cluster on the Cloud with INDIGO components
- Cloudified services for molecular dynamics
- A distributed archive system for the Cherenkov Telescope Array (CTA)
- The Large Binocular Telescope (LBT) distributed archive
- INDIGO's Ophidia for the calibration of astronomical images
- Launching POWERFIT and DISVIS VMs on the EGI FedCloud using INDIGO tools
- POWERFIT and DISVIS web portals: harnessing GPGPUs on the Grid using INDIGO udocker
- Automated deployment of an Ophidia big data analytics cluster
- INDIGO at the Central Institute for the Union Catalogue of Italian Libraries and Bibliographic Information
- EGI and INDIGO integration
- ELIXIR-ITALY: developing a Galaxy instance provider platform
- Multidisciplinary Oceanic Information System
- Deploy a Zenodo-based repository in the cloud using Marathon for DARIAH
- On-demand analysis and big data infrastructures for the CMS LHC experiment
- Theoretical physics on HPC clusters using INDIGO udocker

INDIGO and Industry

- We collaborated with several large companies such as ATOS, INDRA, T-Systems, IBM and others to facilitate the adoption and enhancement of INDIGO components.
- In the HelixNebula ScienceCloud context, we successfully ran production scientific workloads on private and commercial clouds, proving that INDIGO is really IaaS-independent and can work at scale on hybrid clouds.
- SMEs such as QUIBIM adopted INDIGO components to complement their own services.

Collaborations with commercial providers

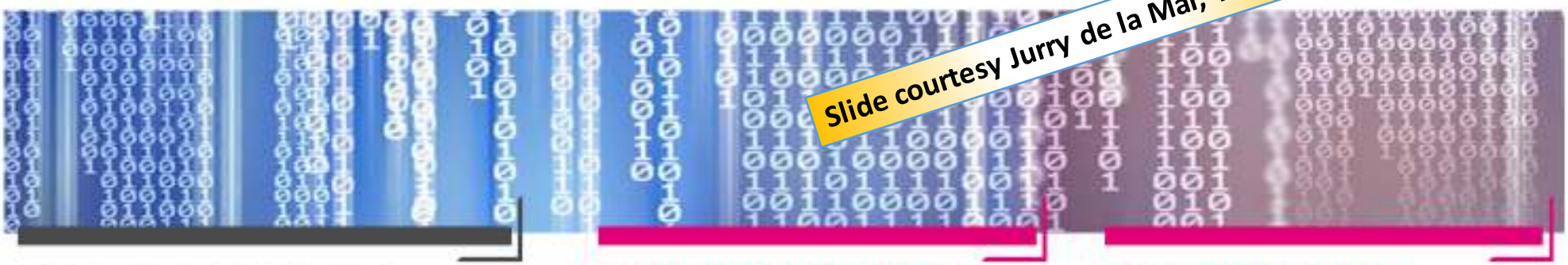
- See https://developer.ibm.com/opentech/2017/05/18/cloud-computing-better-science-recap-egi-conference-indigo-datacloud-summit-2017/ for a summary of the INDIGO Summit 2017 by Dr Sahdev Zala of **IBM**
  - With some details about the ongoing collaboration between IBM and INDIGO-DataCloud
- See https://indico.egi.eu/indico/event/3249/session/48/contribution/98/material/slides/0.pdf for info on the integration of INDIGO tools into the **Open Telekom Cloud portfolio**, the public cloud offering of T-Systems (a Deutsche Telekom unit)

INDIGO-DATACLOUD @OPEN TELEKOM CLOUD

SERVICES RUNNING ON OPEN TELEKOM CLOUD

INDIGO - DataCloud

Slide courtesy Jurry de la Mar, T-Systems



USER INTERFACE & CONTROL

- GUI CLIENTS
  - GATEWAY PORTAL
  - MOBILE CLIENTS
  - WORK-FLOWS
  - SCIENCE GATEWAY
- SUPPORT SERVICES
- ...

PLATFORM & AUTOMATION

- PAAS ORCHESTRATOR (KUBERNETES)
- DATA SERVICES
- INFRASTRUCTURE MANAGER
  - ACCOUNT-ING
  - IAM SERVICE
  - MESOS CLUSTER
  - ...

COMPUTE & STORAGE

- SMART SCHEDULING
- STORAGE SERVICES
- LOCAL REPOSITORY
  - HEAT / IM
  - NATIVE DOCKER
  - QOS SUPPORT
  - ...

T · · Systems ·

INDIGO-Datacloud meets OTC 11/05/2017 7/7

INDIGO and External Projects: Components and Patches Merged in Upstream Open Source Projects

- OpenStack (https://www.openstack.org)
  - Nova Docker
  - Heat
  - OpenID-Connect for Keystone
  - Pre-emptible instances support (under discussion)
- OpenNebula (http://opennebula.org)
  - OneDock
- Infrastructure Manager (http://www.grycap.upv.es/im/index.php)
- Clues (http://www.grycap.upv.es/clues/eng/index.php)
- Onedata (https://onedata.org)
- TOSCA adaptor for JSAGA (http://software.in2p3.fr/jsaga/dev/)
- OCCI implementation for OpenStack (https://github.com/openstack/ooi)
- Extended AWS support for rOCCI in OpenNebula. Python and Java libraries for OCCI support.
- CDMI and QoS extensions for dCache (https://www.dcache.org)
- Workflow interface extensions for Ophidia (http://ophidia.cmcc.it)
- OpenID Connect Java implementation for dCache (https://www.dcache.org)
- MitreID (https://mitreid.org/) and OpenID Connect (http://openid.net/connect/) libraries

INDIGO & EOSC **in production**: >= TRL8*

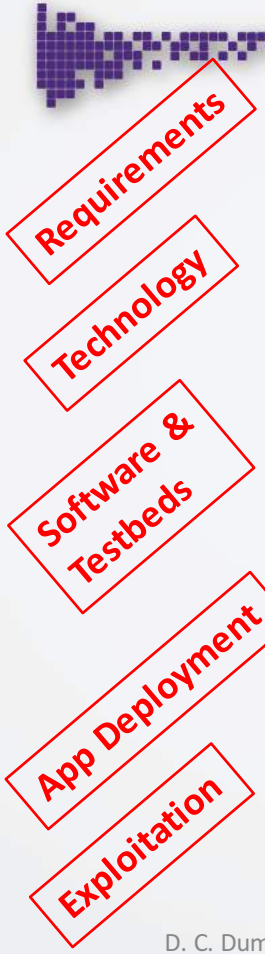
- Several **INDIGO solutions and activities** are in the new **EOSC-hub proposal**, submitted by a consortium of 74 partners **under the coordination of EGI, EUDAT and INDIGO-DataCloud**
- With **INDIGO components** such as Identity and Access Management, Token Translation, Virtual filesystems (Onedata), Advanced IaaS Services, the Infrastructure Manager, the INDIGO PaaS and its orchestrator, web front-end services, user-level containers
- INDIGO was also given the **overall technical coordination of the project** and will contribute to **training, support, external liaison, stakeholder engagement, policy**.

*: **TRL** = **T**echnology **R**eadiness **L**evel. TRL 8 means production-ready at scale

INDIGO & EOSC in evolution: < TRL8

- **Novel features evolving INDIGO components** are a key part of two proposals to the **EINFRA-21-2017 call (eXtreme-DataCloud and DEEP-Hybrid DataCloud**), both recently approved and due to start in late 2017:
  - Intelligent dataset distribution and data lifecycle management
  - Smart caching
  - Orchestrating Computing Workflows based on policy driven or adaptive data movements
  - Flexible metadata management for big data sets
  - Access to bare-metal resources on the Cloud
  - PaaS-Level access to HPC resources
  - Extensions to the INDIGO Orchestrator for hybrid IaaS deployments and scale out to 3rd party clouds
  - Extensions to the INDIGO Virtual Router Appliance
  - Real-time, streaming-based data ingestion and processing

Conclusions

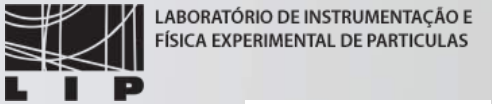


- **In 30 months**, the INDIGO-DataCloud project realized a **comprehensive involvement of many Research Communities and providers** for the definition and tracking of their requirements.
- We identified **technology gaps** linked to several concrete use cases in multiple fields, defined, published and implemented the **overall INDIGO architecture**.
- We **produced two major software versions and 14 minor updates**, releasing 40 open modular components. We did that exploiting key European know-how, **reusing** and **extending** open source software, and **contributing** to upstream projects. We established software development and management processes, and defined development and pre-production distributed testbeds.
- **Production deployment of many applications making use of the INDIGO software** is well underway, and INDIGO components have been selected for production use in big infrastructures, commercial companies, external projects.
- **Several opportunities for further exploitation of INDIGO components** are being explored and implemented, in the context of the EOSC and beyond.

Many Thanks

https://www.indigo-datacloud.eu

Better Software for Better Science.



@indigodatacloud

www.indigo-datacloud.eu

https://www.facebook.com/indigodatacloud/

Thank you

INDIGO - DataCloud

https://www.indigo-datacloud.eu

Better Software for Better Science.

 @indigodatacloud

 www.indigo-datacloud.eu

 https://www.facebook.com/indigodatacloud/

Backup slides

The PaaS Orchestrator

- It **takes deployment requests**, expressed through templates **written in TOSCA** YAML, and deploys them **on the best cloud site** available.
- The Orchestrator **collects all the information needed** to deploy a service consuming others PaaS µServices APIs:
  - QoS/SLA Service: get the prioritized list of SLAs of the user
  - Monitoring Service: get the status of the underlying IaaS platforms and their resource availability;
  - CloudProviderRanker Service: sort the list sites on the basis of rules defined per user/group/use-case;
- The orchestrator **delegates the deployment** to some other component (see later)
- This is an essential components to manage multiple clouds.

The Cloud Information Provider

- A service **retrieving available images and containers** at each IaaS site:
  - Interacting with the local Cloud Middleware Programmable via APIs
  - Formatting output as JSON using a template
- Importing images' definitions inside CMDB using REST API
  - Purging obsolete images

The Monitoring Service

- INDIGO Functionality:
  - Monitor the **PaaS Core Services**
    - Using e.g. Heapster in the Kubernetes cluster.
  - Monitor the **customized virtual infrastructures**
    - Using e.g. Zabbix
  - Monitor the **state of the sites**
    - Service Availability and Response Time Monitoring
- Monitoring information is **centralized using Zabbix** as collector
- Monitored information **exposed via a REST API** to be consumed by other services.

The Accounting Service

- Accounts for resource usage on the INDIGO PaaS and provides that data to other INDIGO-DataCloud services
  - QoS/SLA service will use information gathered by the Accounting service (and the monitoring pillar) to monitor SLA violation
- Usage data is extracted from the system where the resources are used and sent to a central repository
- The repository aggregates the data from across the infrastructure to produce totals based on a number of fields - such as user, site, month, year, etc.

The Infrastructure Manager (IM)

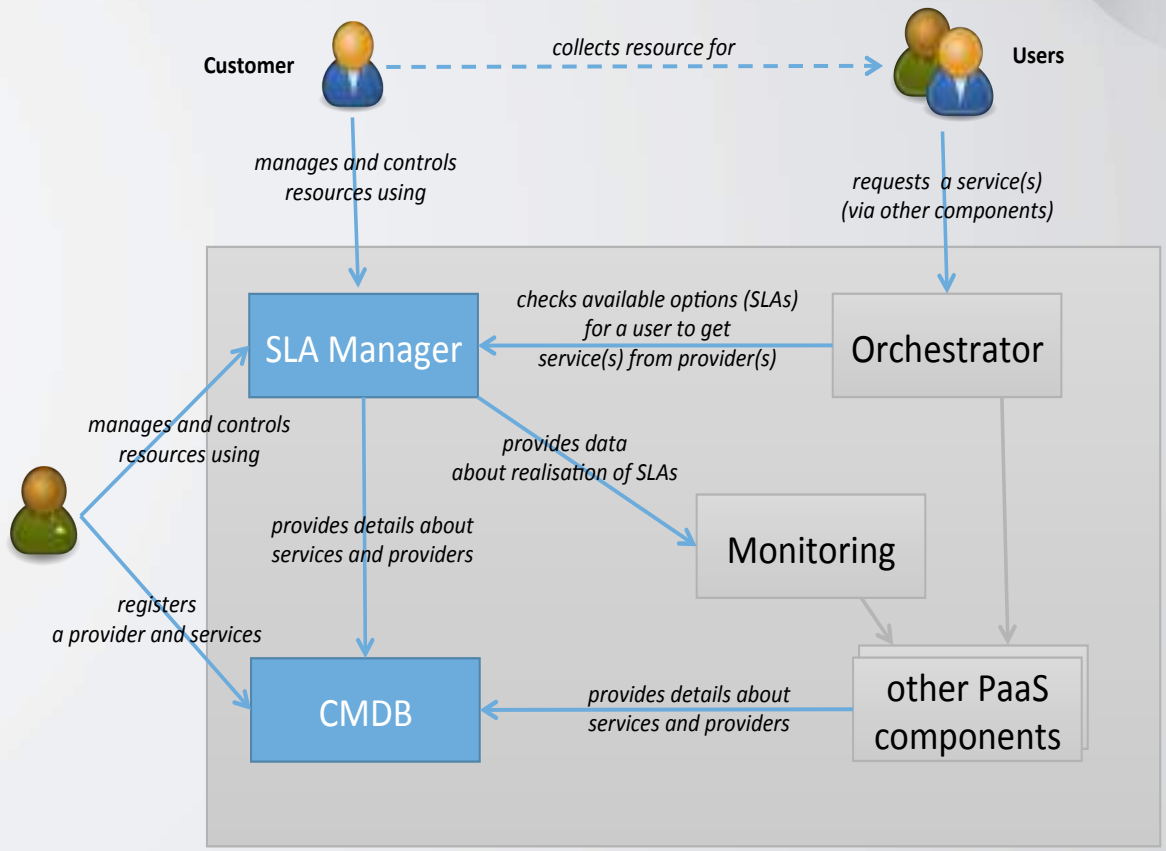
- The IM is a tool that **deploys complex and customized virtual infrastructures on multiple back-ends**.
- Powered by **Ansible**, the IM provides recipes for common deployments.
- **Support for TOSCA** 1.0 YAML specification with the custom node types described in the INDIGO project.
- Support for the Identity and Access Management Service (**IAM**).
- Support for the **Token Translation Service** (TTS) to support IAM authentication on OpenNebula Clouds.
- Improvements to access OpenStack Clouds that support IAM.
- IM can also deploy the virtual infrastructures described in the **TOSCA templates in external Clouds** (either public or on-premises).
  - Implements a TOSCA runtime that can work with a large set of cloud providers: OpenNebula, OpenStack, **Google Cloud Platform, Microsoft Azure, AWS**, EGI FedCloud (via OCCI), FogBow.

Managed Services / Application Deployment Service

- This service is in charge of **scheduling, spawning, executing and monitoring** applications and services on a distributed infrastructure.
- The core of this component consists of an elastic **Apache Mesos** [http://mesos.apache.org] cluster with slave nodes dynamically provisioned and distributed on the IaaS sites.
- Apache Mesos provides efficient resource isolation and sharing across distributed applications (frameworks).
- The INDIGO PaaS uses:
  - **Marathon** to deploy, monitor and scale **Long-Running services**, ensuring that they are always up and running.
  - **Chronos** to run user **applications** (jobs), taking care of fetching input data, handling dependencies among jobs, rescheduling failed jobs.

The Configuration Management Database (CMDB)

- CMDB, the **Configuration Management Database**, is a crucial component for proper management of services according **to ITIL®, ISO 20000, FitSM**.
  - It provides **configuration information about registered providers and services** certified in the infrastructure and other important data needed to operate and integrate the services.



The SLA/QoS Manager

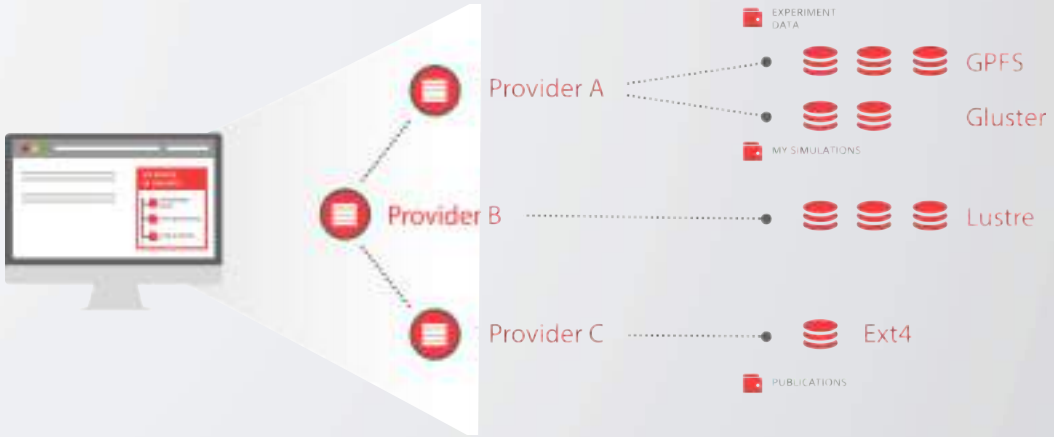
- The SLA Manager enables the possibility to define agreements between a Customer and a Provider for using resources and services in the form of an SLA. This opens the ways for more **SLA-aware infrastructures.**

Automated Scaling Service

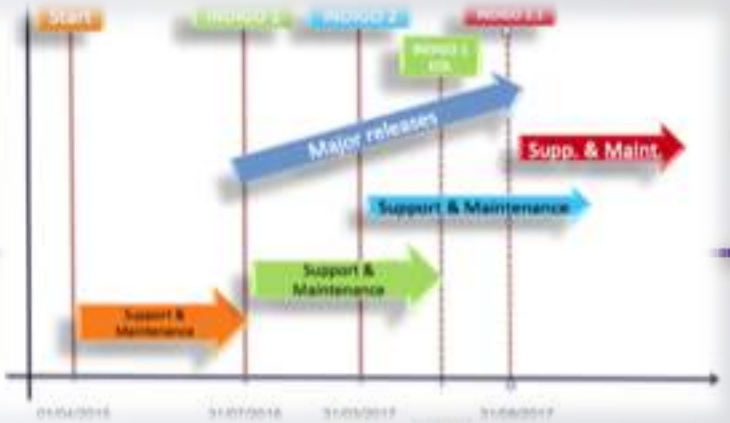
- In INDIGO, this is called CLUES. It is an **elasticity manager system** for HPC **clusters** and for **Cloud infrastructures** that features the ability to power on/**deploy** working nodes as needed (depending on the job workload of the cluster) and to power off/**terminate** them when they are no longer needed
  - It **supports** a large set of LRMS (Local Resource Management Systems): **SLURM**, **Torque**, **SGE**, **HTCCondor** and **Apache Mesos**.
- It is **integrated** with INDIGO-DataCloud's **PaaS Orchestrator** to automatically add/remove working nodes for the deployed virtual computer clusters (horizontal elasticity).

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



Release Timeline

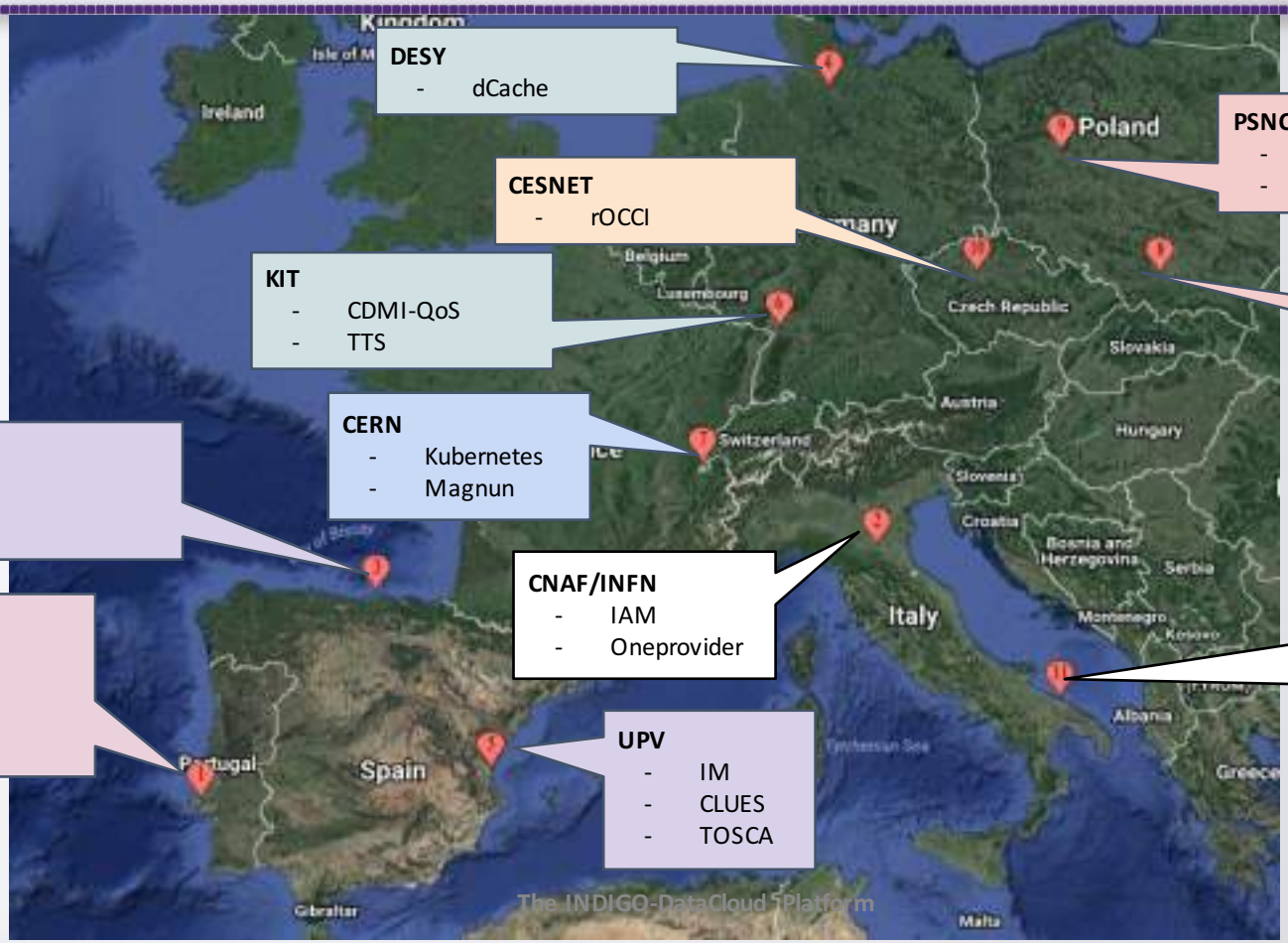


INDIGO - DataCloud

		Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
INDIGO-1	Full updates																		
	Standard updates																		
	Security updates																		
INDIGO-2	Full updates																		
	Standard updates																		
	Security updates																		

	Release Date	End of Full Updates	End of Standard Updates	End of Security Updates & EOL
INDIGO-1 MidnightBlue	08/08/2016	31/01/2017	31/03/2017	**31/05/2017**
INDIGO-2 ElectricIndigo	14/04/2017	30/09/2017	*30/11/2017*	*31/01/2018*

The INDIGO Development and Integration Infrastructure



DESY
- dCache

PSNC
- Indigokepler
- indigo-omt

CESNET
- rOCCI

KIT
- CDMI-QoS
- TTS

Cyfronet
- Onedata

CERN
- Kubernetes
- Magnun

IFCA/CSIC
- OOI
- OPIE

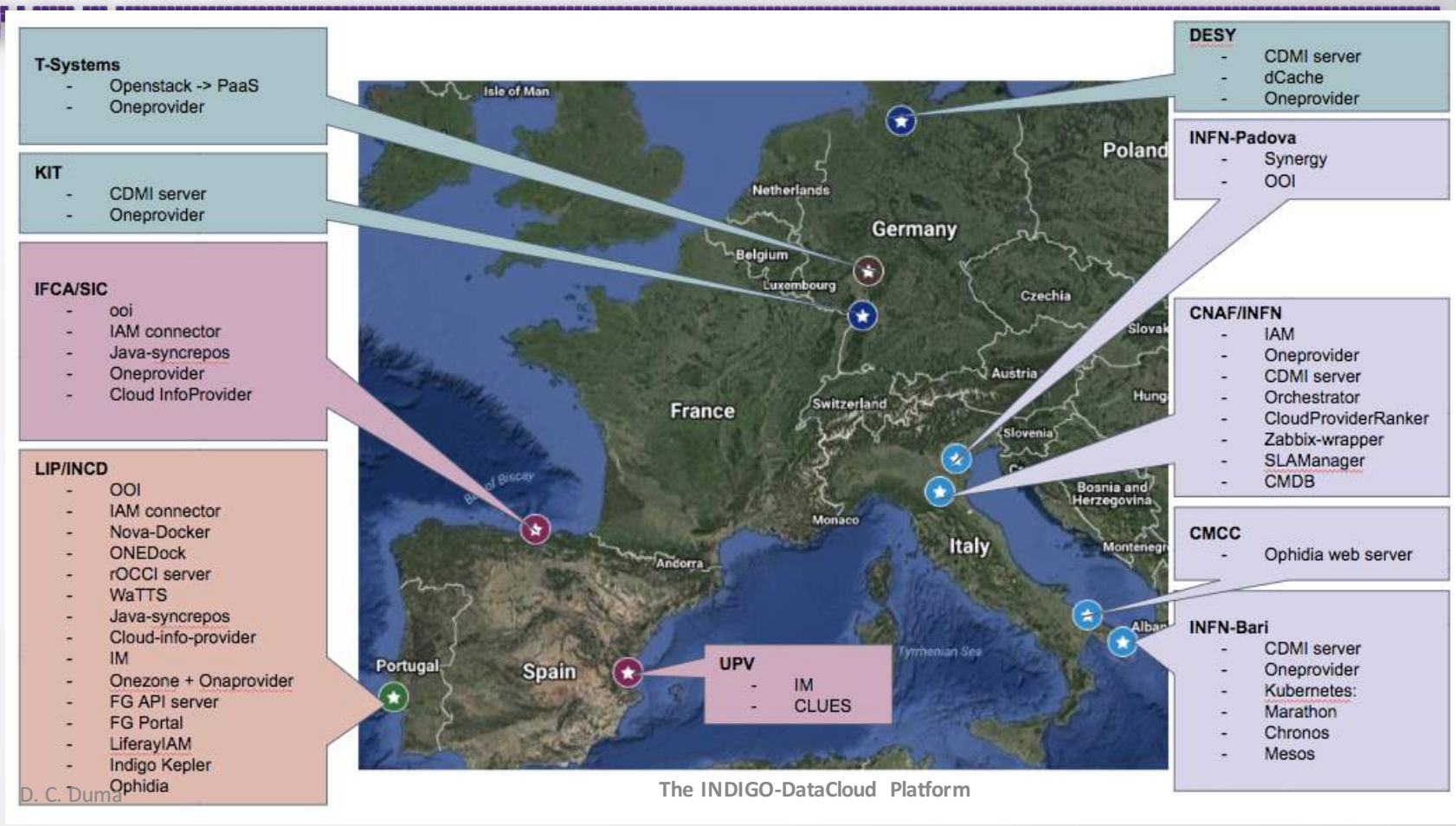
CNAF/INFN
- IAM
- Oneprovider

LIP/INCD
- OpenNebula: ONEDock
- Nova-Docker
- FutureGateway

INFN Bari
- Kubernetes
- Mesos
- Chronos

UPV
- IM
- CLUES
- TOSCA

The INDIGO Pilot Preview Testbed



The INDIGO-DataCloud Platform