INFN Cloud Intro

Barbara Martelli, Giacinto Donvito

DICE Datathlon

INFN Cloud, https://www.cloud.infn.it/

- In **production** since March 2021.
- The **seed** of a National Datalake for research and beyond, building on existing, renewed or new e-Infrastructures.
- The **base of the evolution** of the INFN Distributed Computing vision.
- Built on a **thin middleware layer** running on top of *federated clouds*, decoupling physical and logical views via a **service composition** mechanism.
- The **INFN foundation** of all the NRRP computing-related initiatives.
- A **multi-site, federated Cloud** infrastructure integrating HPC and HTC resources.

INFN Cloud

ARCHITECTURAL FOUNDATIONS

NO VENDOR LOCK-IN

Open-source, vendor-neutral architecture

FEDERATION

of existing Cloud infrastructures for both compute and data

DYNAMIC ORCHESTRATION

of resources via the INDIGO PaaS Orchestrator

CONSISTENT AUTHN/AUTHZ

at all cloud levels via OpenID-Connect/OAuth2

EPIC Cloud

Enhanced PrIvacy and Compliance Cloud is an ISO certified cloud platform	A region of INFN Cloud with a certified Information Security Management System
EPIC Cloud offers an IaaS Community Cloud for the communities of	Biomedical and genomic researchers Industrial researchers

Site locations: Bologna (active now), Bari and Catania sites will be added in October enabling for high availability and disaster recovery

Resource available today: about 1 PB storage, 2k CPUs, 16 TB RAM

Ongoing expansion with 3M euro of NRRP resources and 4M euro of funds from other projects

Motivation

- The GDPR states that Clinical and medical data (for instance, genomic) is personal data; i.e., it fits in the Art.9 special categories of personal data.
  - Genomic data is mostly impossible to be anonymized → GDPR shall always be applied
  - ISO/IEC 27001 is the main certification mechanism compliant with GDPR requirements (Art. 43, 58, 63)
- In order to comply with the requirements of health research projects INFN is involved in, we created **a region of the INFN Cloud infrastructure**, applied specific organizational and technical security measures, and certified it ISO/IEC 27001, 27017, 27018.
  - This is the **EPIC Cloud**: a reference Cloud implementation for the treatment of sensitive data at INFN.

From the Data Controller side, the fact that EPIC Cloud is ISO-certified is a way to demonstrate that processing is performed in accordance with the GDPR.

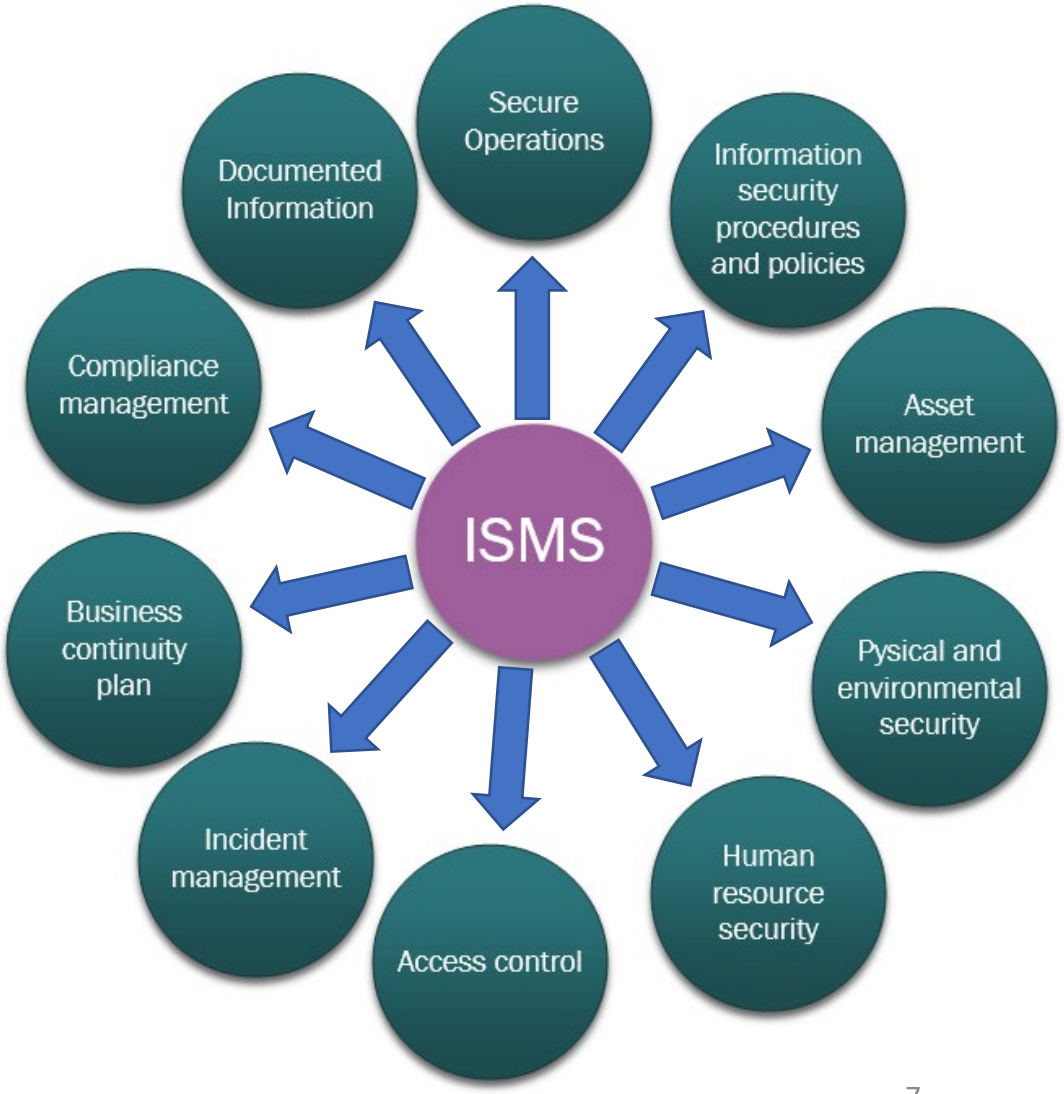
Information Security Management System (ISMS)

- Information Security Management is about preserving the Confidentiality, Integrity and Availability (CIA) of information and associated information facilities (systems, services, infrastructure or physical locations)
- It ensures business continuity by minimizing business damage by preventing and reducing the impact of security incidents
- Other properties can also be involved, such as authenticity, accountability, non-repudiation, reliability and FAIRness
- The objectives of the ISMS are NOT fixed, they depend on the context and are defined by the organization

ISMS: What's all about

Information Security Management System

- It is an organizational framework linking all the elements relevant to the information security, in order to assure that policies, processes and security objectives are implemented, communicated and assessed.
- It needs to continually improve -> **Deming Cycle**
- It is centered to the risk assessment process -> all decisions are based on the output of this process
- Goal: achieving the optimal CIA **balance**, i.e., ensuring Confidentiality of information, while still ensuring the information remains accessible to authorized persons and is not altered



Deming Cycle and Risk Assessment

Plan
(context, leadership, planning & support)

Establish the organization

- Define the scope
- Analyze the context and define the security requiremnts and objectives
- Define the Security Policy
- Determine roles and responsibilities
- Define the risk assessment approach

Do
(risk assessment & operations)

Risk Assessment

Control Selection

Control Implementation (Operations)

Check
(performance evaluation)

Monitoring KPI

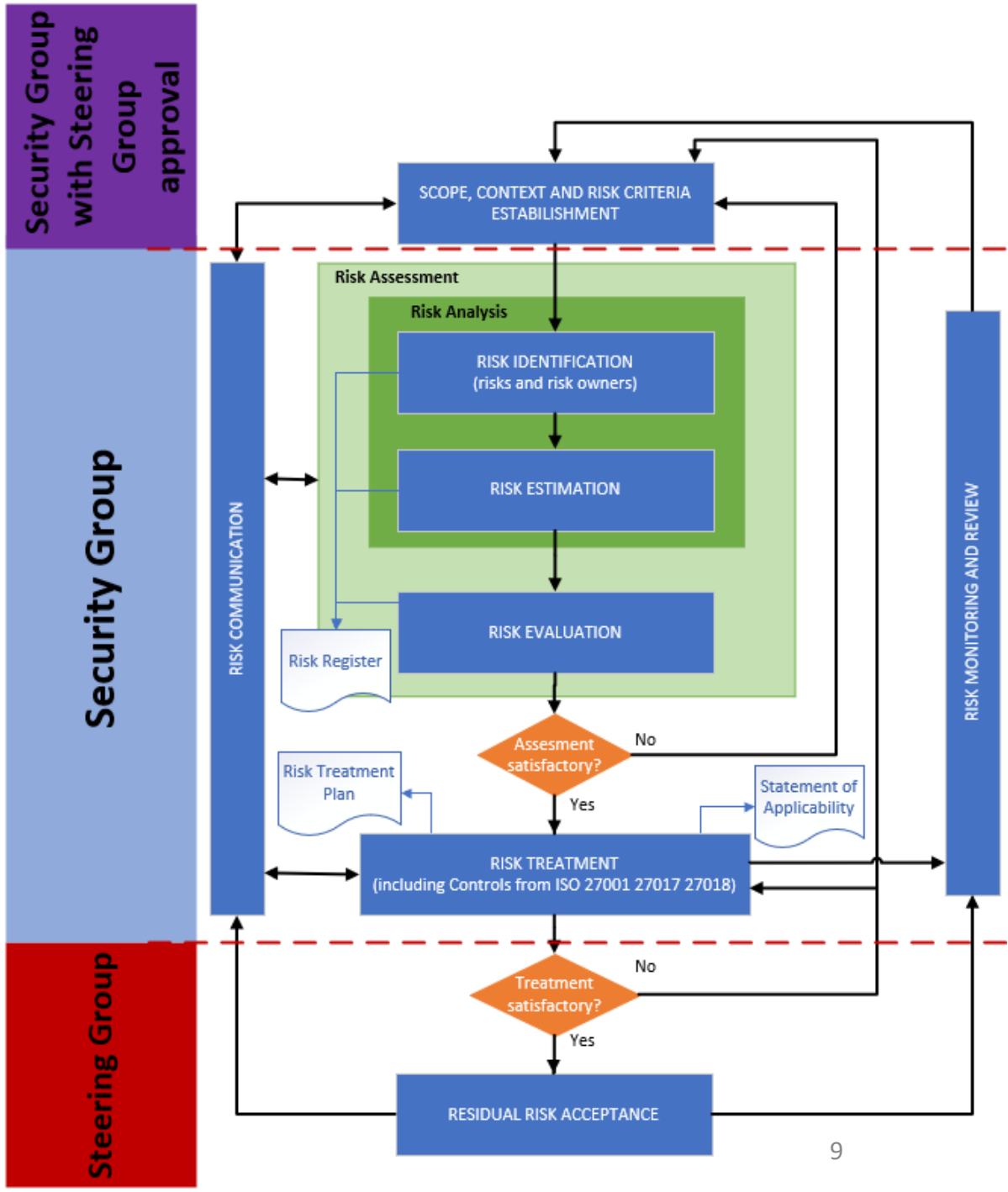
Internal Audit
External Audit

Act
(continual improvement)

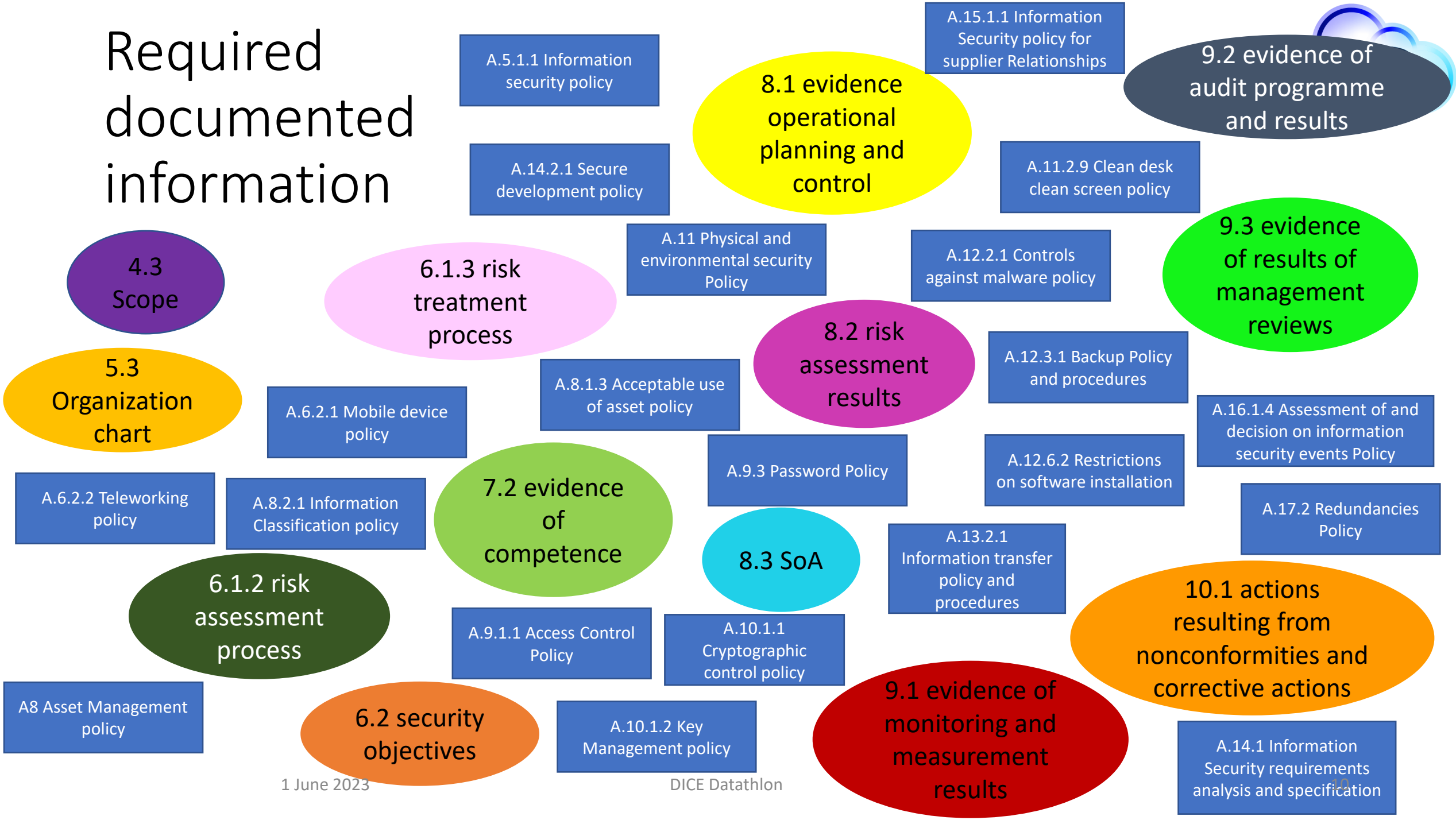
Effectiveness

Risk Management Process in EPIC

- Inspired by:
  - ISO 27005 guideline with modifications
    - we don't start with asset identification, but we use a scenario-based risk assessment
  - and ISO 31000
- Iterative process aimed at supporting the decision-making process
- In EPIC is performed once a month and whenever a relevant change in the system occurs
  - ISO 27001 clause 8.2 requires to perform it *"at planned intervals or when significant changes are proposed or occur, taking account of the criteria established in 6.1.2 a)."*
- Established criteria to define Risk Owners



Required documented information



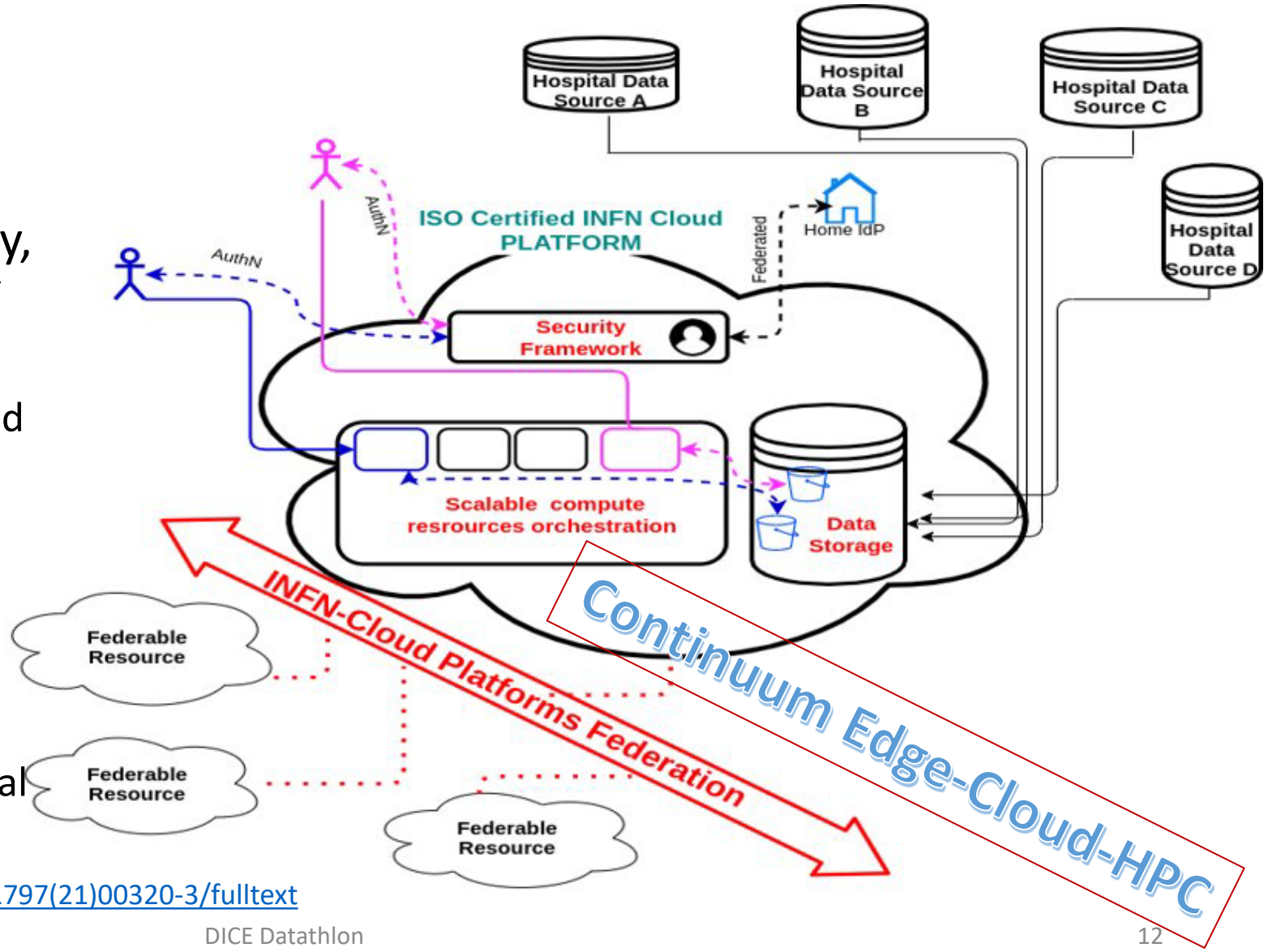
Technology

It is based on the same technologies of INFN Cloud (OpenStack, K8S, Indigo IAM, OneData, Laniakea), with various enhancements introduced to meet higher security and privacy standards. For example:

- The IAM provides 2FA, integration with web services, SSH and VPN (OpenVPN)
- Enhanced ONEDATA with more auditing functionalities
- Network segregation between OpenStack tenants is guarantee by ACLs
- At-rest and in-transit encryption
- Standard shared responsibility model:
  - User manages data, applications, runtime, middleware and OS
  - EPIC manages networking, storage, servers, virtualization
- Advanced logging and auditing services
  - centralized syslog managed applying the ***segregation of duties*** principle

EPIC Cloud enables the federated datalake for health-related use cases

- In general, we receive heterogeneous requirements form life-science communities: someone need central repository, others need to store data locally
- Possible scenarios:
  1. Central harvesting of data collected remotely
  2. Edge-level anonymization and central ingestion and analysis of data
  3. Edge-level Feature Extraction and central ingestion and analysis of features
  4. Federated learning (training at local sites and algorithm publishing)



https://www.physicamedica.com/article/S1120-1797(21)00320-3/fulltext

Thanks!

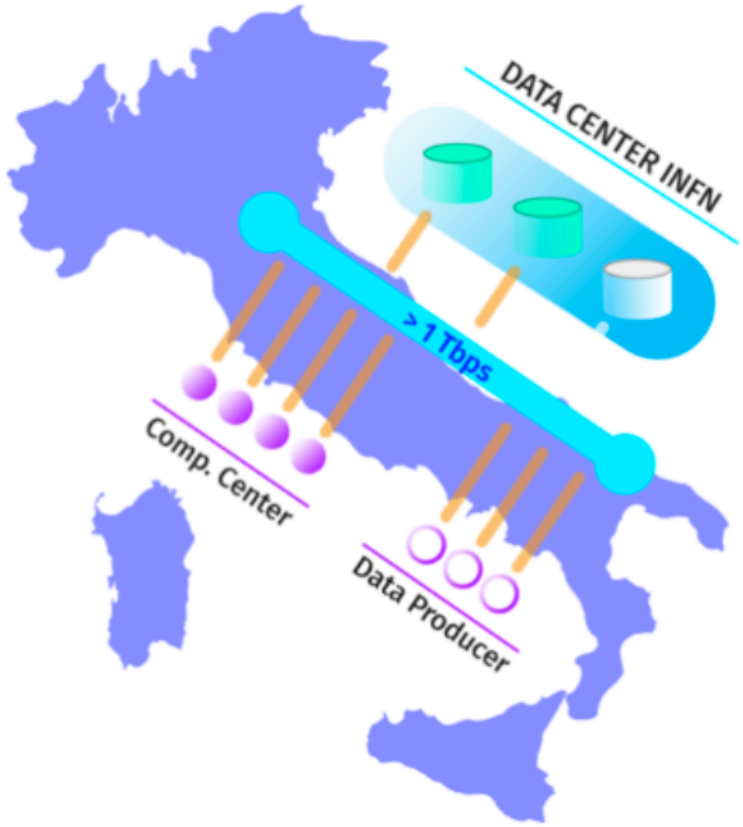
Backup slides

Federation

INFN CLOUD IS DESIGNED AS A FEDERATION OF PRE-EXISTING INFRASTRUCTURES

- The **Backbone** of the INFN Cloud is made up of two closely linked federated sites, BARI and CNAF.
- A scalable set of **satellite sites**, geographically distributed across Italy and loosely coupled, expand the resources offered by the backbone.

INFN Cloud core services and some centralised, fully managed, high-level services are hosted on the Backbone. This allows us to leverage **high-availability** and **disaster recovery** capabilities to ensure that these critical services are always available and operating at peak efficiency.

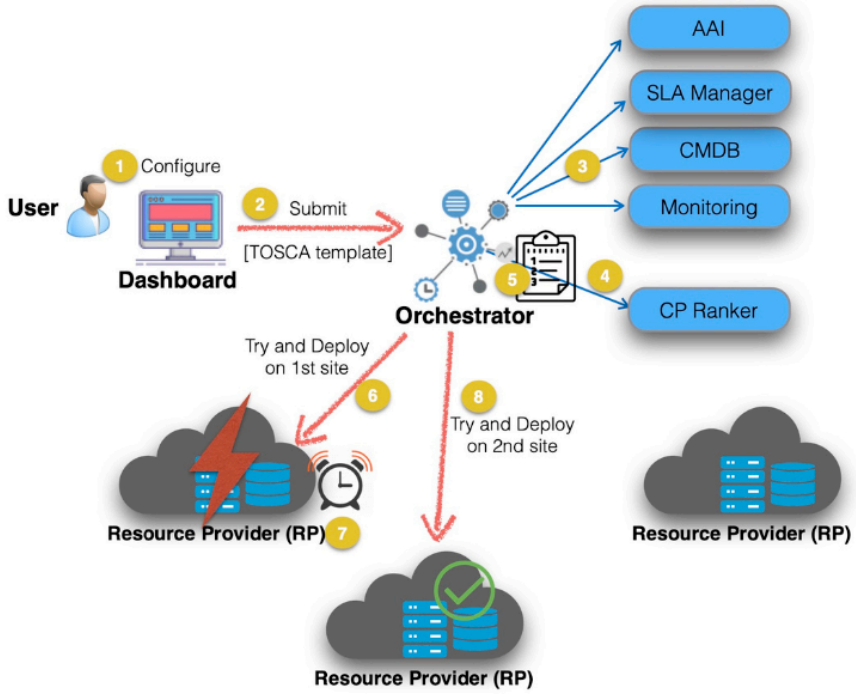


Middleware

THE FEDERATION MIDDLEWARE

The INDIGO PaaS Orchestrator enables the federation of distributed and heterogeneous compute environments: clouds, docker orchestration platforms, HPC systems.

- **Smart scheduling** → Automatic selection of the best provider
  - based on compute/storage requirements vs provider capabilities including the following criteria:
    - Resource quotas (SLA)
    - Monitoring data
    - Support for specialized hardware (GPU, Infiniband)
    - Data location
- Support for **hybrid deployments** and network orchestration
- **Client interfaces** for advanced users (REST APIs, CLI, python bindings) and end-users (web dashboard - no skills required)

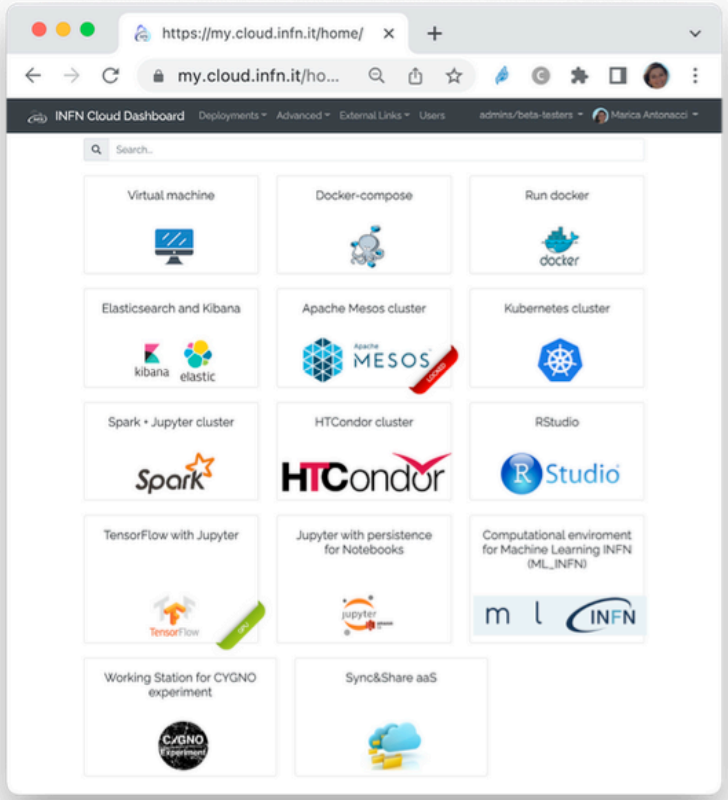


Front-ends

THE PAAS DASHBOARD

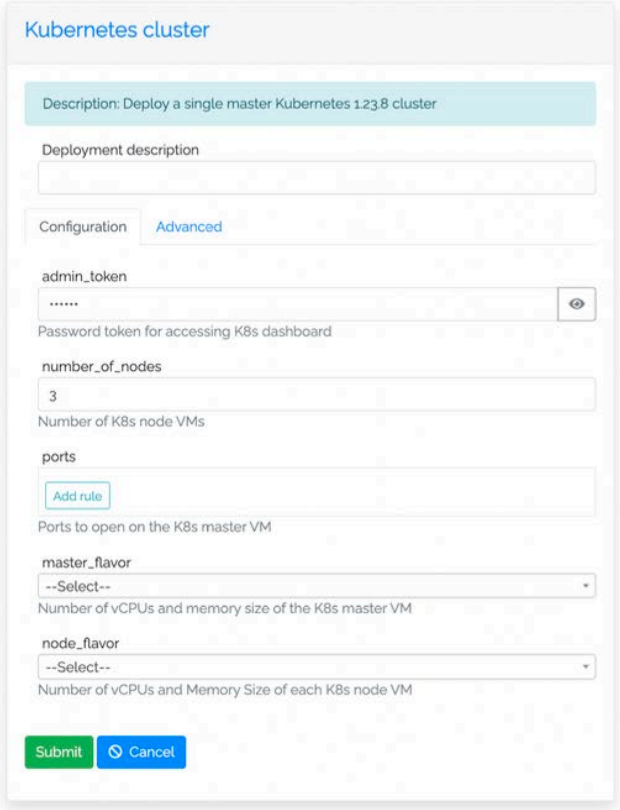
The INDIGO PaaS Dashboard is a web-based user interface that enables users to manage and monitor their deployments without requiring any TOSCA knowledge.
The dashboard hides all technical details and provides an intuitive interface for managing service deployments.

- OpenID-Connect Authentication
- Multi-tenancy
- Secrets management (via Vault integration)
- Dynamic view of service catalog (depending on the user group membership)



Self provisioning

REQUEST SERVICES WITH JUST A FEW CLICKS



Customize your deployment

through the deployment input parameters

Choose the Scheduling strategy

- automatic: let the Orchestrator select the best provider
- manual: choose the provider from the drop down menu automatically created by the Dashboard with the list of providers returned by the SLA Manager service