

RIA-653549

INDIGO – DataCloud Technical Overview

Giacinto.Donvito@ba.infn.it INFN-Bari



Gap analysis Goals Architecture WPs activities Conclusions

INDIGO-DataCloud RIA-653549







Support federated identities and provide privacy and distributed authorization in open Cloud platforms

Performance issues limiting massive adoption of virtualized Cloud resources in large data centers.

Orchestrate and federate Cloud [public or private], Grid and HPC resources

The lack of flexible data sharing between groups' members and the difficulty in obtaining easy access to data generated by collaborating users working with different infrastructures or sites

Avoid software and vendor lock-in INDIGO-DataCloud RIA-653549

Gap Analysis

The barriers that limit the adoption of true PaaS solutions, such as the use of custom, non-interoperable interfaces and the limited availability of APIs for technology-independent storage access







Static allocation and partitioning of both storage and computing resources in data centers Exploit specialized hardware, such as GPUs or low-latency interconnections Manage dynamic and complex workflows for scientific data analysis Provide APIs to exploit the capabilities of the infrastructure and write applications, customizable portals and mobile views The current inflexible ways of distributing and deploying applications. The lack of an open and solid platform that permits to exploit distributed computing and storage resources through transparent

network interconnections. INDIGO-DataCloud RIA-653549

Gap Analysis





Development of a Platform based on open source software, without restrictions on the e-Infrastructure to be accessed (public or commercial, GRID/Cloud/HPC) or its underlying software (OpenStack or OpenNebula) providing the interface between e-Infrastructures and

Platforms.

the form of science gateways and access libraries.

Provide high-level access to the platform services in Streamline the adoption of the software products.

INDIGO-DataCloud RIA-653549





INDIGO - DataCloud

Develop an open framework, based on open source software

Taking into consideration the medium and long-term exploitation and sustainability

Exploit existing solutions, according to user requirements, and having in mind the expected evolution of technology.

Keep a direct contact with e-Infrastructure providers, to assure a successful deployment of the INDIGO Platform

The framework will offer to final users, as well as to developers, a low learning curve

INDIGO-DataCloud RIA-653549

INDIGO approach



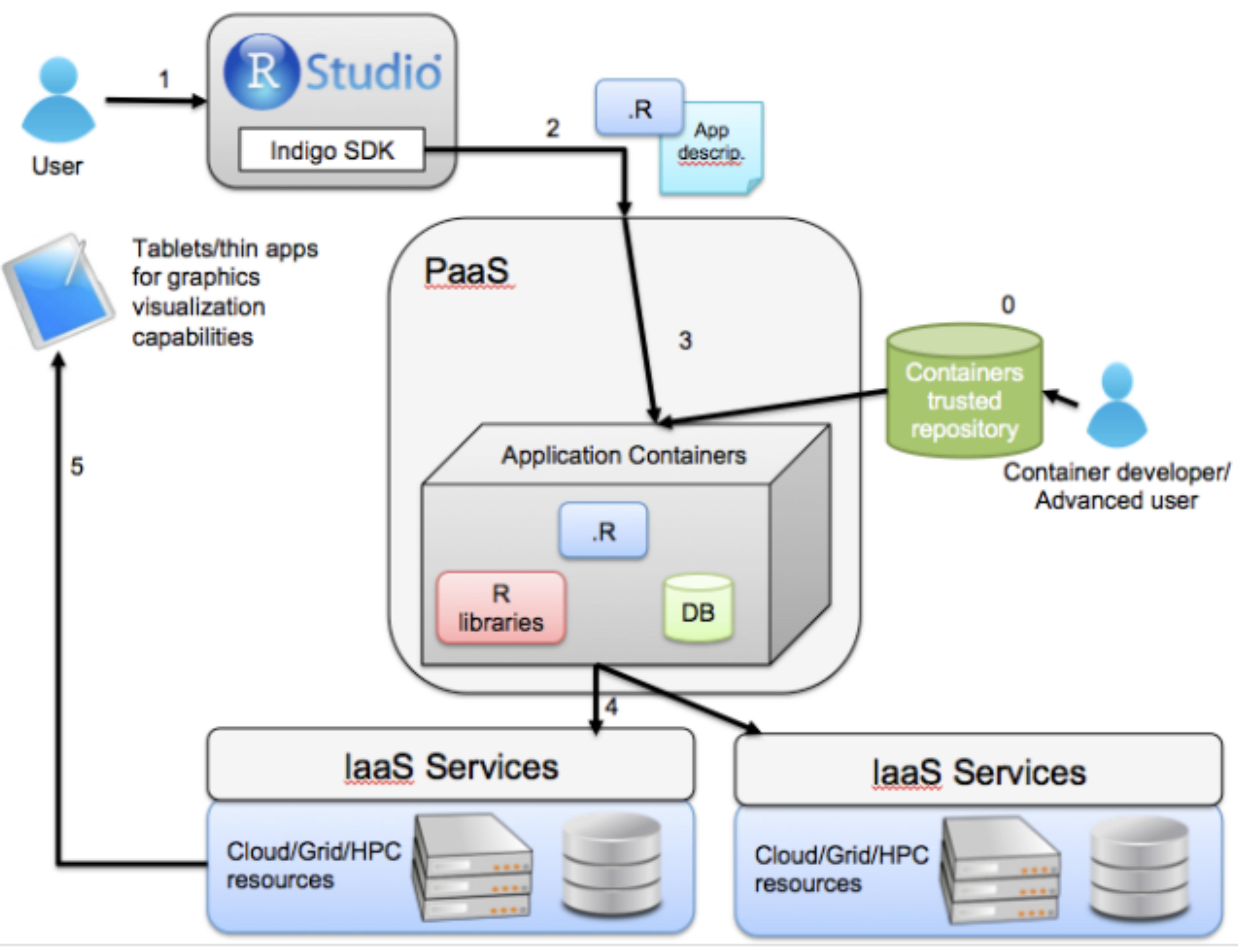
Whenever possible exploit general solutions instead of specific and dedicated tools/services or put effort in increasing the generality of tools developed in a given community this will be important for sustainability of the architecture Existing software like ROOT, OCTAVE/MATLAB, MATHEMATICA or R-STUDIO, will be supported and offered in a transparent way while running on powerful remote e-Infrastructure

resources.

INDIGO-DataCloud RIA-653549

INDIGO approach







INDIGO-DataCloud RIA-653549

Use Cases

Figure 1: use case of supporting R-Studio through INDIGO

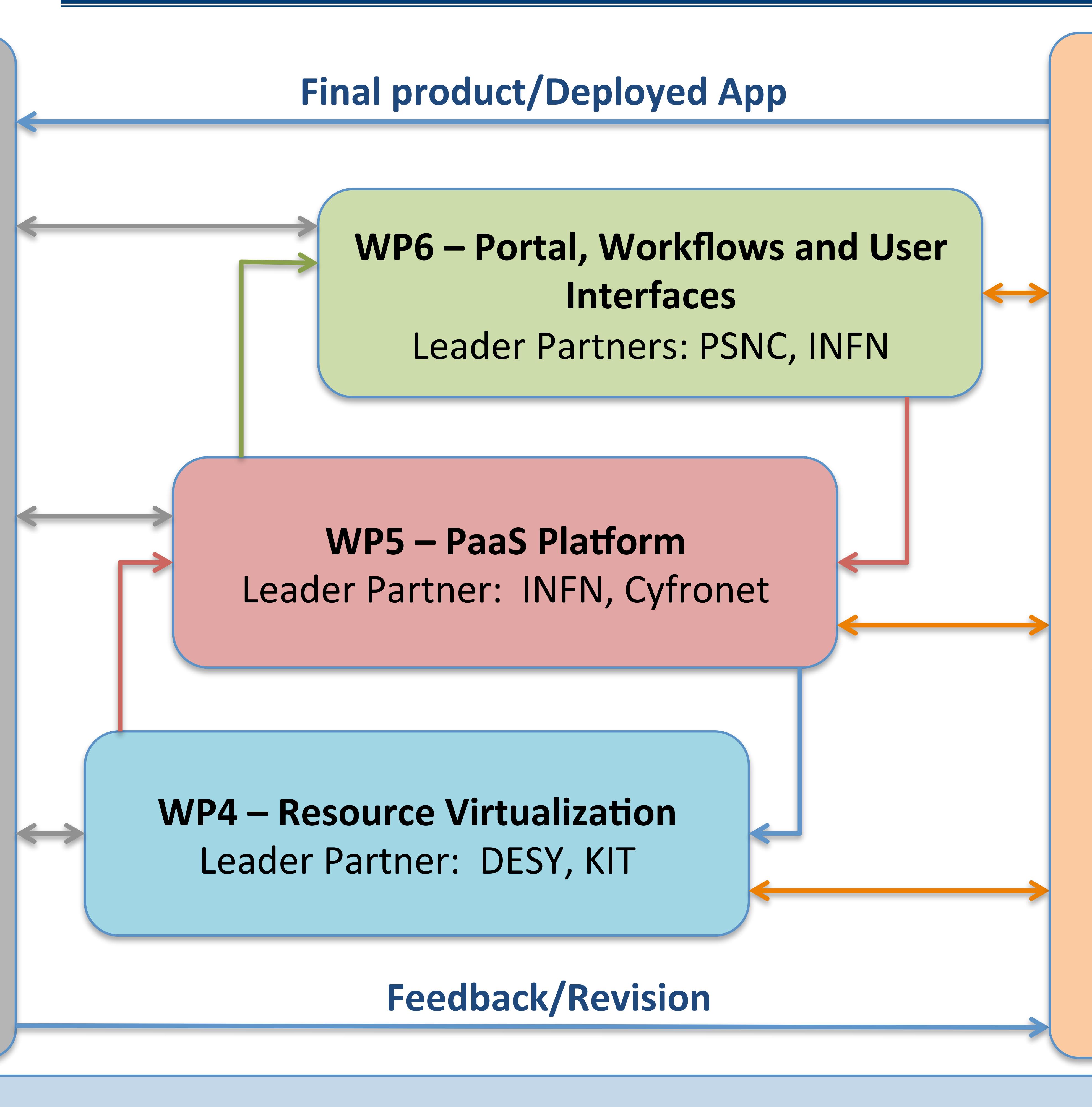




ities Safet Safet Safet 0 C **C** Π C R **t**O J D S S ų. σ Δ D D \mathbf{O} σ

INDIGO-DataCloud RIA-653549

Work Packages



WP1 – Administrative and Technical Management Leader Partners: INFN, CSIC

J ω S Ω D \mathbf{O} DD U S **CDS**

9

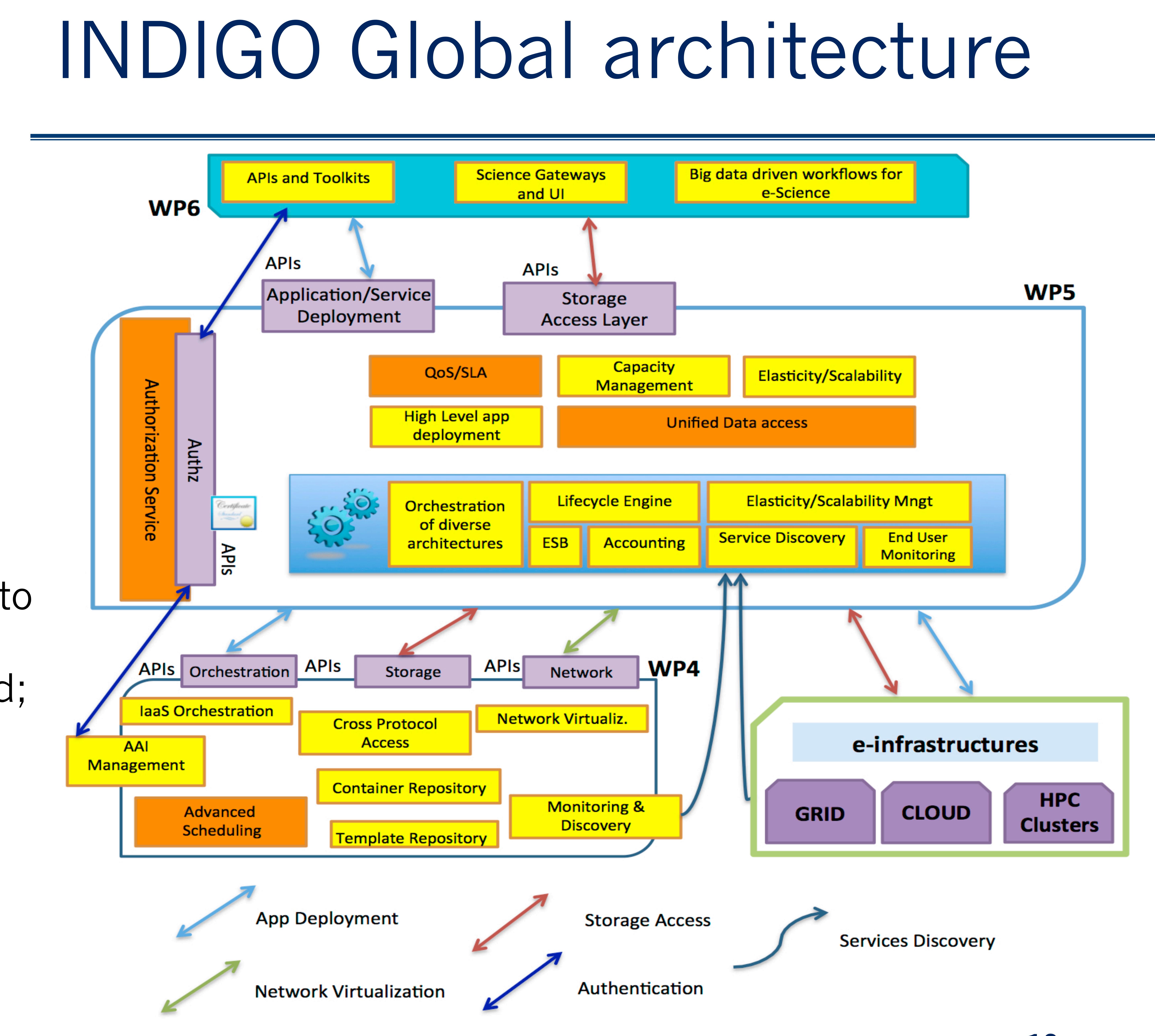


INDIGO global architecture.

Color codes:

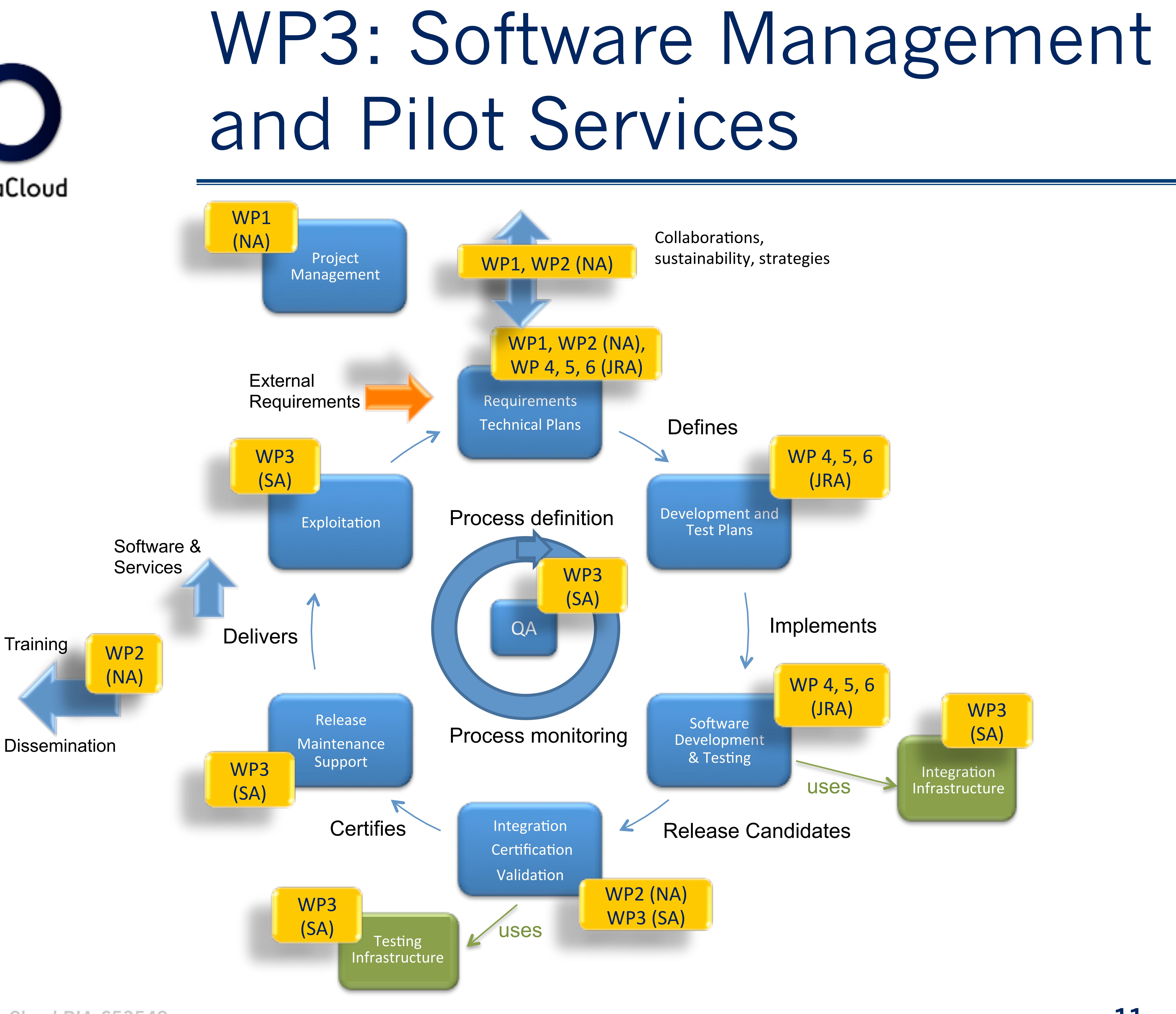
Yellow: implementation based on already available solution to be improved/changed;

Orange: Completely new services to be Implemented









INDIGO-DataCloud RIA-653549





INDIGO - DataCloud T3.1 will deal with the software quality assurance, compiling and enforcing the necessary quality criteria, indicators, and tests necessary to ensure high quality software components ready for production.

T3.2 will make the certified software components available as a set of coherent high quality releases, supported by an efficient maintenance process.

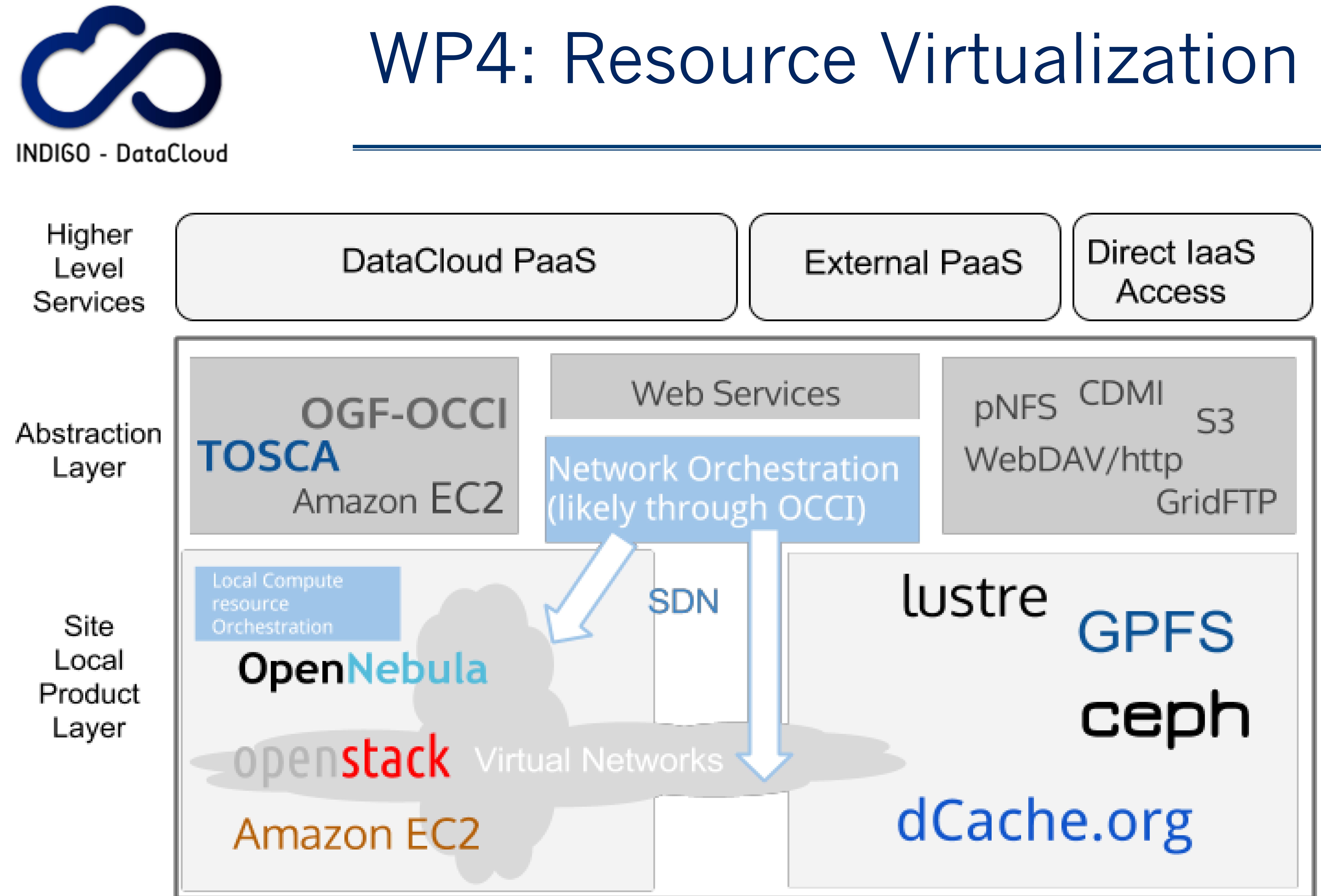
T3.3 will provide the pilot infrastructures and services for integration and testing supporting the tasks T3.1 and T3.2 activities.

T3.4 will interface with major production e-infrastructures, collect their feedback, requirements, and will enable a path towards production exploitation.

INDIGO-DataCloud RIA-653549

WP3: Software Management and Pilot Services





INDIGO-DataCloud RIA-653549



Common subtask:

AAI Management for the virtualized computing cloud infrastructure To ensure the integration of federated AAI technologies into OpenStack, OpenNebula, CEPH, dCache and other supported INDIGO products, allowing users to access infrastructure resources using their **home** or guest **IdP** account.

INDIGO-DataCloud RIA-653549

WP4: Resource Virtualization

Service Discovery and Monitoring To extend existing local-site monitoring services for all INDIGO products to provide to higher-level services of WP5 and WP6 with monitoring and accounting information through a query-API.



Cloud Computing Virtualization

Providing support for containers as a portable and performant platform for the execution and deployment of applications, and by providing local site orchestration features (e.g. HEAT or OneFlow) that simplify the management of the lifecycle of laaS: both containers and VMs.

INDIGO-DataCloud RIA-653549

WP4: Resource Virtualization

improving compute orchestration and scheduling center and improving its response to the users.

Improving the on-demand compute capabilities of data-centers by To improve the existing cloud schedulers in OpenStack and **OpenNebula** to include the support for postponing low priority workloads (by killing, preempting or stopping running containers or VMs) in order to allocate higher priority requests, thus enabling the advanced scheduling policies, optimizing the usage of the data

15



Cloud Storage Virtualization

QoS Support in storage service regardless of the underlying implementation.

supporting the defined service levels.

protocols.

INDIGO-DataCloud RIA-653549

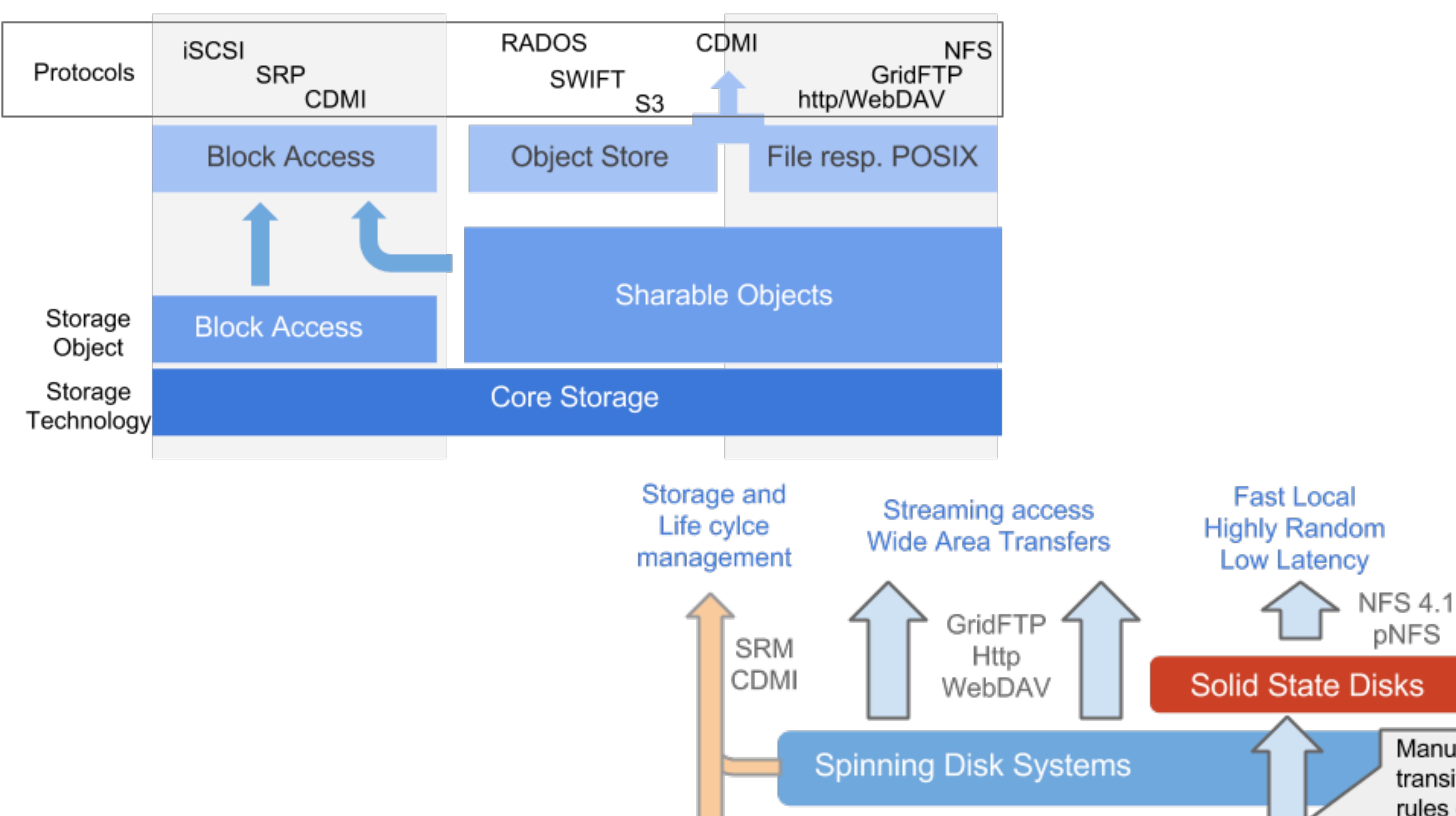
Cross Protocol support for storage solutions Use cases often require storing files with one access protocol and subsequently accessing the same data with a different protocol. This requires enabling access to identical data via different

WP4: Resource Virtualization

will enable users to specify service quality policies for their data. In collaboration with RDA, we envision standardizing the associated terms and definitions, so users can expect the same quality of Evaluate/extend available protocols (e.g. CDMI, WebDAV, SRM)

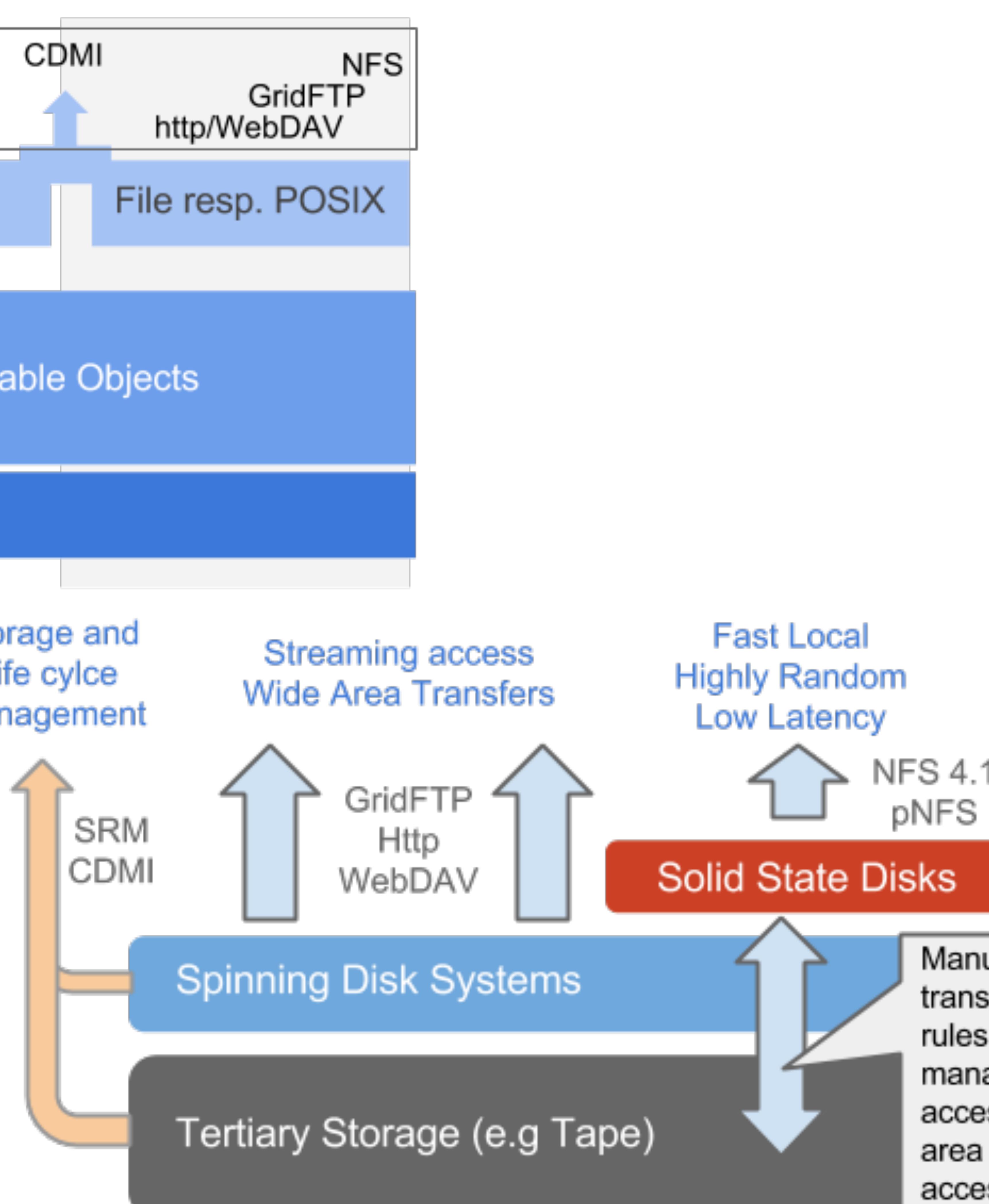






INDIGO-DataCloud RIA-653549

WP4: Resource Virtualization



Manual or automatic transition based on rules (lifecycle management) or access profiles (wide area vs. fast random access)

17



Network Virtualization

INDIGO-DataCloud RIA-653549

Enhancing the capabilities of Local virtual networks active elements such as switches or routers.

Use of SDNs to set up virtual networks spanning multiple sites Evaluating available technologies to enable spanning of virtual networks across multiple sites, including advanced services, such as cross-site load balancing or providing assured bandwidth for use by various PaaS applications.

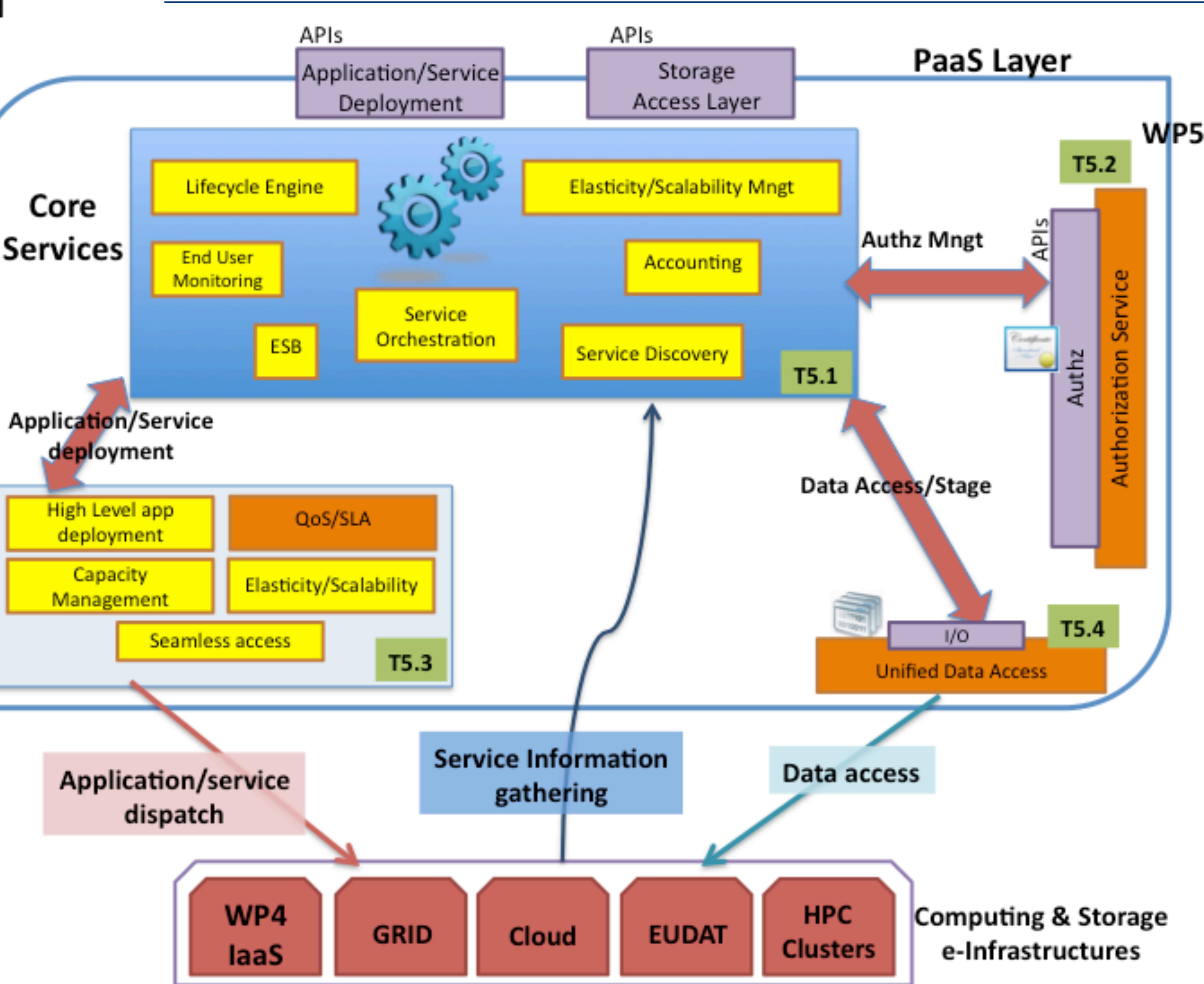
Provisioning of such spanning networks should be available on an automated basis, using common cloud management platforms such as OpenNebula or OpenStack.

WP4: Resource Virtualization

features comparable to real-life physical networks, including a predefined complex topology of such a network or the presence of

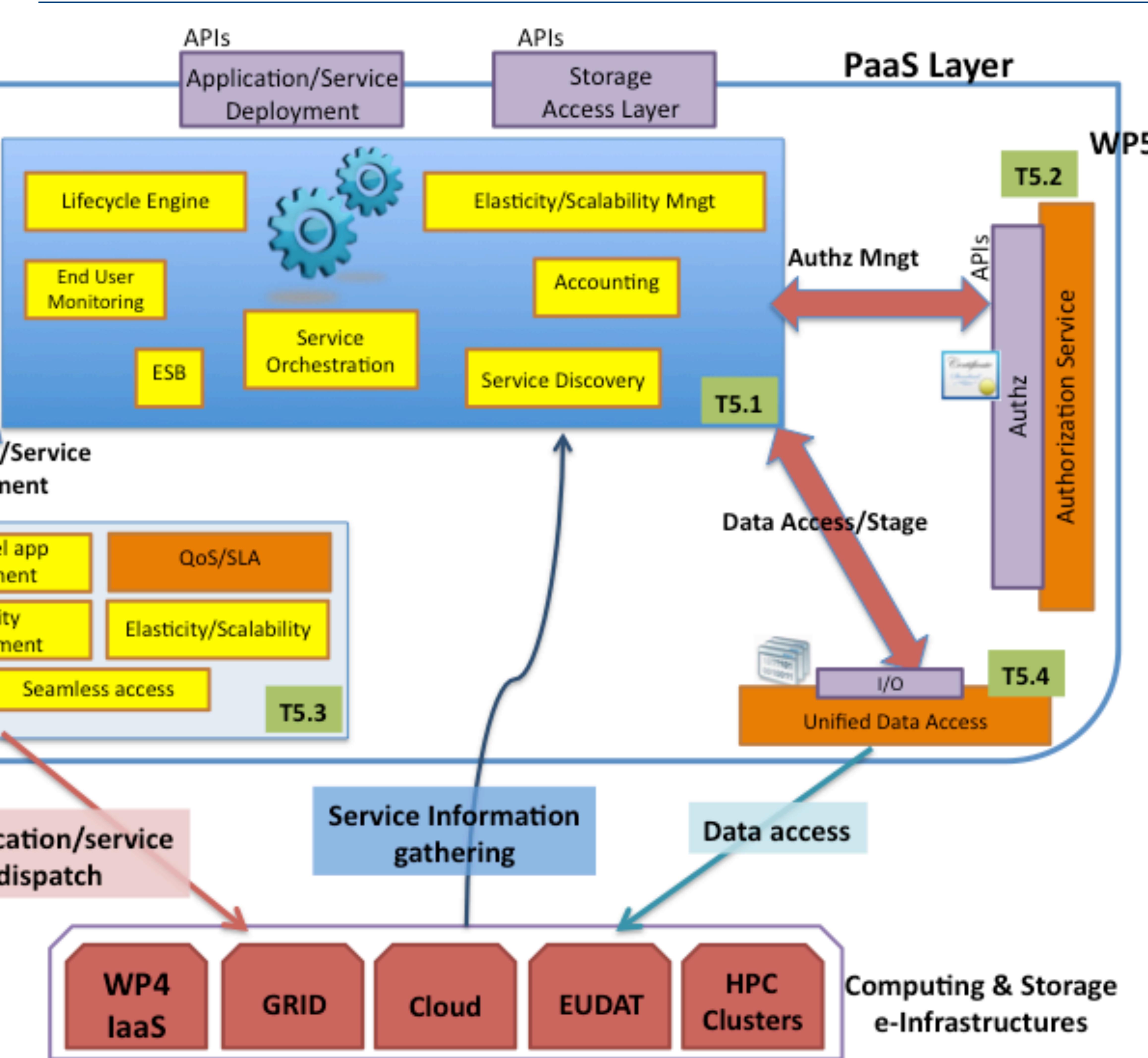






INDIGO-DataCloud RIA-653549

WP5: PaaS Platform Development







features: Transparency Error management Elasticity/SLA management.

INDIGO-DataCloud RIA-653549

WP5: PaaS Platform Development

PaaS architecture and implementation Design and implementation of a PaaS layer allowing scientific communities to exploit, in a powerful and high-level way, several heterogeneous computing and data e-infrastructure such as: laaS Cloud, Helix Nebula, EGI Grid, EGI Federated Cloud, PRACE, HPC, EUDAT, etc. It will be possible to process large amounts of data and to exploit the efficient storage and preservation technologies and infrastructure already available in the European e-infrastructure, with the appropriate mechanisms to ensure security and privacy. Implementation of an advanced PaaS Layer providing the following





Security and Authorization Will focus on authorization issues in distributed, multi-tenant cloud infrastructures, leveraging and integrating with standard and already available authentication technologies (e.g., X.509, SAML2, **OpenID-Connect**). user identities, enrollment and attribute management (e.g., group) membership management, role assignment); access policy definition, distribution and composition; controlled privilege delegation; credential translation.

INDIGO-DataCloud RIA-653549

WP5: PaaS Platform Development





High-level geographical application/service deployment will implement a solution, Geo-deployment Service, to deploy in a transparent and powerful way both services and applications in a distributed and heterogeneous environment made by several different infrastructures (EGI Grid, EGI Fed Cloud, laaS Cloud, Helix Nebula, PRACE, local HPC clusters, etc). well beyond the simple scheduling of the application over grid or cloud environments because it will provide the capabilities to deploy a wide range of diverse applications and services with a powerful set of APIs that hide the complexity of the underlying infrastructures

INDIGO-DataCloud RIA-653549

WP5: PaaS Platform Development



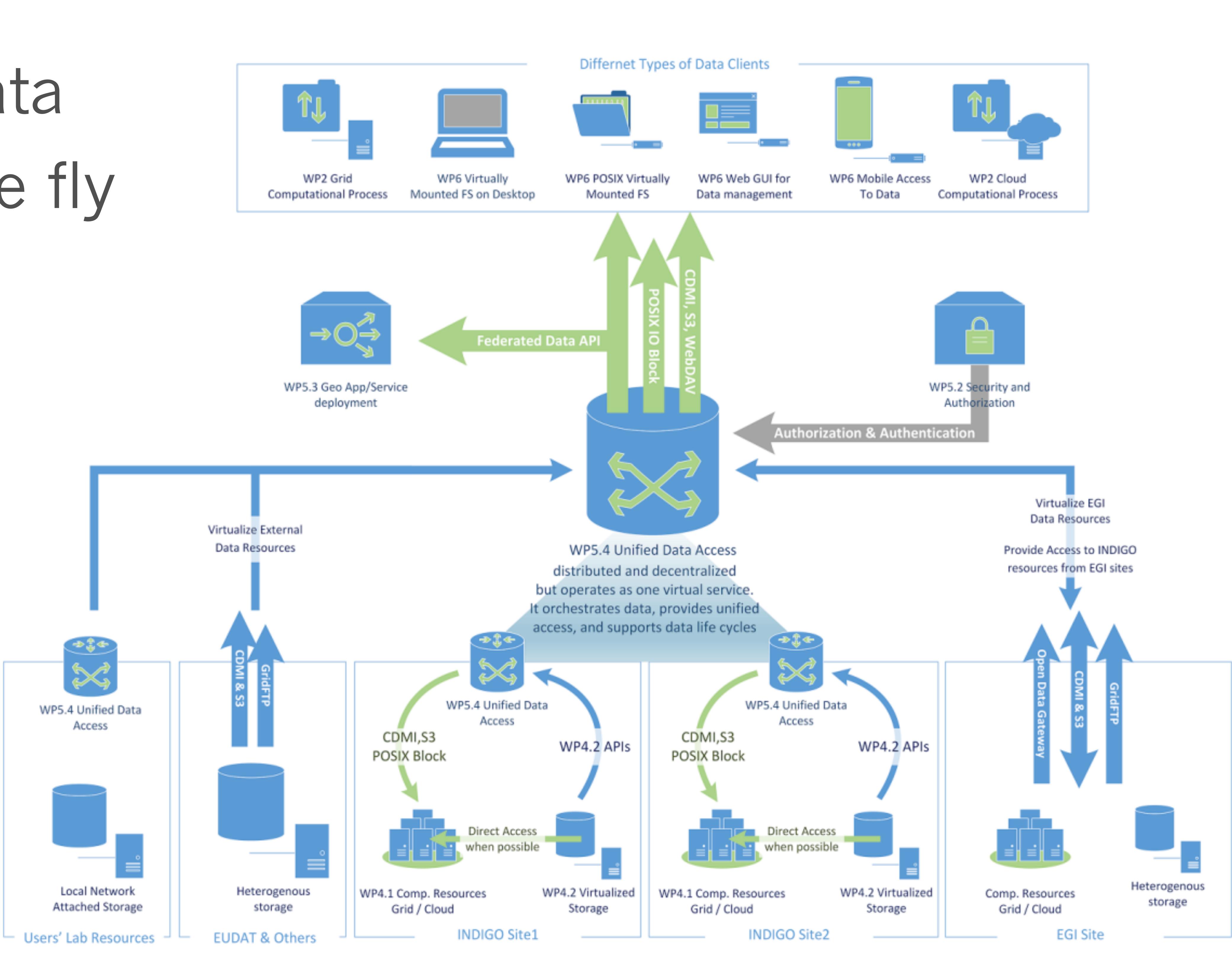
Unified Data Access Unification Federated data access Interoperability and Open Data Optimization and Data on the fly

INDIGO-DataCloud RIA-653549

WP5: PaaS Platform Development

WP5.4 Unified Data Access

Local Network Attached Storage







Develop Toolkits (libraries) that will allow the platform usage from the level of the Scientific Gateways, desktop and mobile applications

Provide and develop the Open Source Mobile Application **Toolkit** for the iOS, Android and Windows Phone platform that will be the base for development of the Mobile Apps. Provide User Friendly front end's, that will prove the usability

of the PaaS proposed:

INDIGO-DataCloud RIA-653549

WP6: Science Gateways, Workflows and Toolkits

Provide both a general-purpose multi-domain Science Gateway and customized examples for selected user communities/scenarios, that will make use of the proposed Toolkits, including **Data Analytics Gateways for e-Science**;



Develop example cross platform native Mobile Apps for selected use cases, based on the Mobile App Toolkit; Manage the execution of complex workflows using PaaS layers, Support for both interactive and batch parallel data analytics workflows.

Provide the dynamic scientific workflows services in a Workflows as a Service model.

multidimensional data).

INDIGO-DataCloud RIA-653549

WP6: Science Gateways, Workflows and Toolkits

Provide workflow interfaces extensions for distributed and parallel data analytics on large volume of scientific,





INDIGO aims to fill important gaps in the field of cloud computing for e-Science Enabling resource provider to improve the efficiency of their cloud infrastructures Enabling users to exploit available infrastructures in an easier and efficient way INDIGO aims to develop general purpose solutions That could be used not only by the user-communities already joining the project, but more widely from scientific communities that has similar needs INDIGO aims to build a sustainable PaaS level cloud solution for escience based on widely used and supported technologies

INDIGO-DataCloud RIA-653549

Conclusions