ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

DESIGN PHASE KICK-OFF EVENT AND AWARD CEREMONY

08 June 2020

Contact: info@archiver-project.eu
Project website: www.archiver-project.eu

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

Event Outline

[2.00 pm - 2.10 pm] Welcome from Port d'Informació Científica (PIC), Barcelona - Design Phase leader

[2.10 pm - 2.20 pm] Project overview - João Fernandes (CERN)

[2.20 pm - 2.50 pm] Use cases overview - Buyers Group representatives (CERN, DESY, EMBL-EBI,PIC)

[2.50 pm - 3.00 pm] Break

Award ceremony:

[3.00 pm - 3.15 pm] Presentation from Arkivum - Google

[3.15 pm - 3.30 pm] Presentation from GMV – PIQL – AWS – SafeSpring

[3.30 pm - 3.45 pm] Presentation from Libnova – CSIC – University of Barcelona – Giaretta Associates

[3.45 pm - 4.00 pm] Presentation from RHEA System Spa – DEDAGROUP – GTT

[4.00 pm - 4.15 pm] Presentation from T-Systems International – GWDG – Onedata

[4.15 pm - 4.30 pm] Feedback session & closing remarks - Marion Devouassoux (CERN)

PIC
port d'informació
científica

Helping to turn Information into Knowledge

Welcome

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

Phase 1 Awards Ceremony
June 8th, 2020

Prof. Manuel Delfino

IFAE 25 years 1991-2016 Institut de Física d'Altes Energies

EXCELENCIA SEVERO OCHOA

BIST Barcelona Institute of Science and Technology

Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

EXCELENCIA MARÍA DE MAEZTU

CIEMAT física de partículas

PIC
port d'informació
científica

PIC scientific data centre

Port d'Informació Científica (PIC)
(Scientific Information Harbour in English) is maintained through a collaboration of two leading scientific institutes in Spain

IFAE 25 years 1991-2016 Institut de Física d'Altes Energies
EXCELENCIA SEVERO OCHOA
BIST Barcelona Institute of Science and Technology

Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas
EXCELENCIA MARÍA DE MAEZTU
cfp CIEMAT física de partículas

PIC is located on the campus of one of Spain's leading universities

UAB Universitat Autònoma de Barcelona

Project funding is provided by

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
AGENCIA ESTATAL DE INVESTIGACIÓN

Unión Europea
Fondo Europeo de Desarrollo Regional
"Una manera de hacer Europa"

PIC
port d'informació
científica

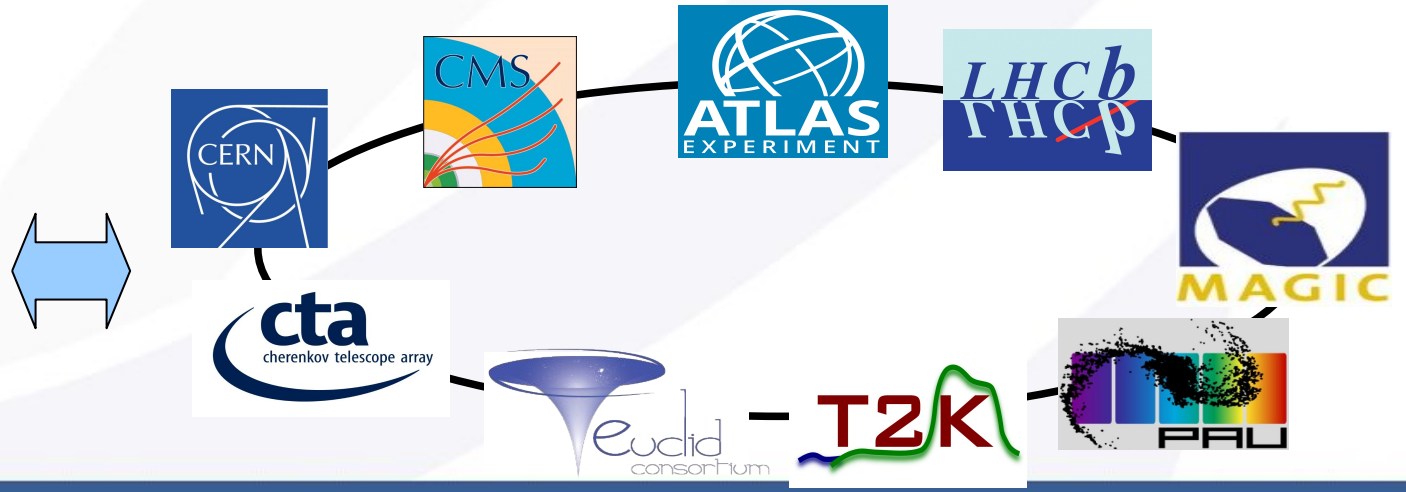
PIC scientific data centre

Port d'Informació Científica (PIC) is the largest scientific data centre in Spain, supporting **research** involving **analysis of massive sets of data.**

It provides data processing and analysis services for international research projects:

- **Spanish WLCG Tier-1 centre for CERN's LHC detectors (ATLAS, CMS and LHCb)** → **~85% of resources**
- **ATLAS Tier-2** and **ATLAS and CMS data analysis facility**
- **Scientific Data Center** for **ESA's EUCLID mission**
- Main data centre for **MAGIC Telescopes and PAU Cosmological Survey**
- Contributing to data processing of ongoing and emerging projects, like **DES and CTA**

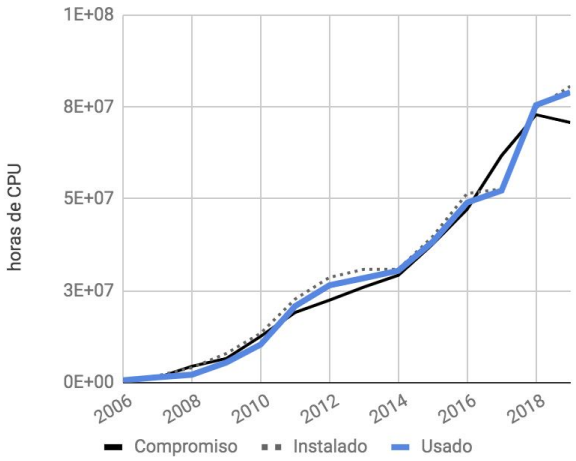
~8500 cores
~10 PB disk
~30 PB magnetic tape



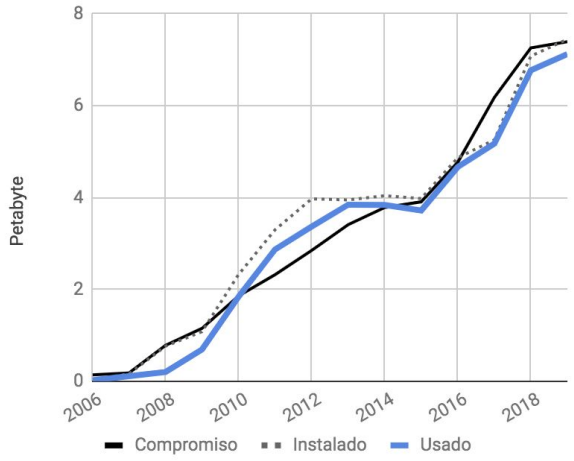
PIC
port d'informació
científica

PIC Tier-1 capacity growth

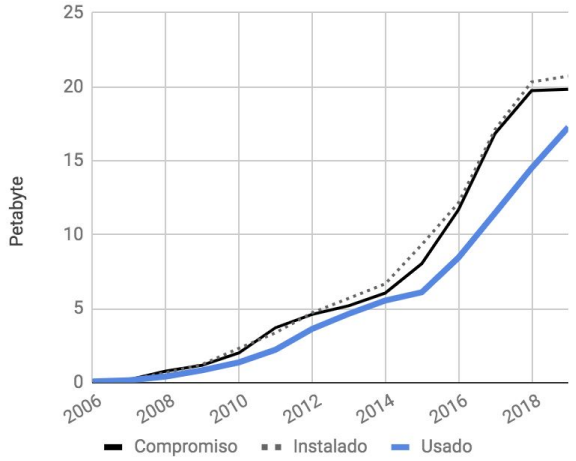
Cómputo



Disco

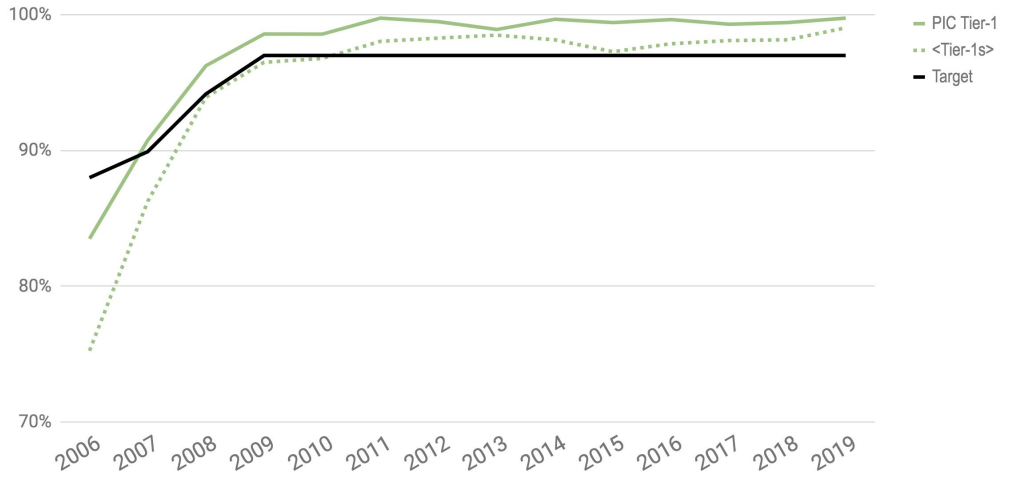


Cinta magnética



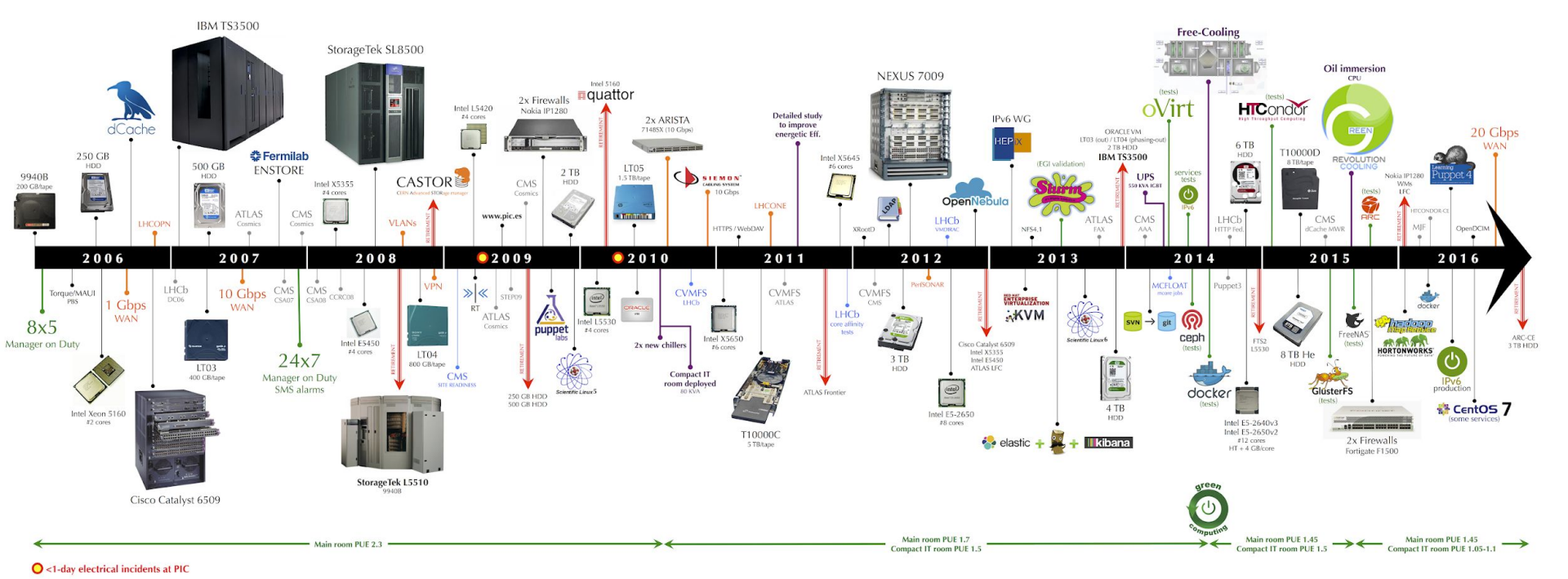
- 37% YOY growth rate stretching over a decade
- Excellent reliability and availability

PIC Tier-1 Reliability



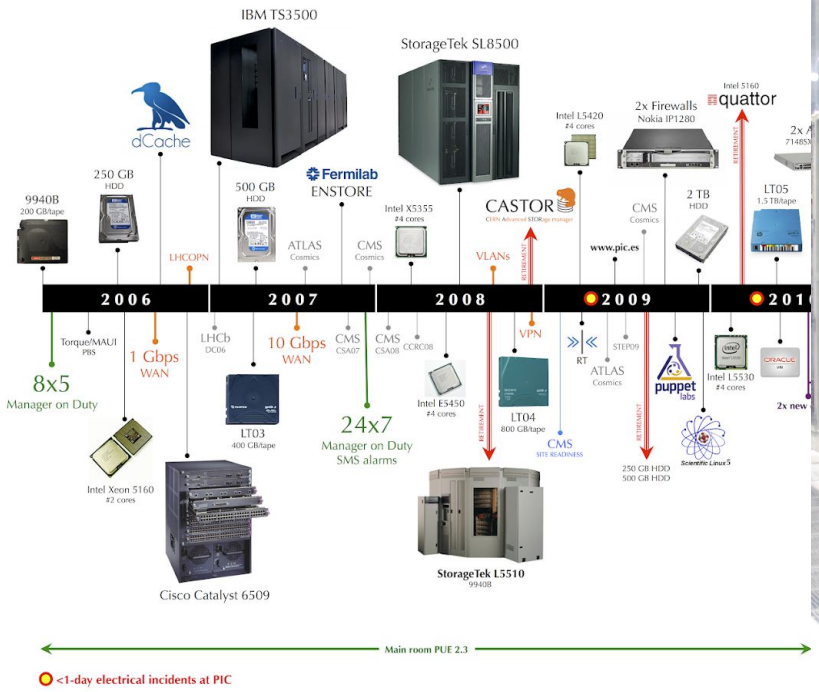
PIC
port d'informació
científica

PIC in constant technological evolution



PIC
port d'informació
científica

PIC in constant technological evolution



Added end 2019:

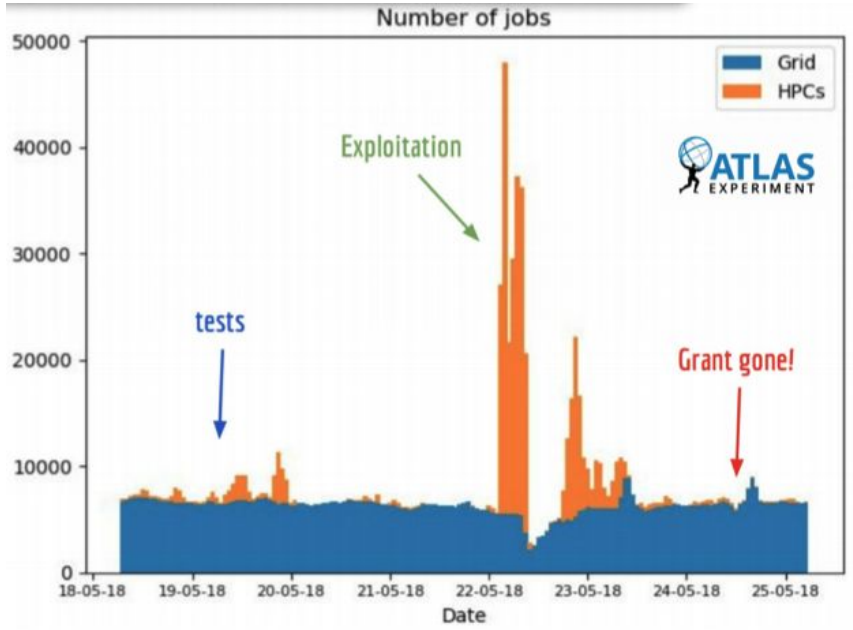
- First module of new IBM Tape Library
- LTO-8 cartridge technology

Unión Europea

Fondo Europeo
de Desarrollo Regional
"Una manera de hacer Europa"

PIC
port d'informació
científica

PIC in constant technological evolution



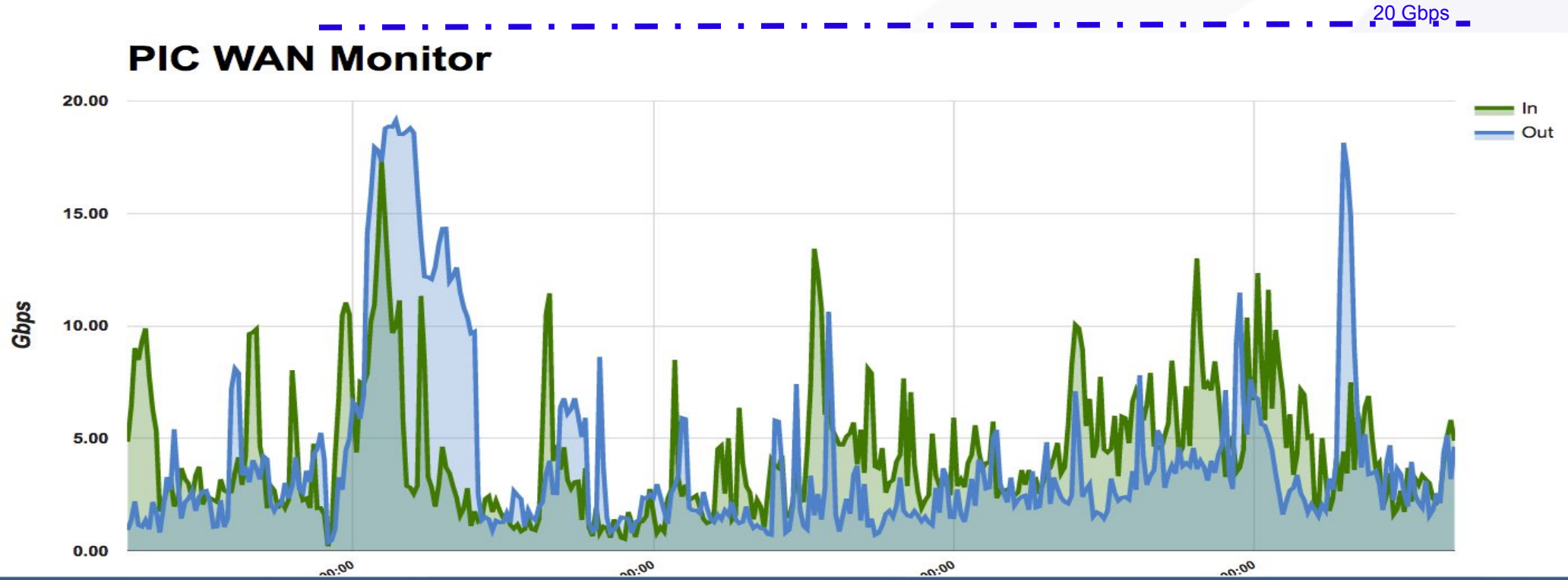
CPU Bursting from PIC out to
Barcelona Supercomputing Center (BSC)

CPU Bursting from PIC out to
AWS instances

PIC
port d'informació
científica

PIC WAN Upgrade

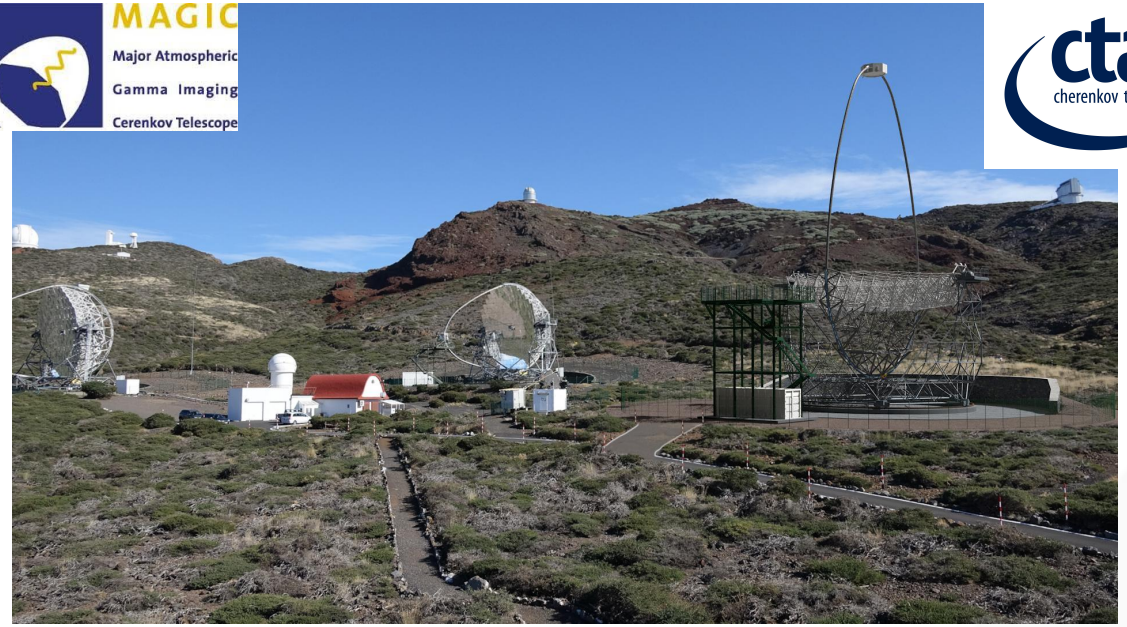
- Network connection upgraded end of 2016
  - from 10 Gbps to 20 Gbps
  - first institution connected to 20 Gbps in Spain
  - preparing deployment of 100 Gbps network upgrade



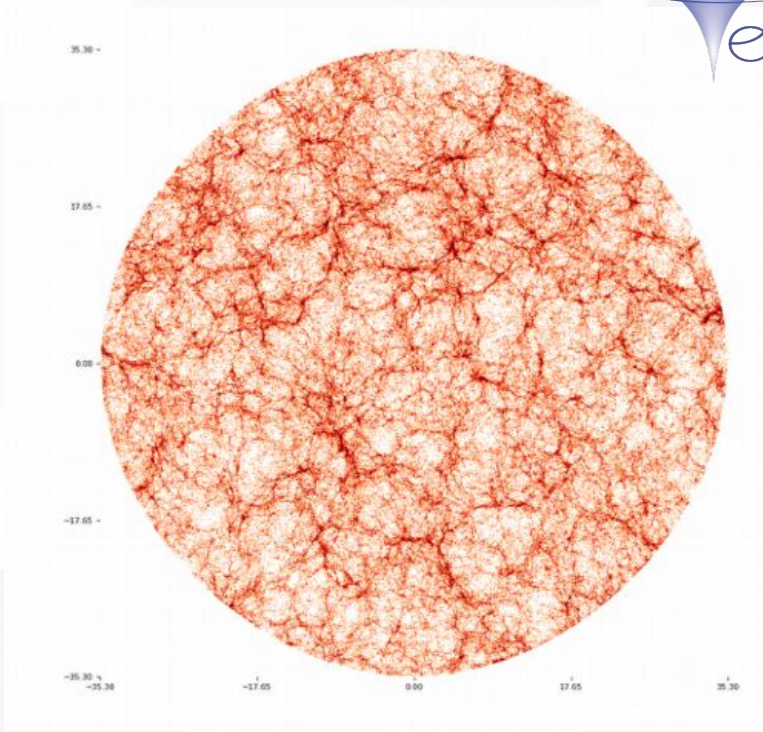
PIC
port d'informació
científica

Astrophysics and Cosmology

MAGIC
Major Atmospheric
Gamma Imaging
Cerenkov Telescope

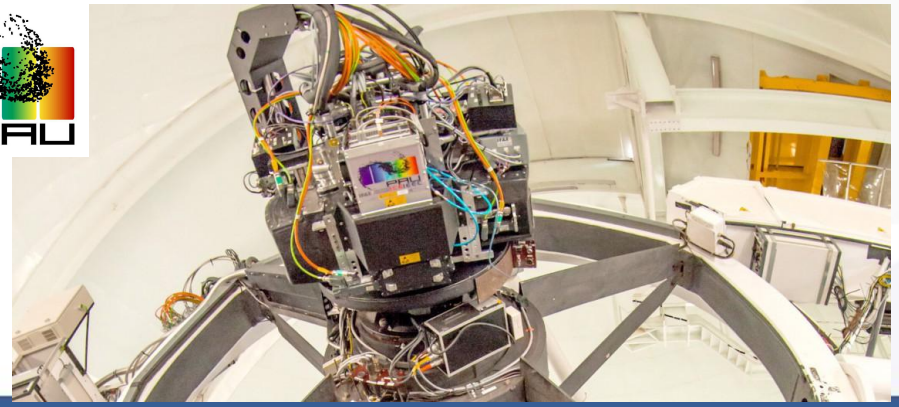


cta
cherenkov telescope array



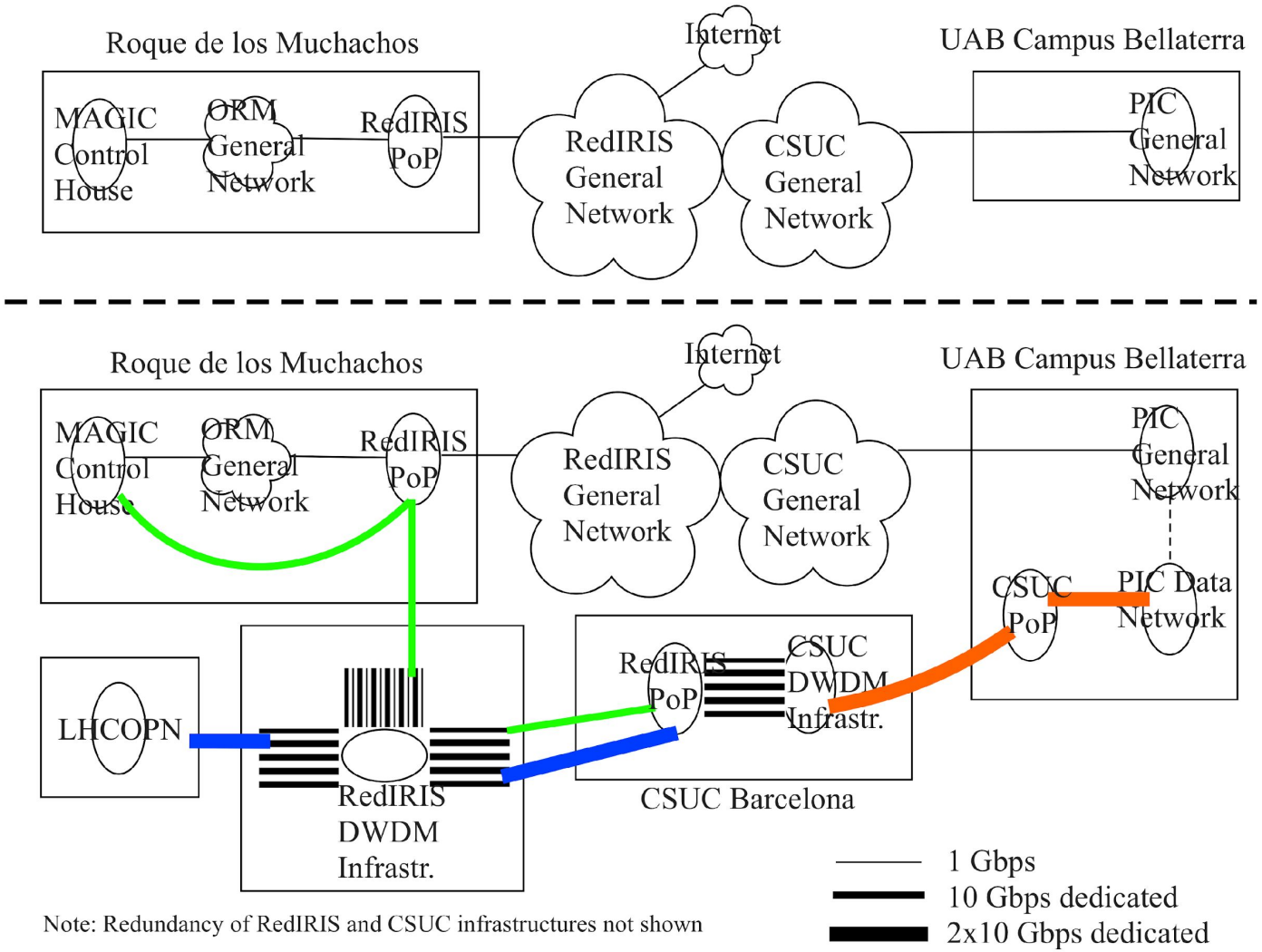
euclid
consortium

PAU



PIC
port d'informació
científica

10 Gbps light path to ORM in La Palma

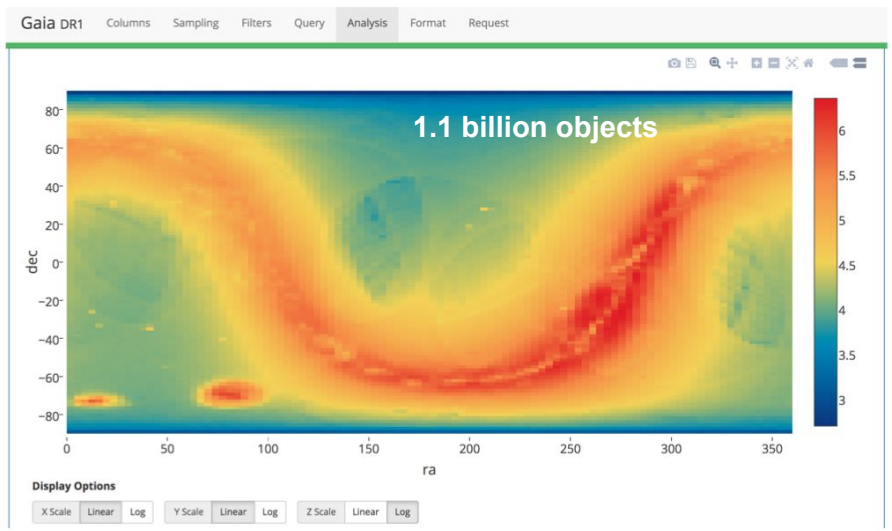


Note: Redundancy of RedIRIS and CSUC infrastructures not shown

PIC
port d'informació
científica

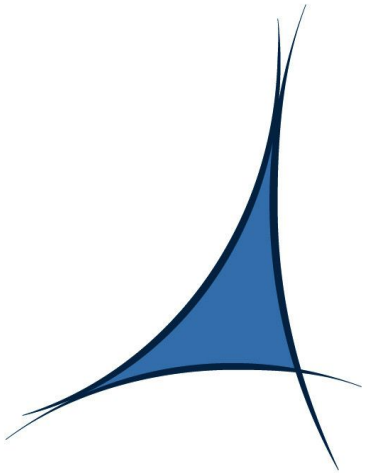
cosmohub.pic.es

- **CosmoHub** on Hadoop: a web portal to analyze and distribute massive cosmological data



- Holds the **largest virtual galaxy catalogue** to date, the Euclid Flagship mock galaxy catalogue, which contains 7.4 billion galaxies covering 1/8 of the sky (full catalog → ~60 billion entries)
- Also holds the input for the Flagship catalogue, a 44 billion dark matter haloes catalogue generated from a 2.3 trillion DM particle simulation by U. Zurich
- Enabling Notebooks over Big Data platform (using Spark/Zeppelin) for users

ARCHIVER
HELIX NEBULA
ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS



PIC
port d'informació científica

Helping to turn Information into Knowledge

IFAE
25 years 1991-2016
Institut de Física d'Altes Energies

EXCELENCIA SEVERO OCHOA

BIST Barcelona Institute of Science and Technology

Ciemat
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

EXCELENCIA MARÍA DE MAEZTU

cfp CIEMAT física de partículas

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

ARCHIVER

Archiving and Preservation for Research Environments

João Fernandes (CERN)
ARCHIVER Project Coordinator

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

Project Objective

Focus: Archiving and Data Preservation Services using commercial cloud services via the European Open Science Cloud (EOSC)

Procurement R&D budget: 3.4M euro; Total Budget: 4.8M

Starting Date: 1st of January 2019

Duration: 36 Months

Coordinator: CERN (Lead Procurer)

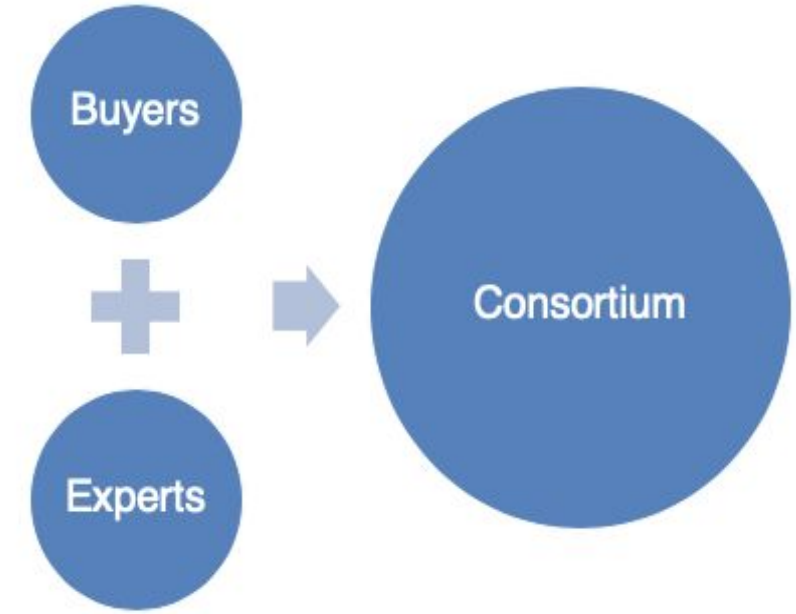
ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

European Commission

Consortium

Includes Buyers and Experts in the preparation, execution and promotion of the procurement of R&D

The “Buyers Group”: Public organisations committing funds to contribute to a joint-R&D-procurement, research data use cases and R&D testing effort



Experts – Partner organisations bringing expertise in requirement assessment and promotion activities, not part of the Buyers Group

European Open Science Cloud

Role of the EOSC:

Data-driven: for 1.7 million European researchers and 70 million professionals in science and technology

Federated virtual environment, free at the point of use for the end researcher

Open services for storage, analysis and re-use of research data

Approach across national borders & scientific disciplines

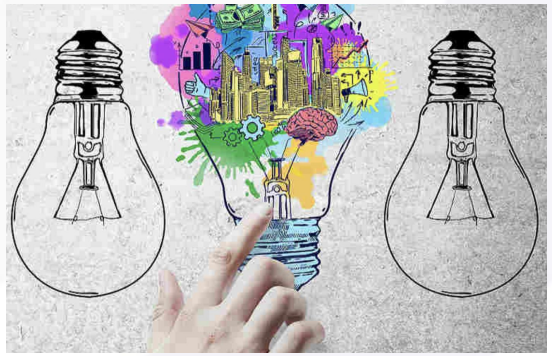
Promote choice of services & deployment models: on-prem, hybrid, off-premise

EOSC legal entity expected by the end of 2020

slide from Rupert Lück (EOSC Sustainability Working Group co-chair, EMBL)

EOSC: Role of ARCHIVER

- Co-create a set of sustainable digital repositories for research
- Foster innovation
- Promote choice 😊
- Stay mainstream by adopting widely used and recognised standards



Early Adopters

https://archiver-project.eu/early-adopters-programme

- **Participants:**
  - Demand side public sector organisations
- **Key advantages**
  - Access and assess if resulting services address archiving and preservation meet their needs
  - Contribute and shape the R&D carried out in the project, contribute with use cases and
  - Have the option to purchase pilot-scale services by the end of the project
- **Confirmed 11 organisations, more are in the process:** → ***High level of interest from the community***

CRG Centre for Genomic Regulation

IEO Istituto Europeo di Oncologia

SURF SARA

aarnet Australia's Academic and Research Network

SWITCH Serving Swiss Universities

KAUST

Universitet Stockholms

FMI Friedrich Miescher Institute for Biomedical Research

UKRI Science and Technology Facilities Council

ARDC Australian Research Data Commons

Jisc

Move from current state of the art

Current Scientific Data Repositories

- ***Growing data volumes***
- ***Basic bit preservation capabilities***
- ***Concerns: technology lock-in (tape), Disaster Recovery/Business Continuity plans needed (COVID-19)***
- ***Most of research data not published***
- ***Fragmentation across scientific disciplines & countries***
- ***Cost underestimation at the planning phase***

- ***PB scale demonstration of scientific data repositories***
- ***Profit from considerable experience of European SMEs preservation experts***
- ***Promote FOSS, open standards & concretely test exit strategies***
- ***Best practices: FAIR, TRUST, DPC(RAM)***
- ***Pan-European: resulting services available in the EOSC***
- ***Cost model adapted to public research***

ARCHIVER "current state of the art" report: https://doi.org/10.5281/zenodo.3618215

R&D Scope

Demand Side Requirements

Layer 4
Advanced services

High level services: visual representation of data (domain specific), reproducibility of scientific analyses, etc.

Layer 3
Baseline user services

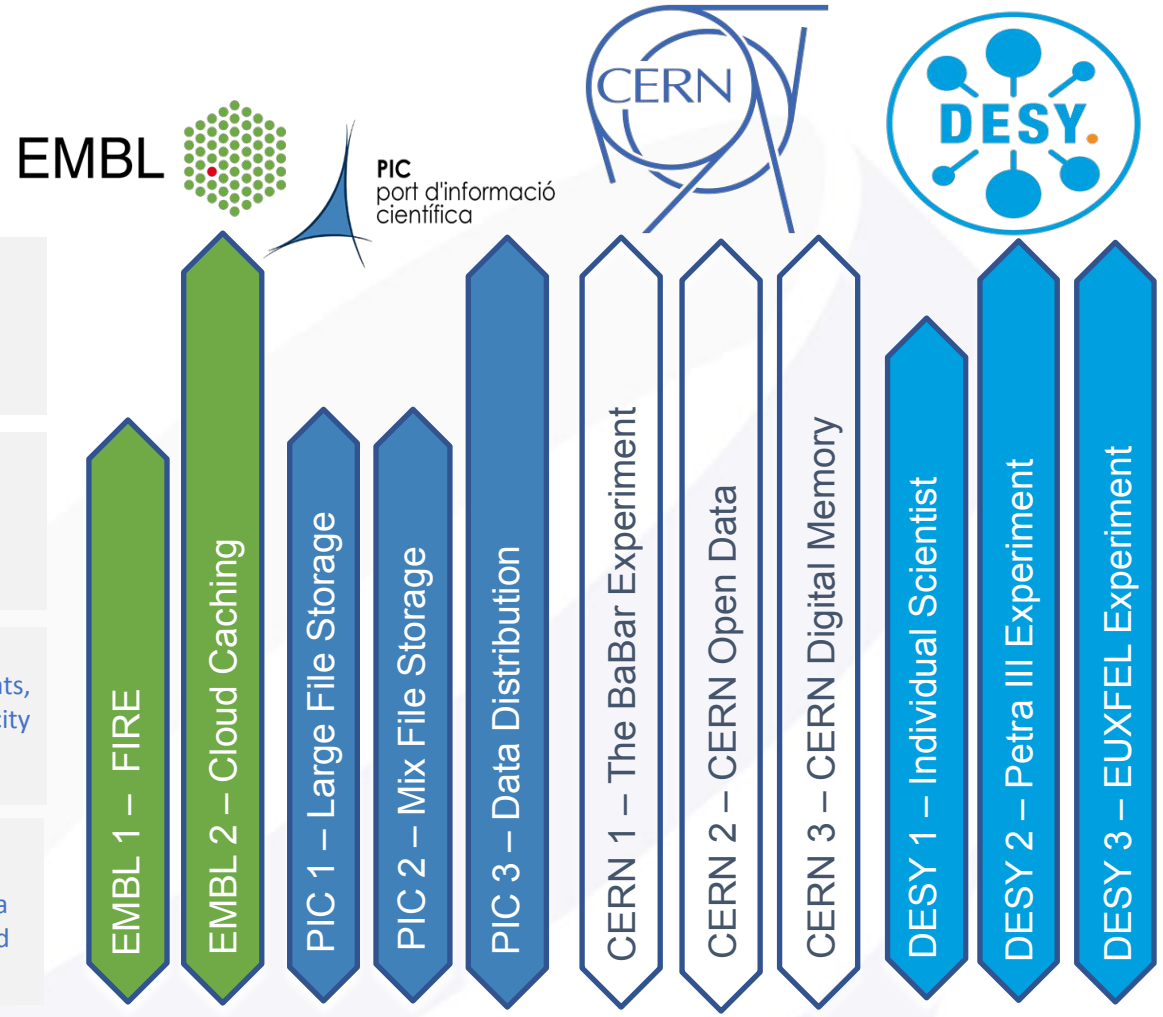
User services: search, discover, share, indexing, data removal, etc.
Access under Federated IAM

Layer 2
Preservation

OAIS conformant services: data readability formats, normalization, obsolesce monitoring, files fixity, authenticity checks, etc.
ISO 14721/16363, 26324 and related standards

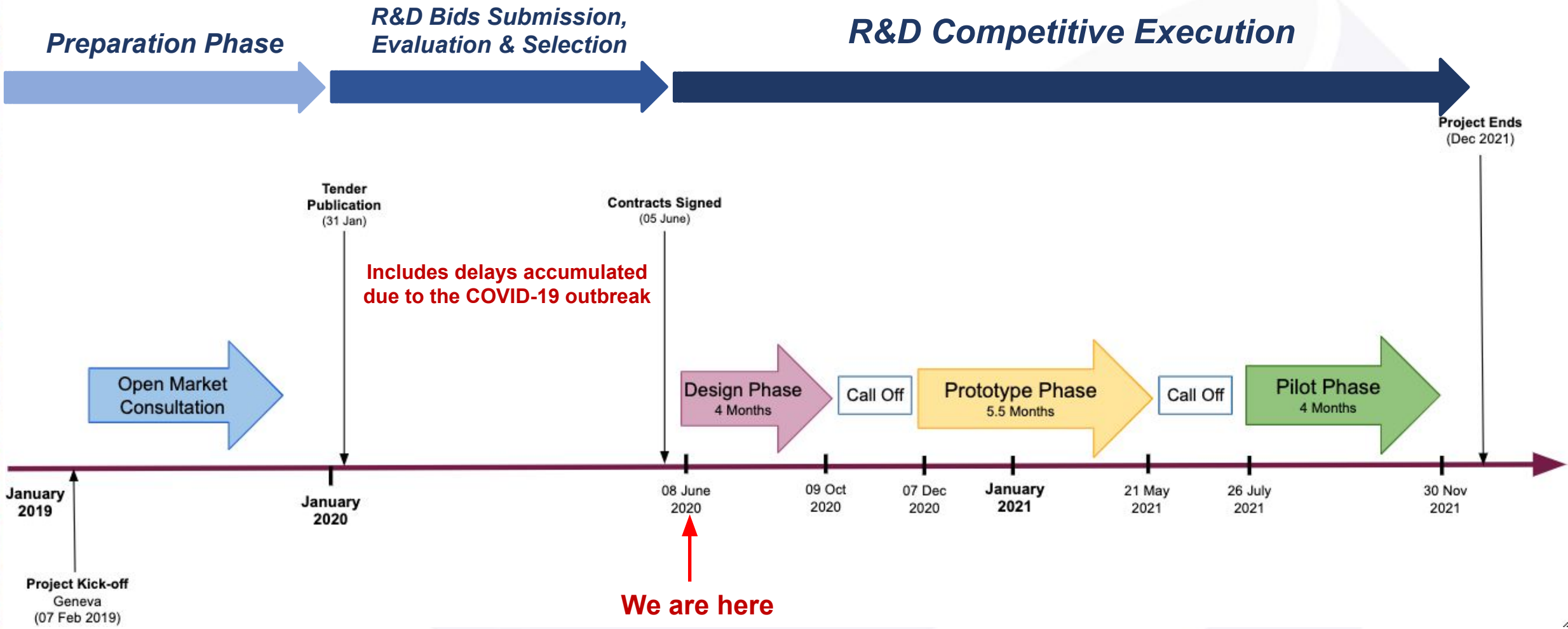
Layer 1
Storage/Basic Archiving/Secure backup

Data integrity/security; cloud/hybrid deployment
Data volume in the PB range; high, sustained ingest data rates. ISO certification: 27000, 27040, 19086 and related standards. Archives connected to the GEANT network



Scientific use cases deployments documented at: https://www.archiver-project.eu/deployment-scenarios

Project Timeline



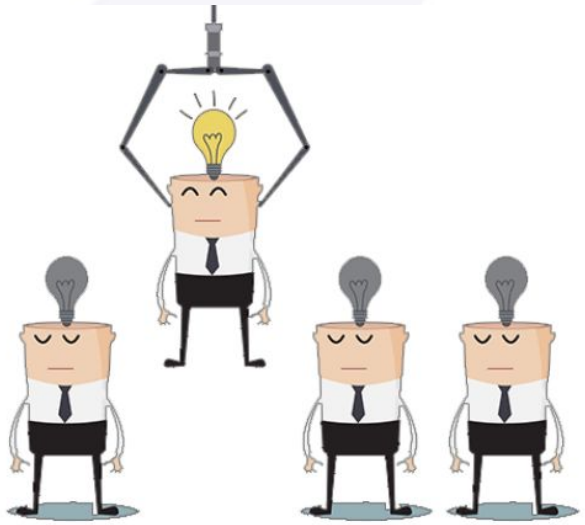
R&D bid submission in numbers

- Information sessions: average of 80 participants

- Downloads of the PCP RfT before closure of submission period:
  - # Downloads: 147
  - # of different organisations / companies: 122
  - # of countries represented: 29

- # R&D bids received: 15
- # of organisations and companies involved: 43

Number of selected consortia: 5



Thank you! 

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

CERN Use Cases Overview

Jakub Urban (CERN)

Tibor Simko (CERN)

Jean-Yves Le Meur (CERN)

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

CERN Use Case - THE BABAR EXPERIMENT

During 2020, the BaBar Experiment infrastructure at Stanford Linear Accelerator (SLAC) will be decommissioned. 2 PB of BaBar data can no longer be stored at the host laboratory.

Currently, a copy of the data is being held by CERN IT (Storage Group).

Objectives:

- To store the second copy of BaBar outside SLAC
- Make the data available for possible comparisons with data from other experiments

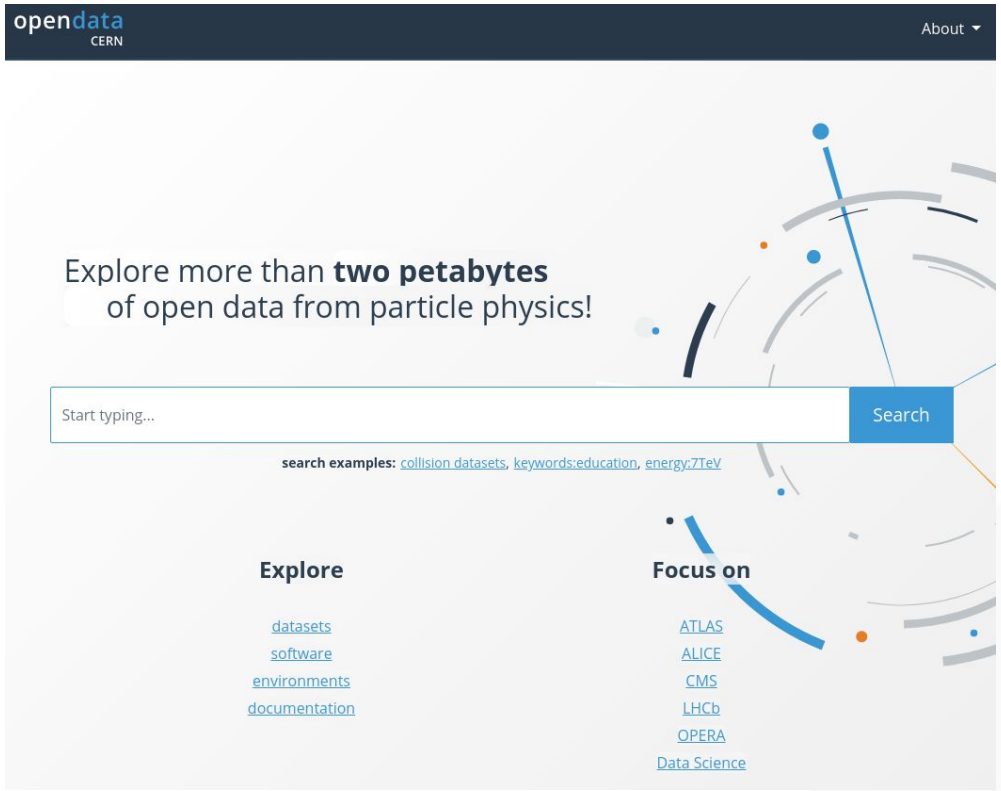
https://www.archiver-project.eu/deployment-scenarios-technical-summaries/babar-experiment

CERN Use Case - THE BABAR EXPERIMENT

- Access control via Federated Authentication
- PB volume; ingestion and recall speeds ~ 10 Gbps
- REST API services for data ingestion and recall
- Web access via a dashboard
- File recalls within a few hours, guaranteed bit preservation
- Provide functionality for data reusability and research reproducibility
- Cost model: over long periods (~ 5 years), estimated 50K€ per PB per year

https://www.archiver-project.eu/deployment-scenarios-technical-summaries/babar-experiment

CERN Use Case - CERN Open Data



Goal: independent preservation

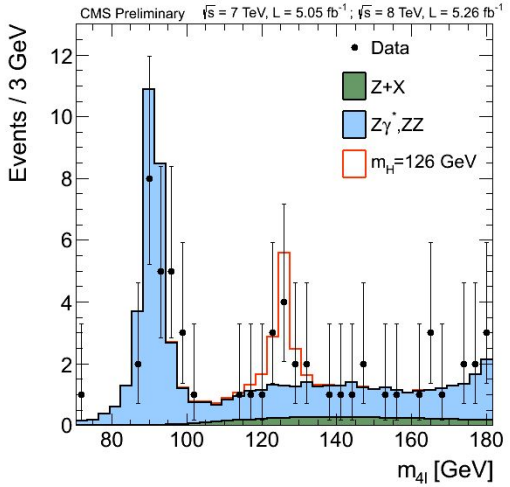
- O(2PB) of data described via JSON Schema
- typical dataset: O(10TB) size, O(3K) files
- 100% open content, easy to push/pull

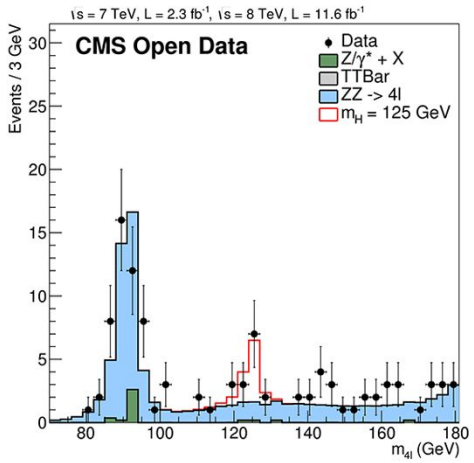
Example scenarios

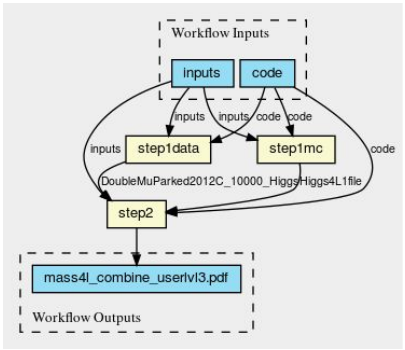
- ingest O(500TB) per month
- recall fast one particular file from a preserved dataset for disaster recovery
- offer public HTTP/XRootD access to preserved content

https://www.archiver-project.eu/deployment-scenarios-technical-summaries/cern-open-data

CERN Use Case - CERN Open Data







CernVM File system

docker

Goal: independent reproducibility

- run selected open data analysis examples
- use Virtual Machines or Docker containers
- offer “compute” to complement “storage”

Example scenarios

- instantiate CVMFS service independently of CERN computing infrastructure
- instantiate condition database during analysis runtime
- run open data analysis workflows

https://www.archiver-project.eu/deployment-scenarios-technical-summaries/cern-open-data

CERN Use Case - Digital Memory

"CERN is not just another laboratory. It is an institution that has been entrusted with a noble mission which it must fulfil not just for tomorrow but for the eternal history of human thought."

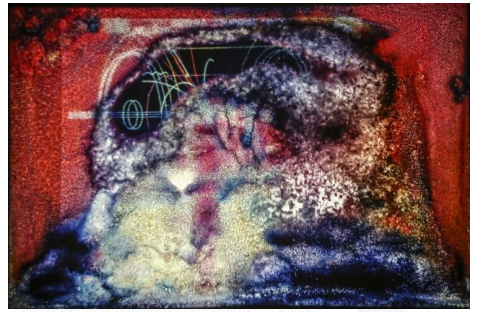
Albert Picot, 3rd Session of CERN Council, Geneva, 10 June 1955

Deployment consisting on a requirement to archive approximately 1.5 PB of digital Memory, containing analogue documents produced in the 20th century as part of the Organization patrimony, as well as digital production of the 21st century (web sites, social media, selected emails, etc.)

Goal : Produce a dark archive in the cloud following standard OAIS practices.

https://www.archiver-project.eu/deployment-scenarios-technical-summaries/cern-digital-memory

CERN Use Case - Digital Memory



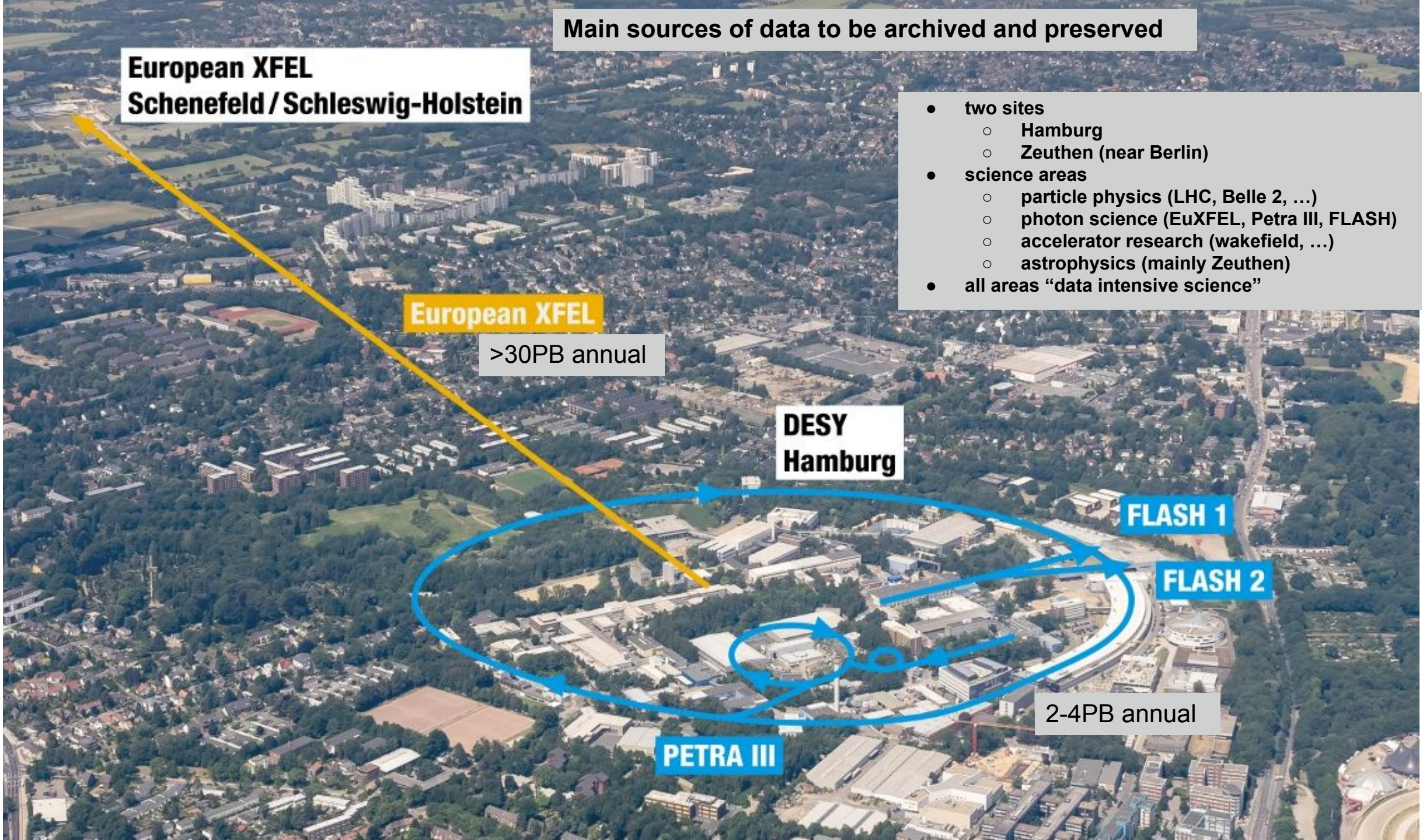
Badly preserved slide revealed as a piece of art

- More than 100 films, 6’000 videos tapes and 450’000 photos already digitized in high-res versions for preservation
- ISO 16363 compliance: create and store Archival Information Packages for the very long term → “AIP Factory”
- Feeding the Archival system from CERN Information Systems (many based on Invenio software)
- Trustworthy Digital Repositories can guarantee Legacy across generations

https://www.archiver-project.eu/deployment-scenarios-technical-summaries/cern-digital-memory

DESY - Archiver use-cases

Sergey Yakubov, Martin Gasthuber
June 8, 2020


Main sources of data to be archived and preserved
European XFEL
Schenefeld / Schleswig-Holstein
European XFEL
>30PB annual
DESY
Hamburg
FLASH 1
FLASH 2
PETRA III
2-4PB annual
two sites
Hamburg
Zeuthen (near Berlin)
science areas
particle physics (LHC, Belle 2, …)
photon science (EuXFEL, Petra III, FLASH)
accelerator research (wakefield, …)
astrophysics (mainly Zeuthen)
all areas “data intensive science”

automation

scale - #objects, volume, bandwidth

API/CLI usage / less interactive

Archiver challenges

individual scientist/ small working groups

- scientist is the archivist
- publication material + condensed data + reference to full datasets
- DOI handling
- mainly interactive access
- few TB, 100MB/sec, 10K objects
- ~0.2-0.5PB annual

- more or less ‘classical preservation model/practices’

mid-size working groups (Petra III experiment)

- nominated member of the group is the archivist (on behalf of)
- raw + derived data + code
- DOI + open-data handling
- comply with site data policy
- few 10TB, 1-2GB/sec, >150K objects
- <50% interactive access
- ~2-4PB annual

large collaboration / site management (EuXFEL organization)

- site nominated archivist responsible for all experiments
- raw + calibration data + code
- DOI + open-data handling
- comply with site data policy
- few 100TB, 2-10GB/sec, >30K obj.
- very low interactive access
- >30PB annual

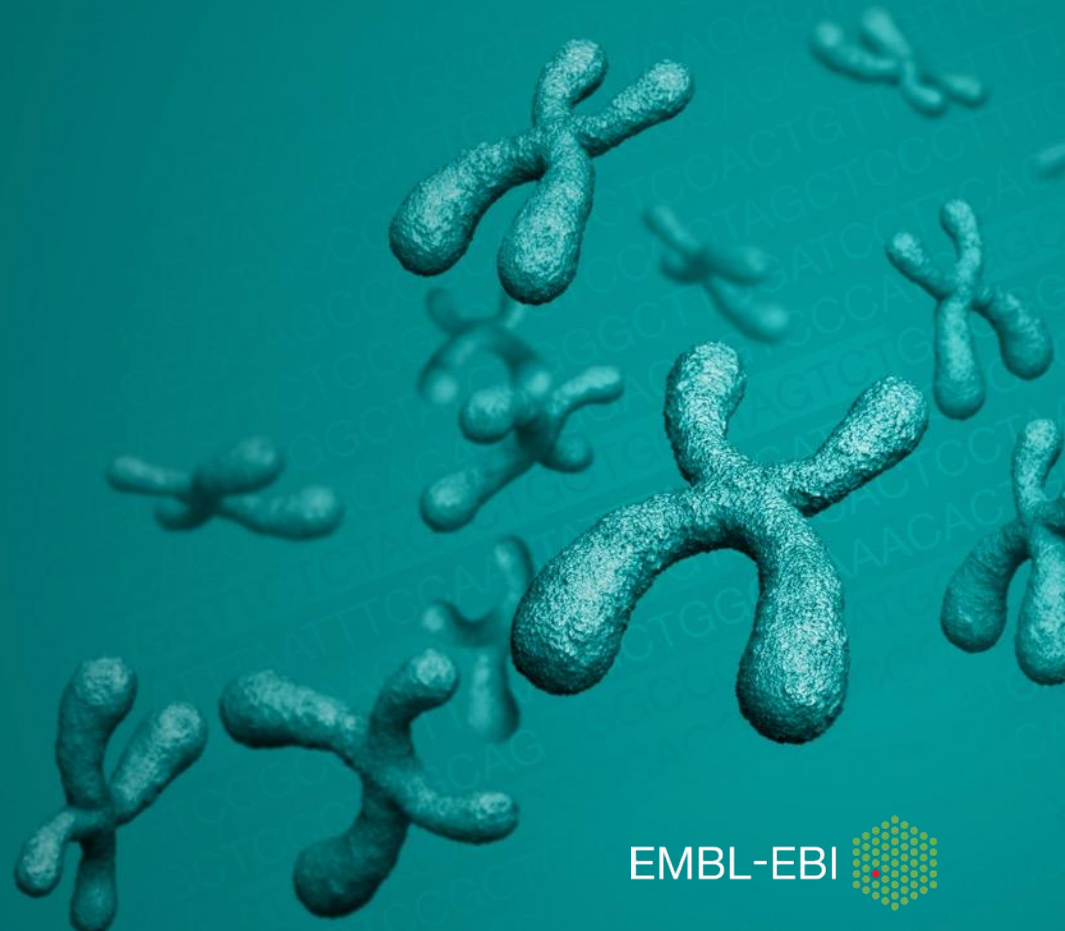
primary bit-stream storage & MD handling/storage on-site, hybrid to ‘private cloud @other labs’ / public cloud (handle ‘open-data’ and higher availability/redundancy, integration in existing preservation process (DPHEP)

The European Bioinformatics Institute

Use-cases for the ARCHIVER project

Tony Wildish

wildish@ebi.ac.uk

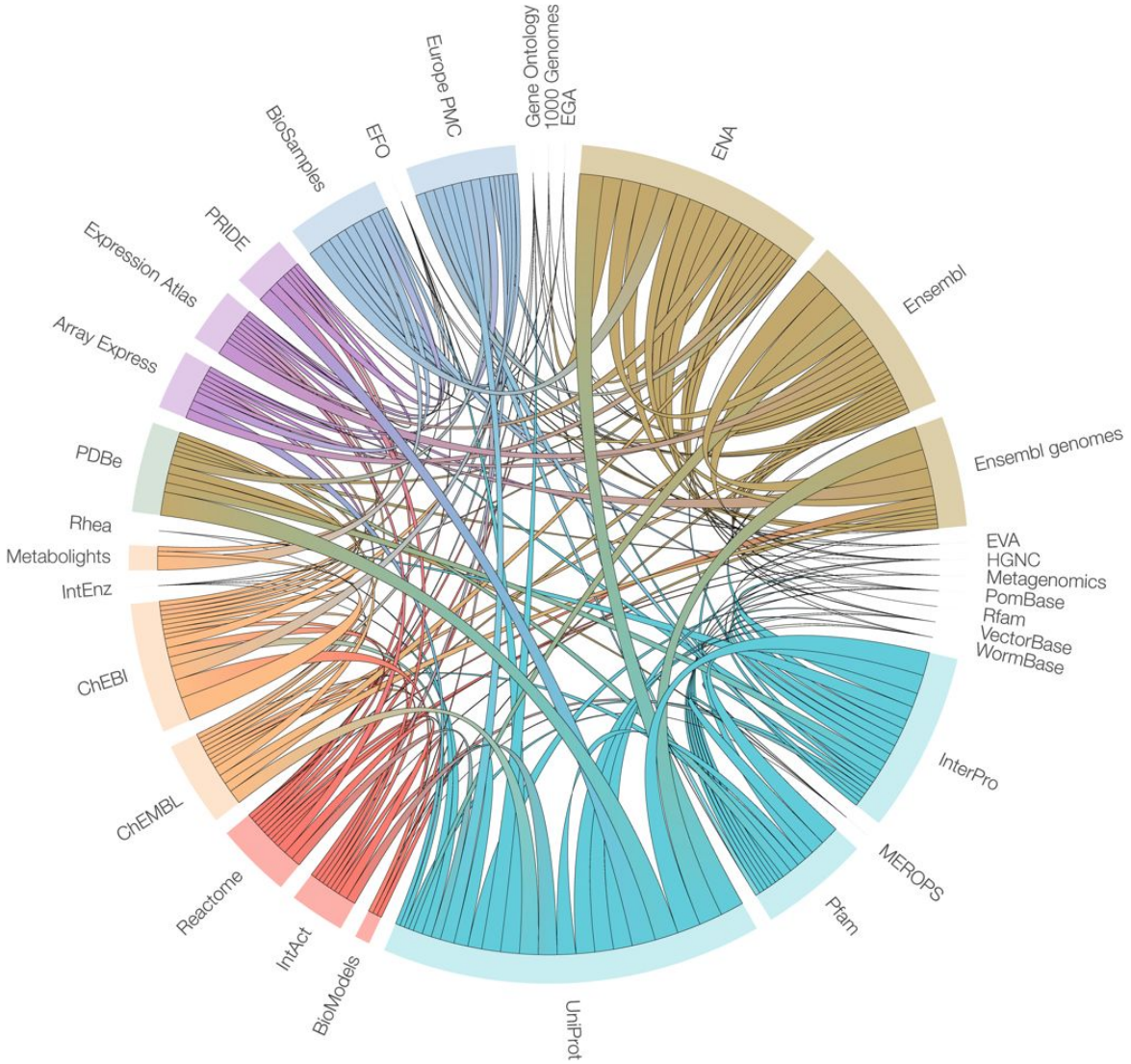


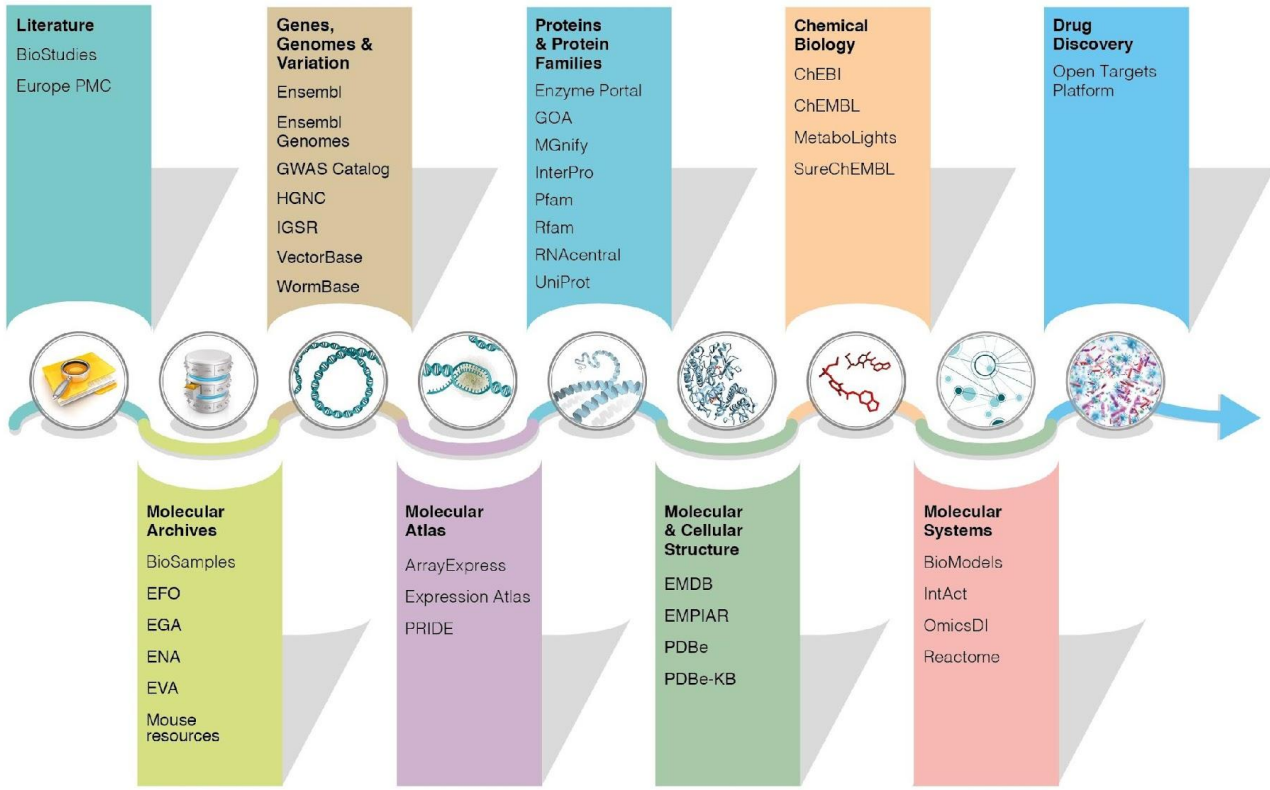
EMBL-EBI

What is EMBL-EBI?

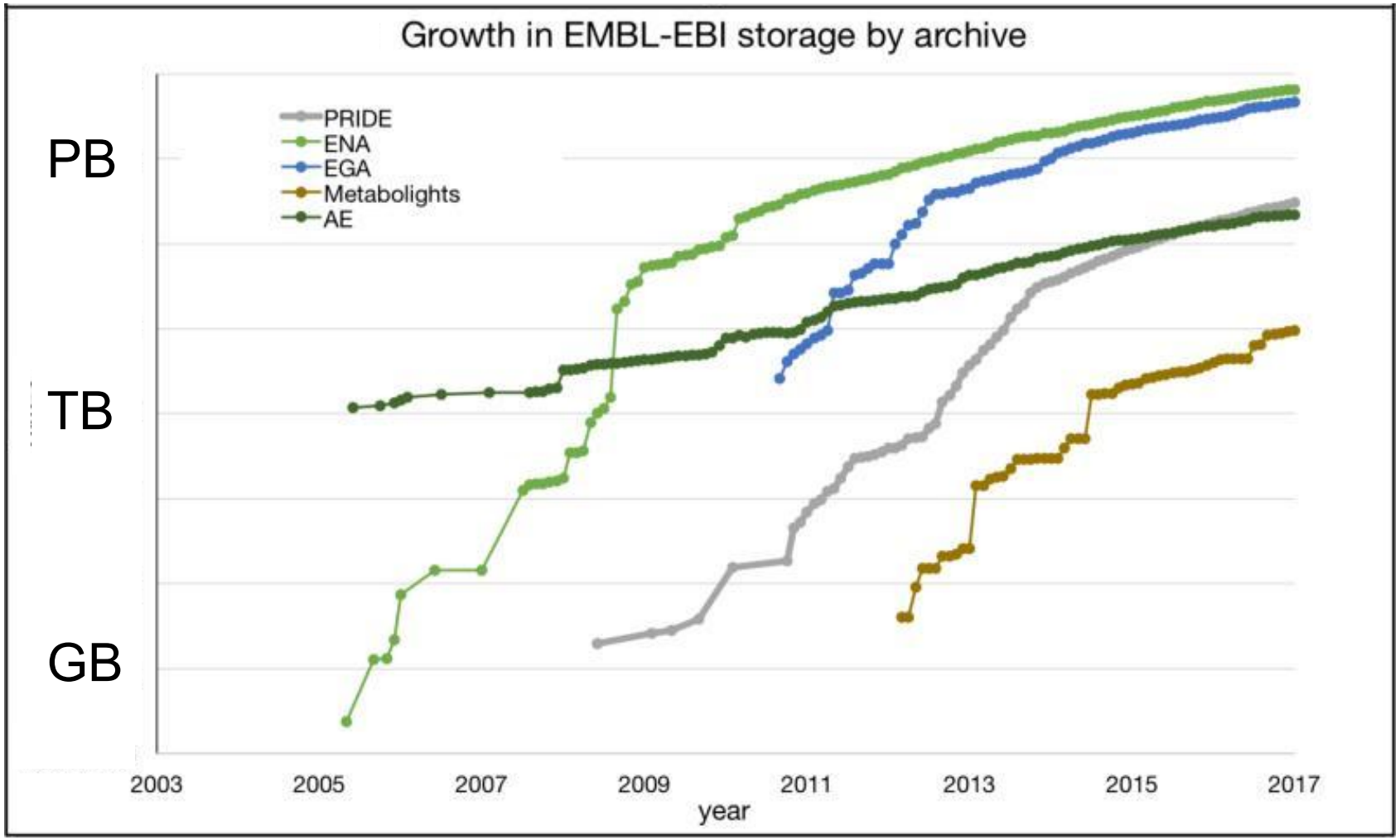
- Europe's home for biological data services, research and training
- A trusted data provider for the life sciences
- Part of the European Molecular Biology Laboratory, an intergovernmental research organisation
- International: 650 members of staff from 66 nations

Data resources at EMBL-EBI





Increasing Data, Increasing Analysis



Storage growth at EBI

- Data volume doubles every two years
- No reason to expect that to slow down

EGA and ENA account for the bulk of the data

- DNA sequences

PIC
port d'informació
científica

ARCHIVER
ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

PIC Deployment Scenarios

V. Acín, J. Casals, M. Delfino, J. Delgado

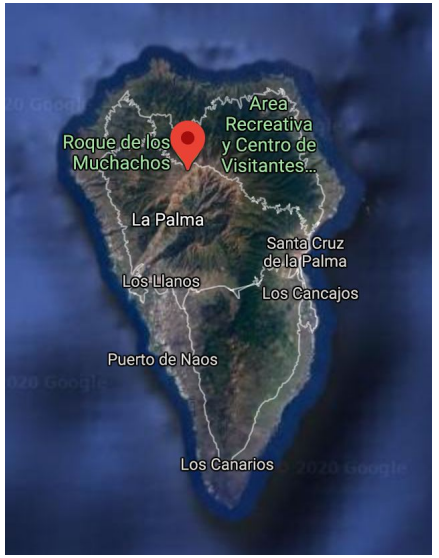
ARCHIVER Phase 1 Award Ceremony
June 8th 2020

Institut de Física
d'Altes Energies

Centro de Investigaciones
Energéticas, Medioambientales
y Tecnológicas

- Actors:
  - **Scientific Instruments**
    (example used will be MAGIC Telescope in La Palma, Canary Islands, Spain)
  - **Private Data Centers** extended by **Contractor Archiving Services**
    (example used will be PIC Data Center + ARCHIVER contractors)
  - **Instrument Scientists**
    (closed group of well identified worldwide users with strict privacy needs)
  - **External Scientists** (other identified scientists, public access)
- Scenarios:
  - **File safe-keeping**: large and mixed-size, various retention policies
  - **In-archive data processing**: avoid external downloads and uploads
  - **Data distribution to Instrument Scientists**: AAI with roles
  - **Data utilization by External Scientists**: multiple AAI schemes

Instrument Example: MAGIC Telescopes located at
Observatorio del Roque de los Muchachos, La Palma, Canary Islands

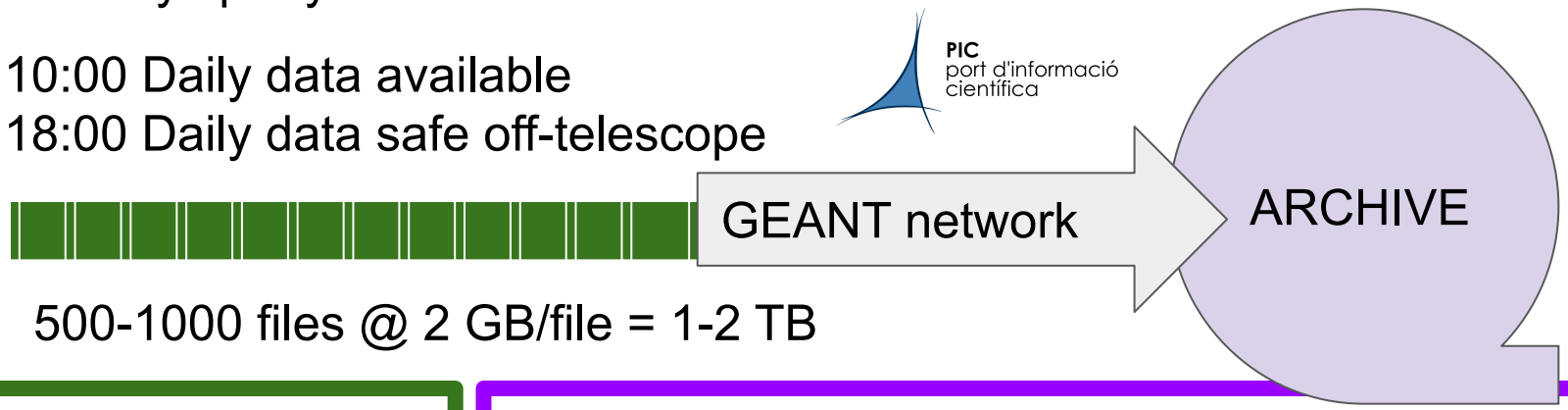


Large-file safe-keeping scenario

MAGIC Telescopes located at
Observatorio del Roque de los Muchachos,
La Palma, Canary Islands

365 days per year:

10:00 Daily data available
18:00 Daily data safe off-telescope



500-1000 files @ 2 GB/file = 1-2 TB

Data characteristics:
Inmutable (read-only)
Binary private format
Single bit error in a file renders it useless
Two metadata items: filename, checksum

Data stewardship (<u>for one yearly instance</u>)
Year 1: Data accumulates: 150k 2 GB files = 300 TB
Years 1-6: Data are bit-preserved

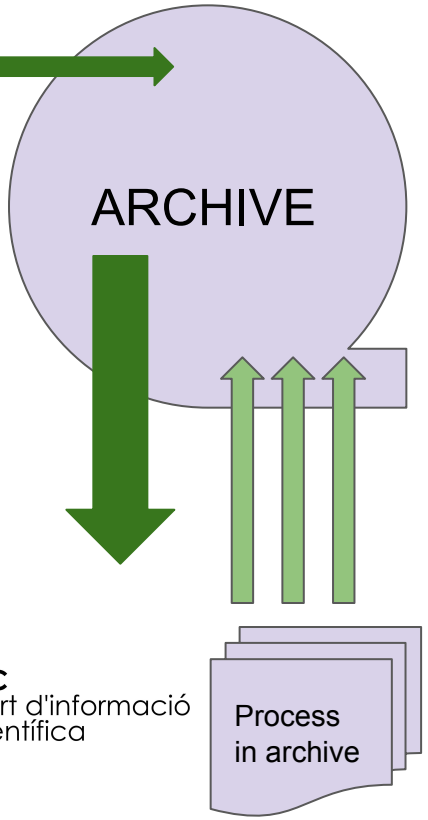
Full 300 TB recalled to PIC at random time(s) in years 2-6

Challenge:
Affordable cost for services with required performance and reliability

Mixed-type file safe-keeping and processing

Data stewardship (type A):
Year 1: Data accumulates: 150k 2 GB files = 300 TB
Years 1-6: Data are bit-preserved

Full 300 TB recalled to PIC at random time(s) in years 2-6

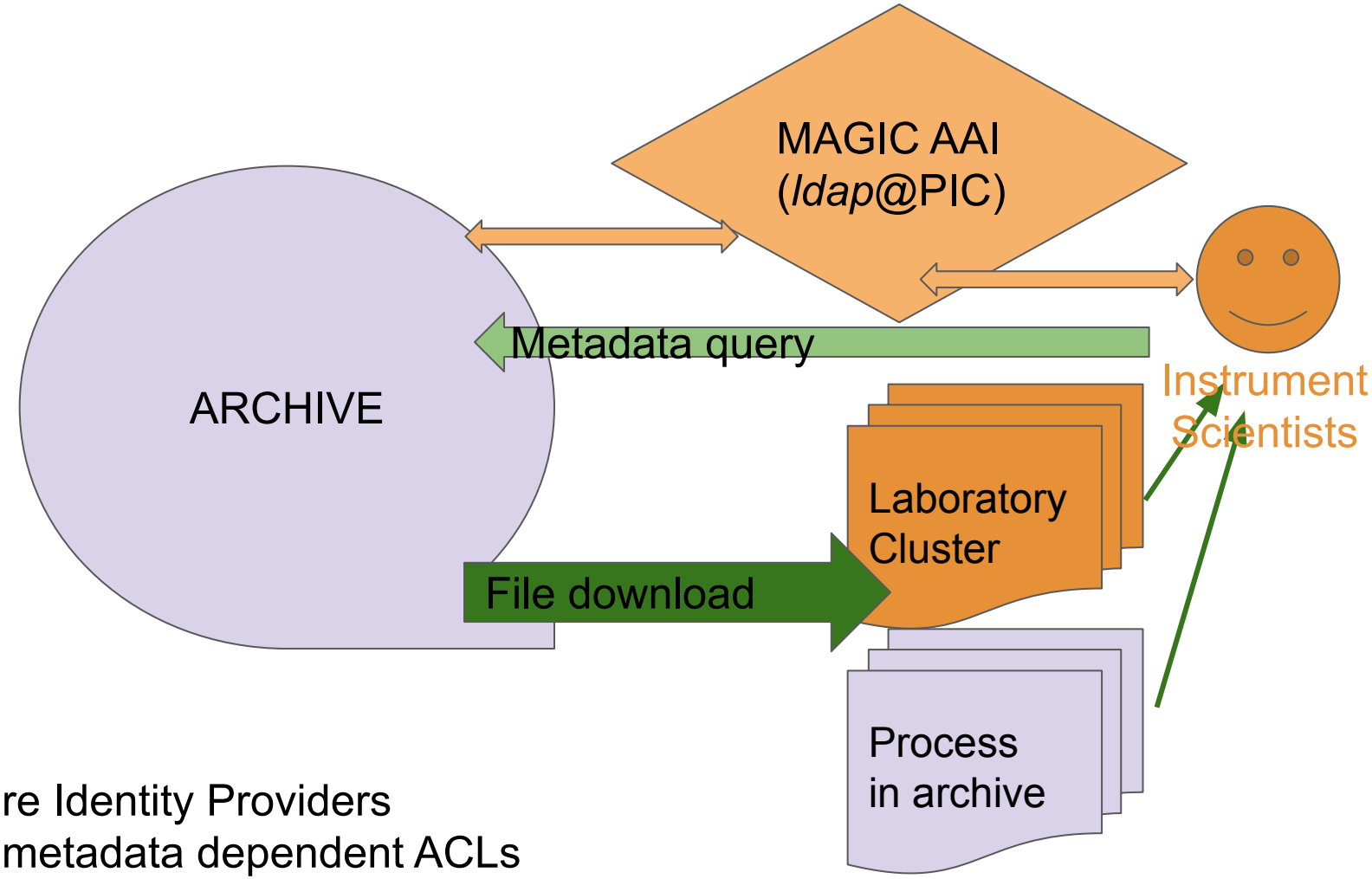


Data stewardship (types B, C, D, …)
Long-term retention (up to several decades)
Metadata driven retention
Metadata driven quality of service
Versioning (same metadata, different data)

Data characteristics:
Inmutable (read-only)
Binary private format
Single bit error in a file renders it useless
Size per file 2 MB to 200 MB
Metadata characteristics:
Dozens of metadata items per file
Allow for metadata to be expanded
OAIS support

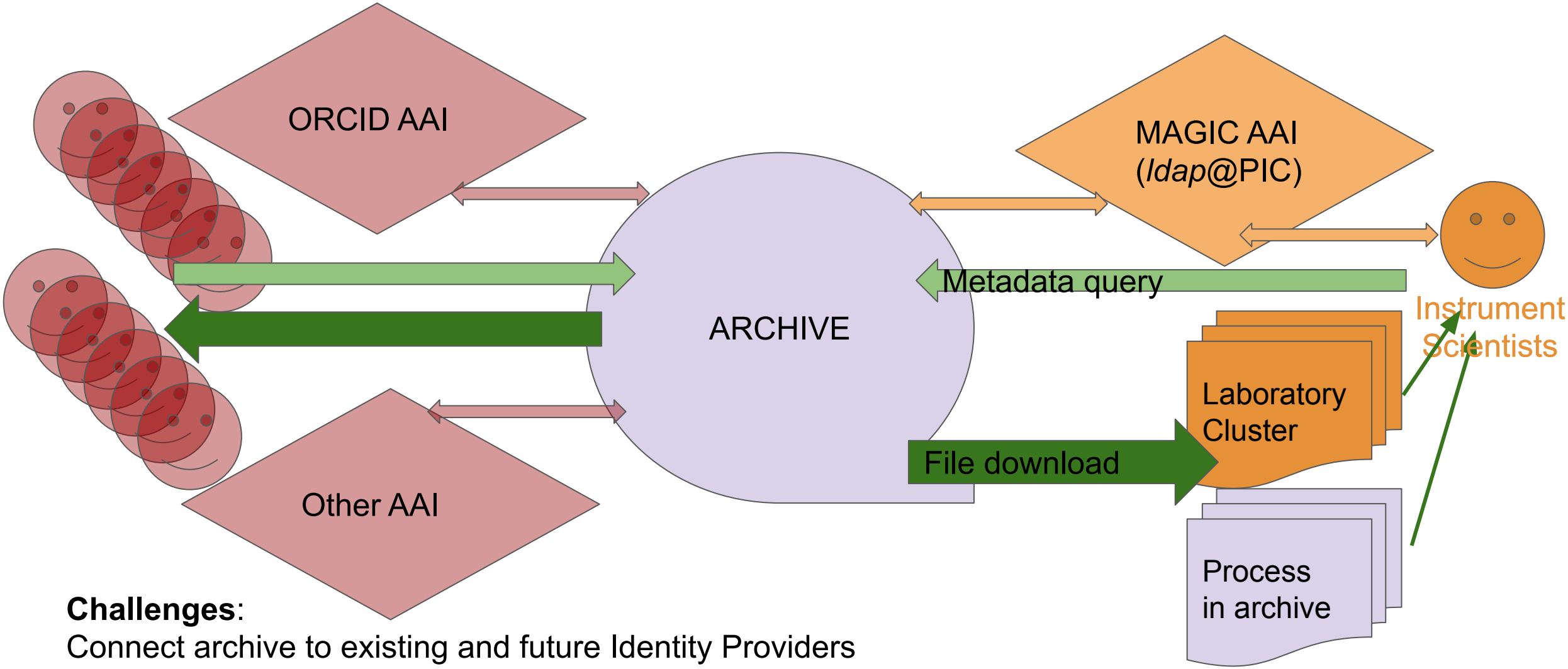
Challenges:
Metadata driven retention, quality of service, access control
Metadata driven recall of Petabyte volumes in millions of files with required reliability and performance
Low barrier for changing vendors (“exit strategies”)
Very elastic High Throughput data processing service for the in archive processing
Affordable cost

Data Distribution scenario: Instrument Scientists



Challenges:
Connect archive to existing and future Identity Providers
Fine granularity access control with metadata dependent ACLs
Affordable cost

Data Distribution scenario: Internal and External



Challenges:
Connect archive to existing and future Identity Providers
Fine granularity access control with metadata dependent ACLs
Affordable cost

BREAK

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

Award Ceremony

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

Arkivum - Google

arkivum

Bringing archived data to life

Google Cloud

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

ARCHIVER Project
Arkivum and Google Solution

01

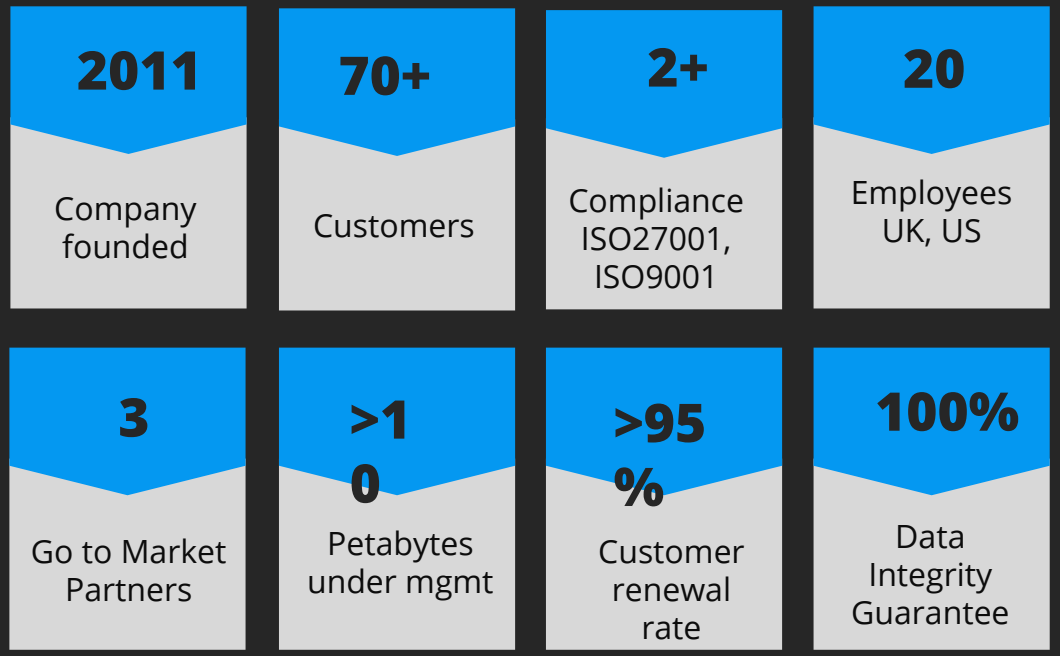
Introduction to Arkivum

Yaron Naor, VP Sales and Business Development

About Arkivum

Arkivum is the trusted software and service partner for long-term data management

We bring archived data to life!



Perpetua is a hosted solution for making your digital content safe, secure, accessible and usable for the long-term

Heritage and Higher Education

The University of Nottingham
University of Colorado Boulder
Princeton University
LSE
The Frick Collection
MoMA The Museum of Modern Art
ITMA Irish Traditional Music Archive Taisce Cheol Dúchais Éireann
University of Westminster
The Open University

Corporate

Jaguar Heritage
EBSCO Information Services
Royal Botanic Gardens Kew
The Archives of the Episcopal Church
South West Heritage Trust

Pharma and Life Science

ThermoFisher Scientific
idorsia
PHLEXGLOBAL
oncologica

* including partner sales

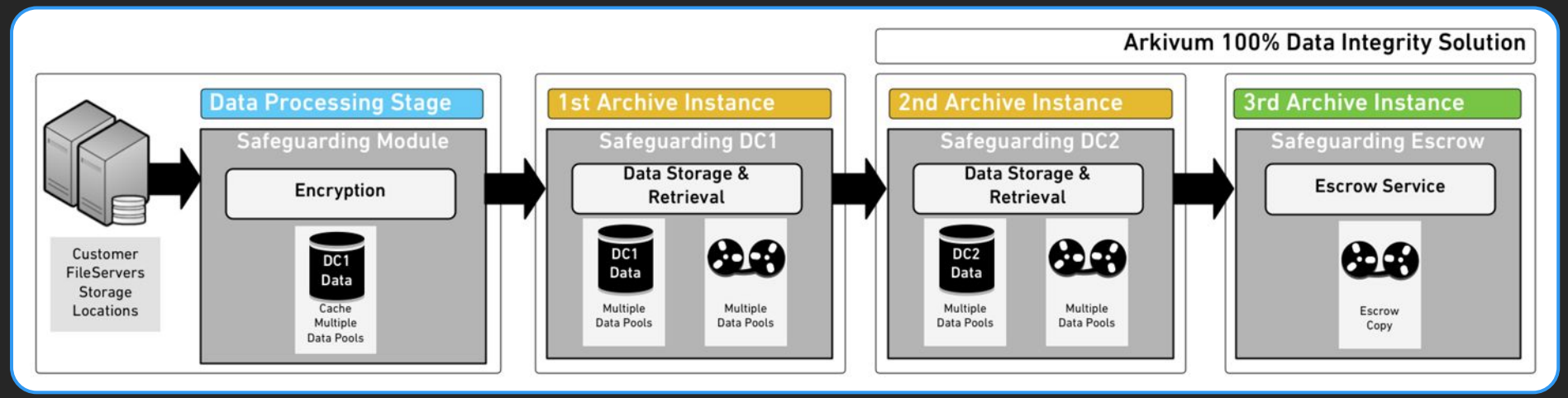
A bit of history

Founding Philosophies and Innovation

*Shortly after Arkivum was started in 2011, the idea of a '**data integrity guarantee**' backed by insurance was created.*

Arkivum provides the subject matter expertise to develop and implement the digital safeguarding and preservation "good practices", leading to:

- *You are not left to figure it out on your own – fully managed service*
- *There is a 100% data integrity guarantee backed by insurance*
- *Escrow copy provides built in exit plan with zero vendor lock-in.*



A new approach to long-term data management…

The Arkivum Approach

Vertical / Silo	→ Integrated / Unified	• Cross enterprise data management • Focus on needs, not data source
Closed / Monolithic	→ Open & Modular	• Reduce vendor dependency • Industry adopted open source
Manual / Reactive	→ Automated & Proactive	• Automating mundane tasks • Enhance governance, reduce costs

Core capabilities today

Designed to meet Real-Life Needs of the Markets We Serve

Safeguarding Data

100% data integrity guarantee

Digital preservation – formats protection from obsolescence

Evidence-ready data handling (authenticity, purge, share)

Open Specifications

Use open standards and specifications (bagit, METS, PREMIS)

Leverage open source technologies (Archivematica, AtoM, MongoDB)

Making Data Usable

Automatic metadata indexing, extraction and enrichment

Powerful search, discovery and share

Agnostic Solution

On-premise, private/public cloud, hybrid

Seamless integration with institutional applications, special collections, scholarly outputs and research data

- NxG Architecture
- Best Of Breed Technologies
- Elastically Scalable

Perpetua is offered as a fully managed service with guaranteed mission-critical SLA

Digital Preservation as a Service

Data Management Overhead Reduction

Subject Matter Expertise

- Minimal IT involvement
- Periodic integrity and fixity checks
- Digital preservation service File format decisioning

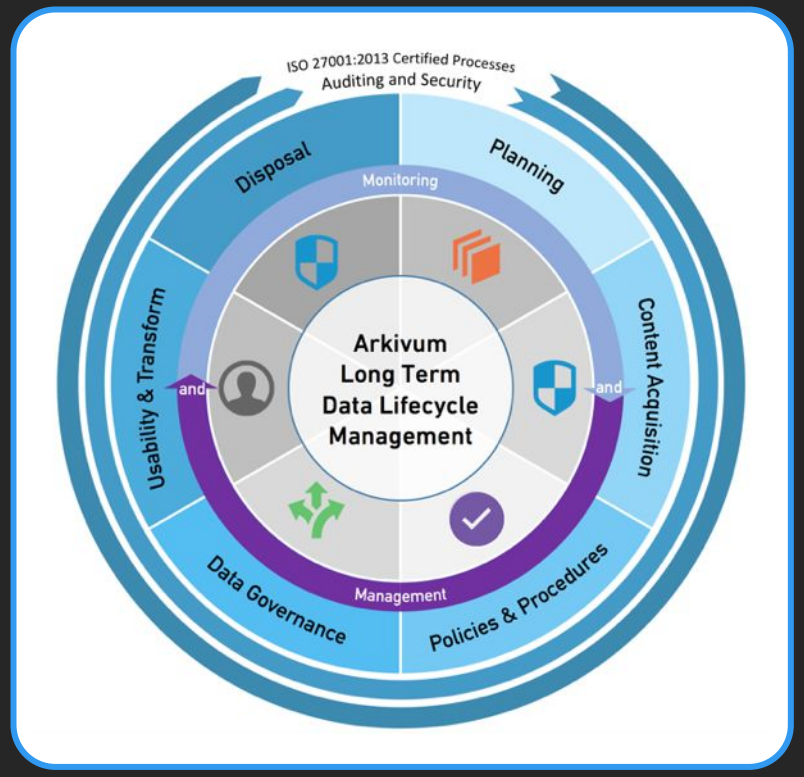
- Data growth requires continuous cost optimization
- Ensuring data safety and usability is a continuous effort
- Bulk migration

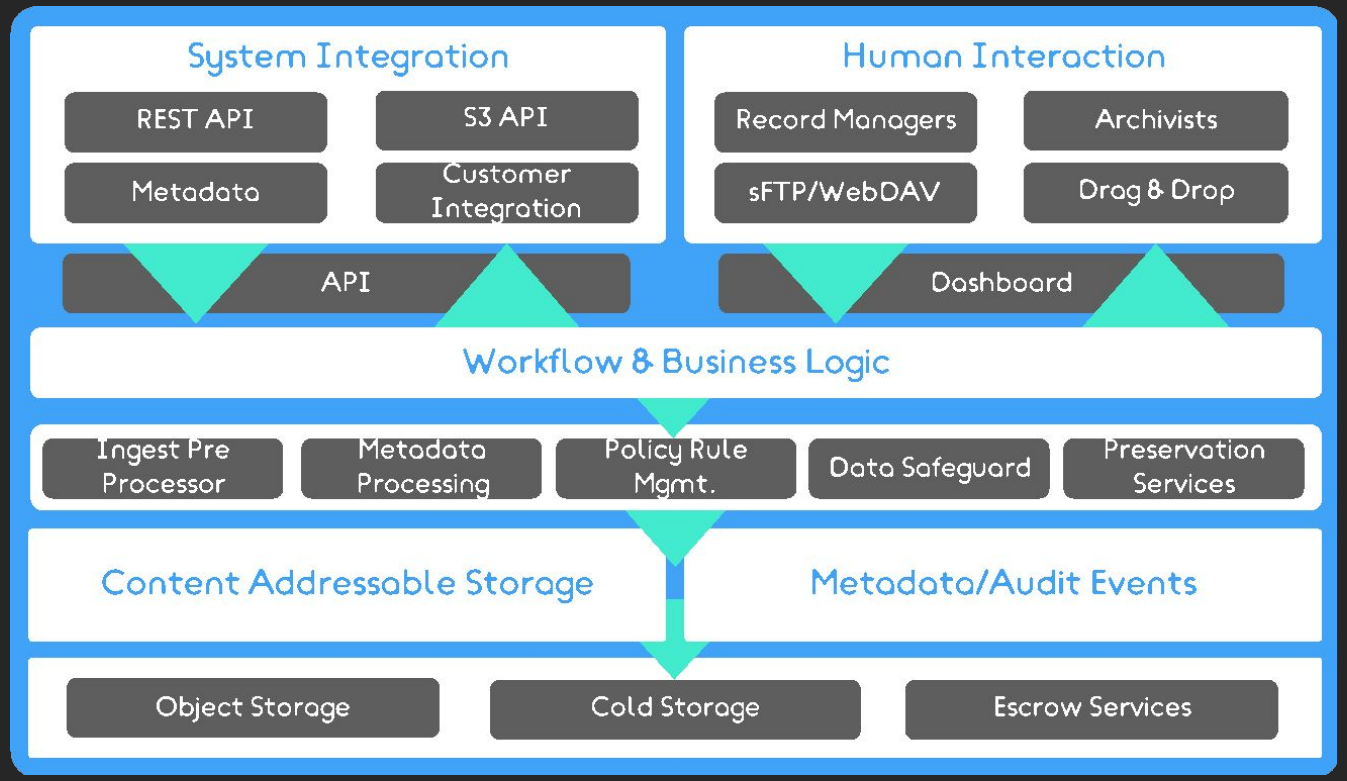
02

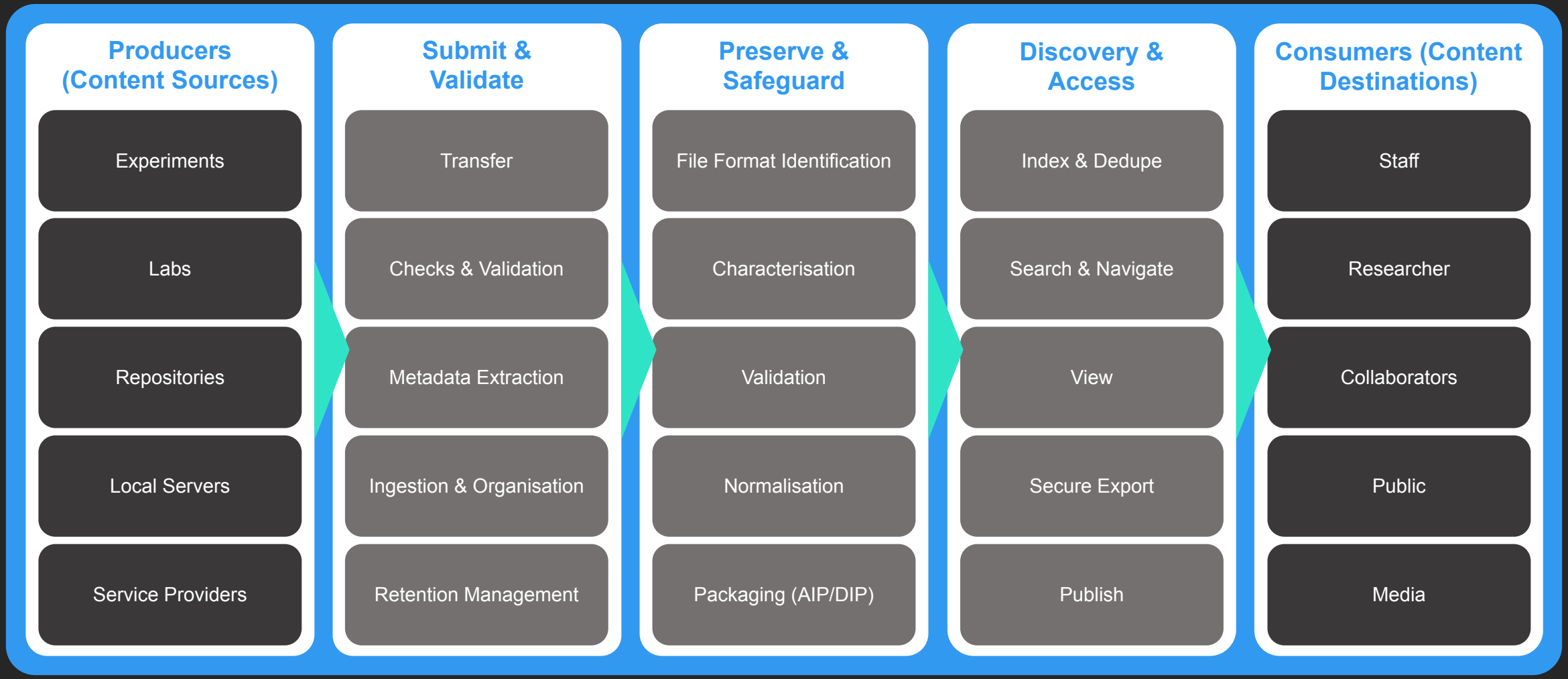
Solution Overview

Matthew Addis, Co-founder and CTO

Arkivum Perpetua: cloud hosted digital preservation and archiving







OAIS, TDR, Core Trust Seal, DPC RAM, Nestor

Open Standards, Open Specifications and Open Source Technologies:

PAR

Preservation Action Registries

Portland Common Data Model

Google Cloud Platform: enabling PB scale archiving and LTDP


Google Object Storage


Google File Storage


Google Compute Engine


Google Kubernetes Engine

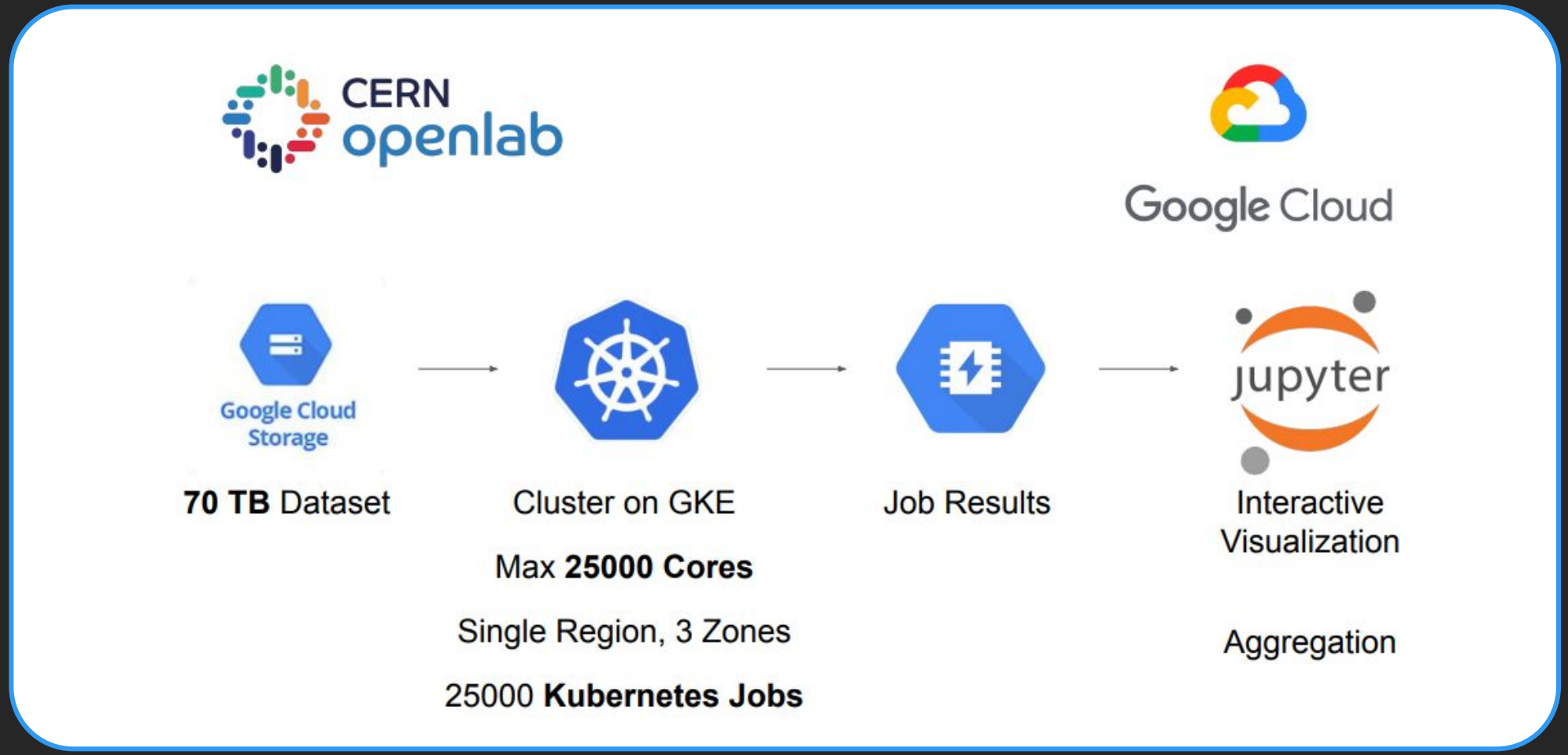

Google Operations


Google Security


High speed network

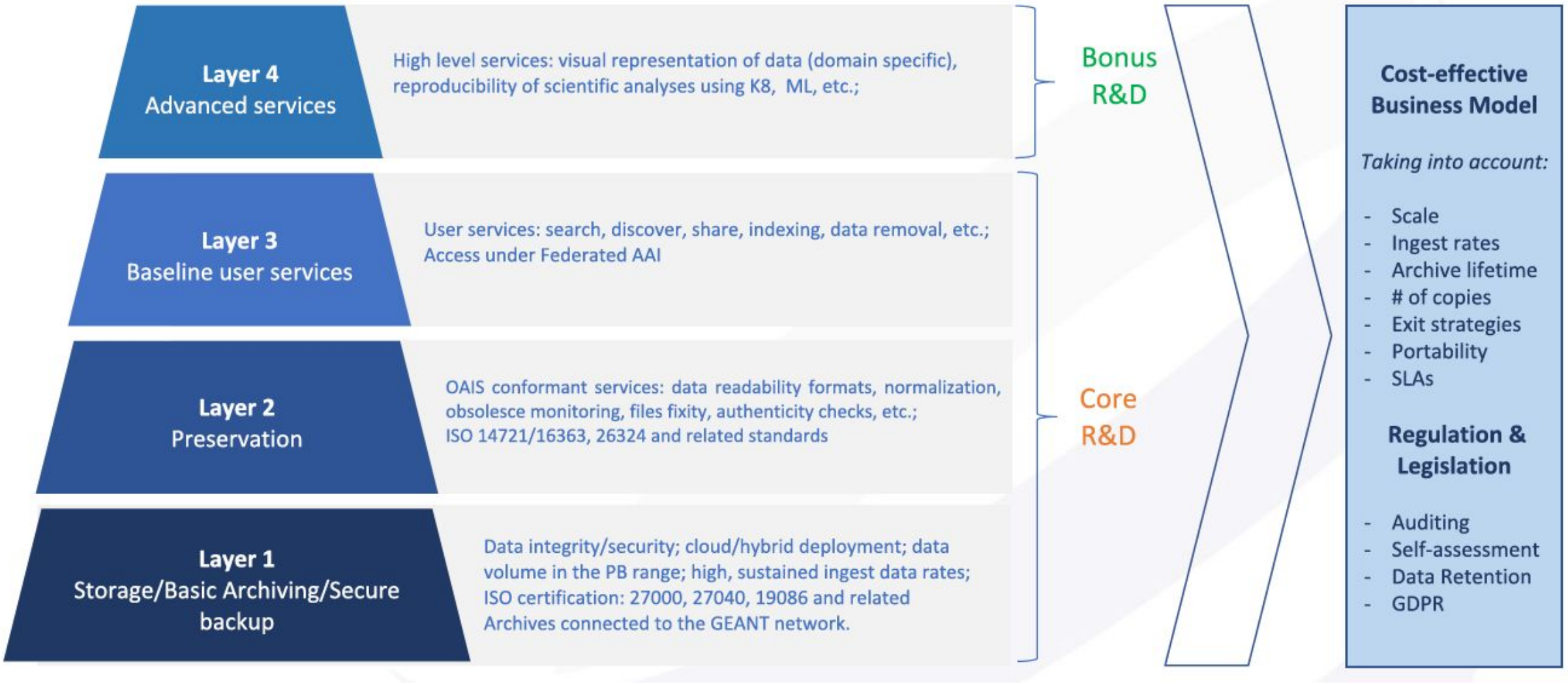
GÉANT
GEANT connected

Google Cloud Platform: hosting scientific applications



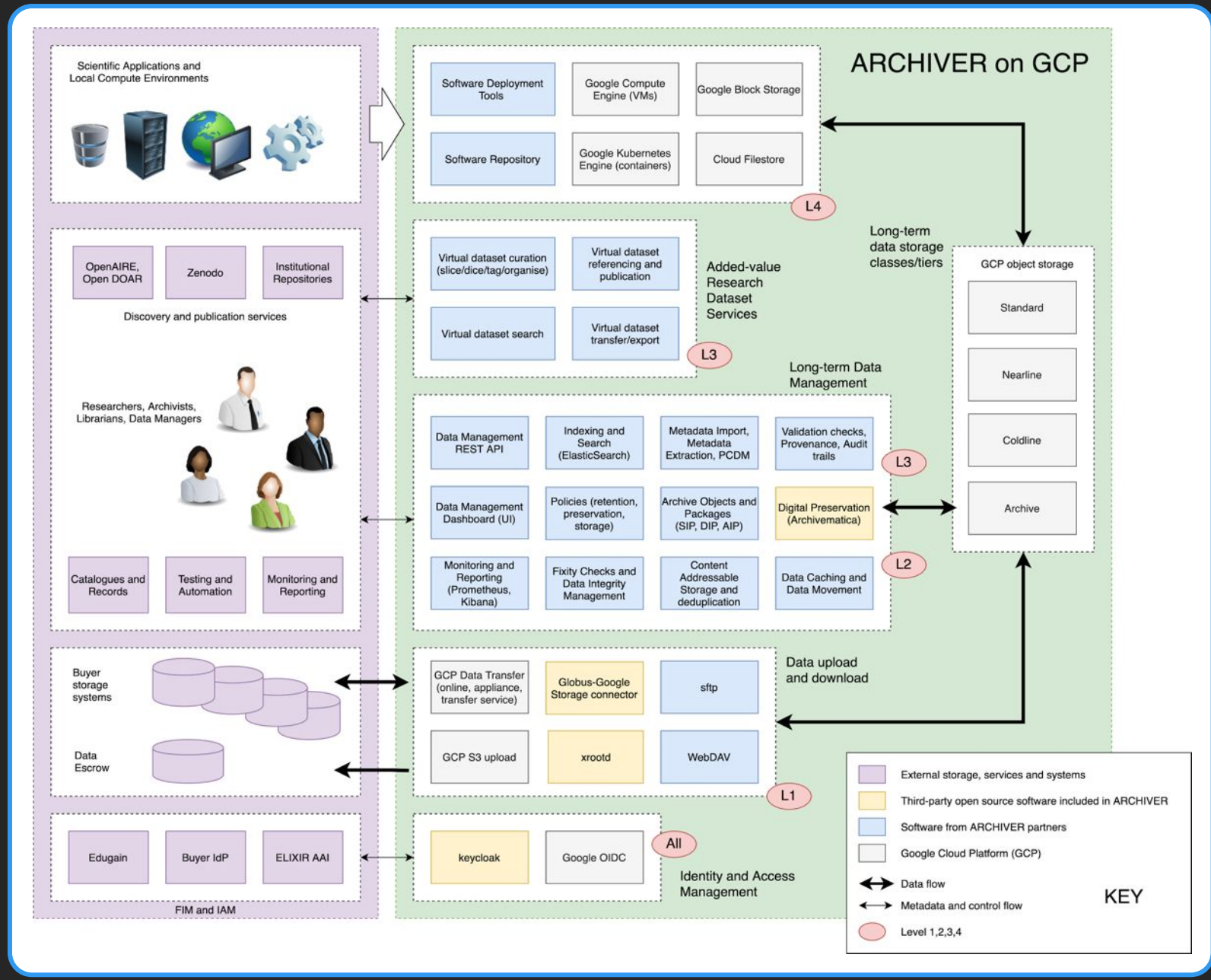
https://indico.cern.ch/event/773049/contributions/3581373/attachments/1939661/3215578/chephiggs.pdf

ARCHIVER requirements:



Arkivum / Google solution:

- Scalable storage and compute
- High speed ingest and access
- Policy based cost optimisation
- OAIS workflows and packages
- Digital Preservation rules and actions
- FAIR datasets and access
- Hosted scientific applications
- Open standards and specifications
- Exit and migration strategies



London Office

Top Floor, The Walbrook Building
25 Walbrook, London EC4N 8AF UK
T: +44 (0)1249 40 50 60
E: hello@arkivum.com

Reading Office

Landmark, 450 Brook Drive, Green Park
Reading, Berkshire RG2 6UU UK
T: +44 (0)1249 40 50 60
E: hello@arkivum.com

Boston Office

745 Atlantic Avenue
Boston, Massachusetts 02111 USA
T: +1 617 306 4563
E: hello@arkivum.com

arkivum

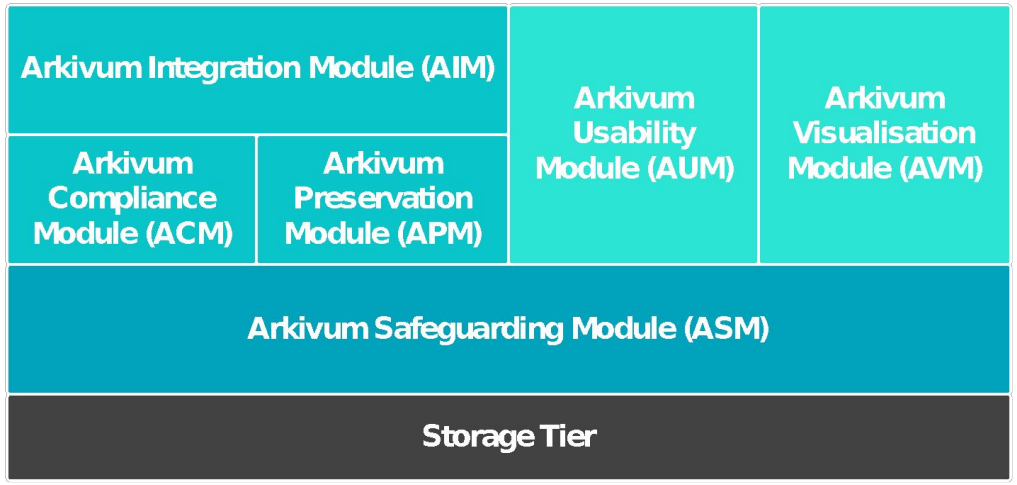
Bringing archived data to life

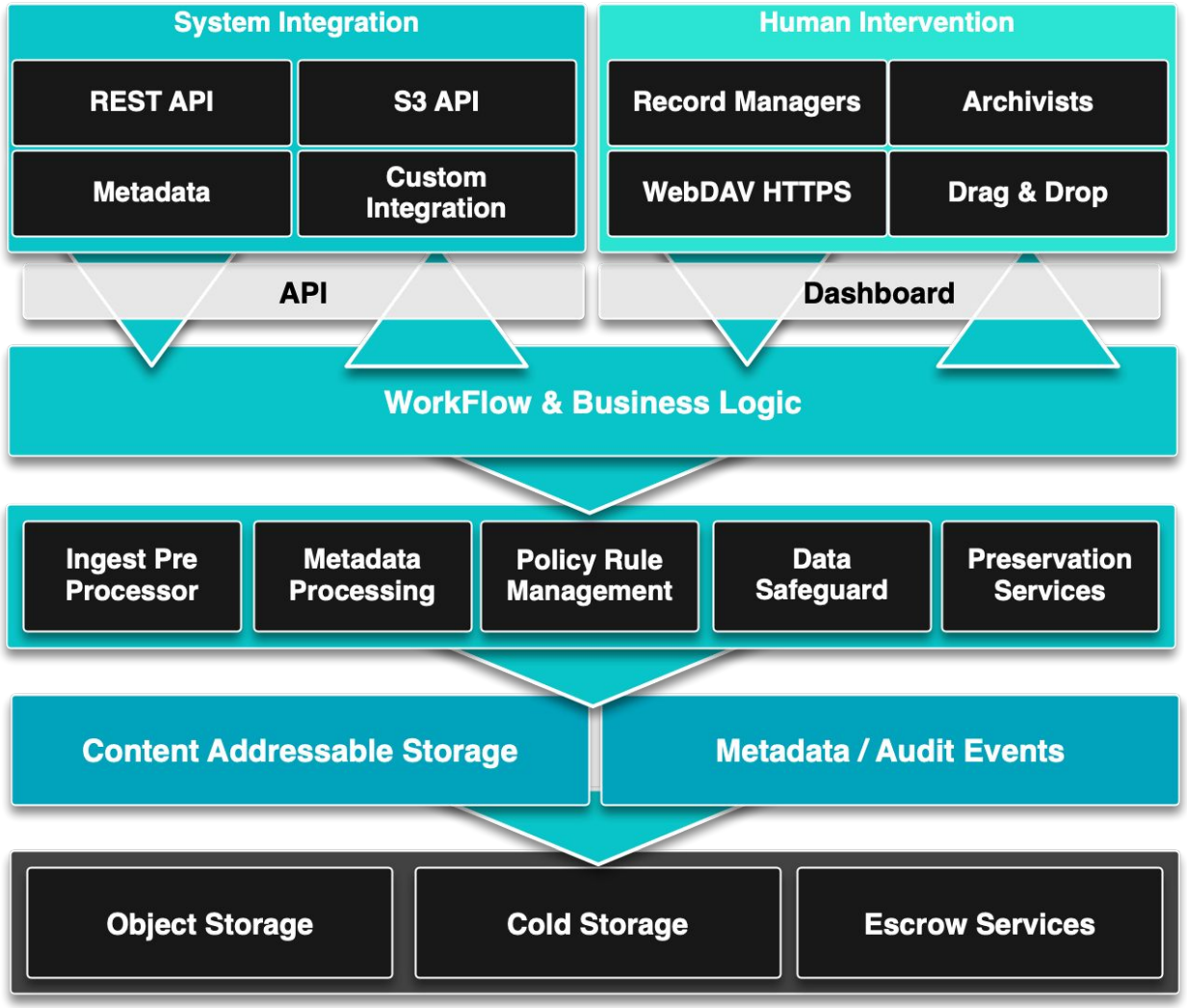
Thank you

www.Arkivum.com

Find us on LinkedIn or on Twitter @Arkivum

Arkivum Perpetua: cloud hosted digital preservation and archiving





ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

GMV – PIQL – AWS – SafeSpring

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

ARCHIVER

GMV & PIQL SOLUTION

WITH SUPPORT FROM AWS AND SAFESPRING

piql

Co-funded by the European Union

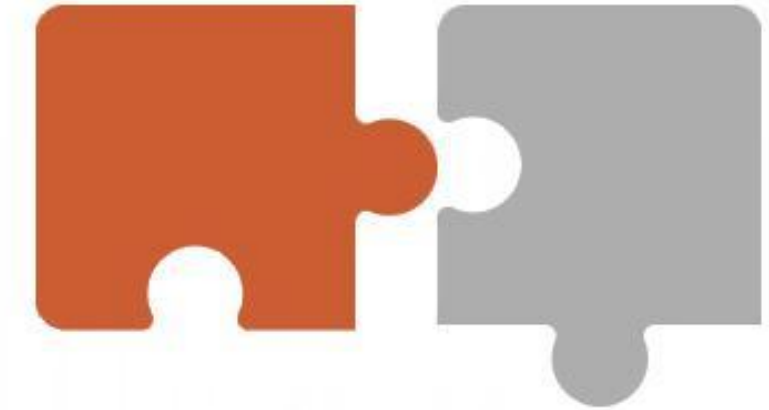
gmv®

Consortium for ARCHIVER PROJECT

piql

Co-funded by the European Union

PRESERVATION COMPANY



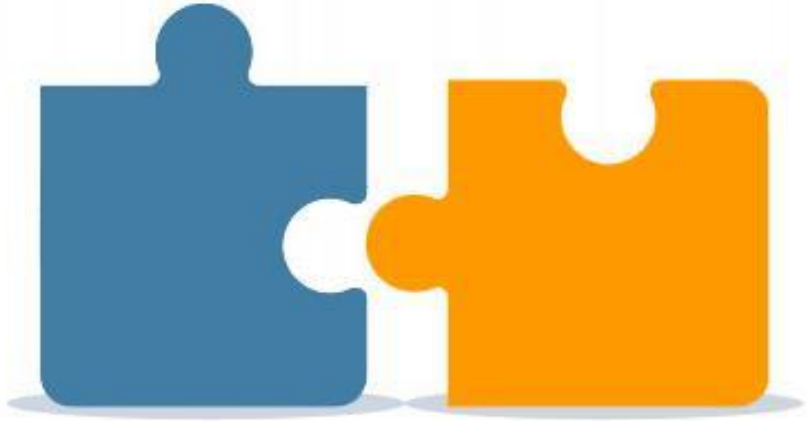
gmv

SYSTEMS INTEGRATION

COORDINATOR

CLOUD PROVIDER
CONNECTION TO GEANT

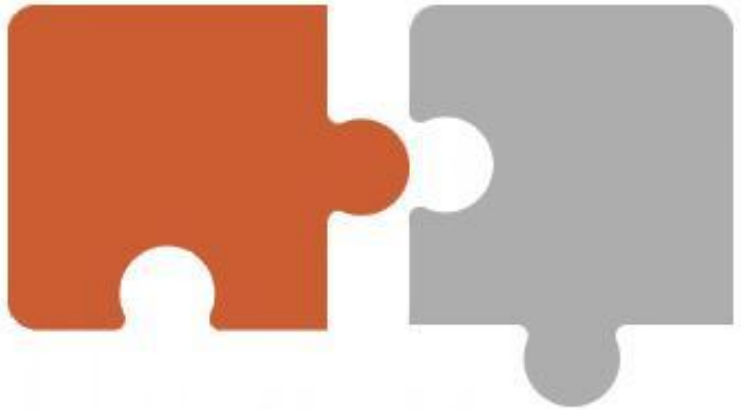
safespring



aws

Outcome for each partner

PRESERVATION COMPANY



SYSTEMS INTEGRATION

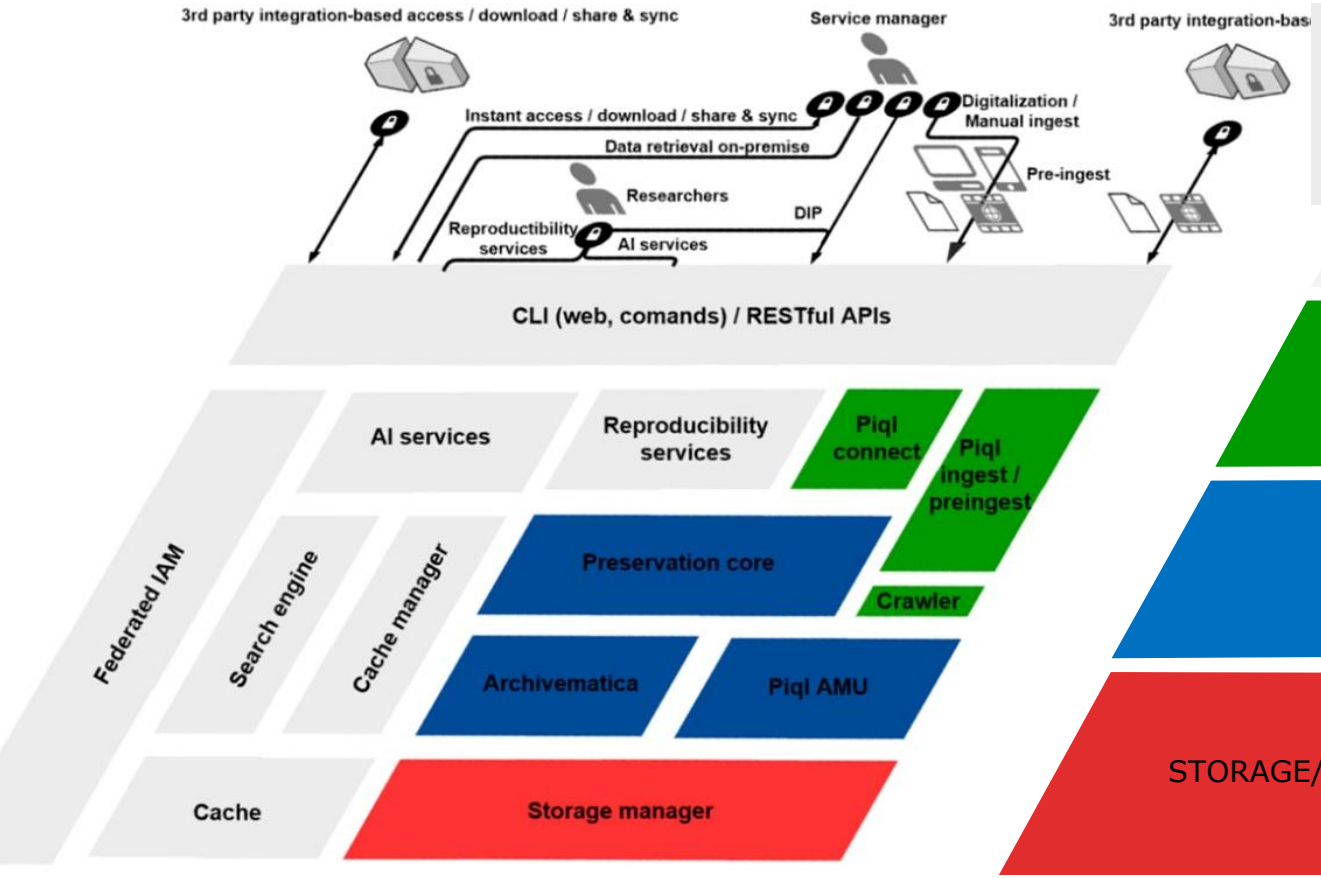
COORDINATOR

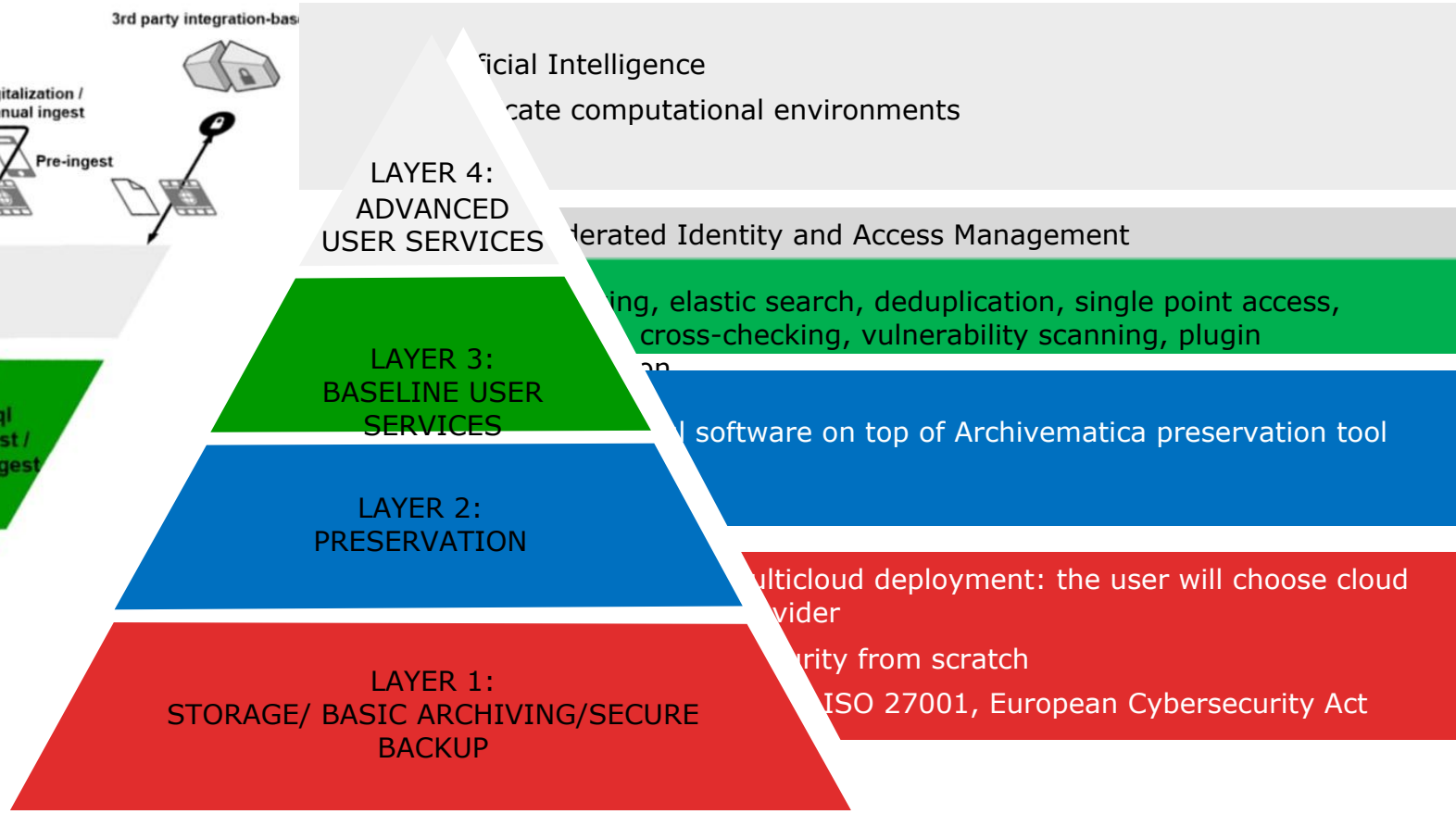
Piql's motto: Define requirements and architecture for preservation of research data, by improving the high volume preservation processes and develop automatic ingestion technologies to be used in the research domain. Lastly, to identify the type of data that would need irreplaceable safeguarding for a lifetime and to be preserved with the unique characteristics of piqlFilm.

GMV motto: create a new layer of services (cibersecurity, AI,..) on top of any preservation system to provide services for preservation companies where the goal is to deal with petabytes data.

...Under open source models

Architecture

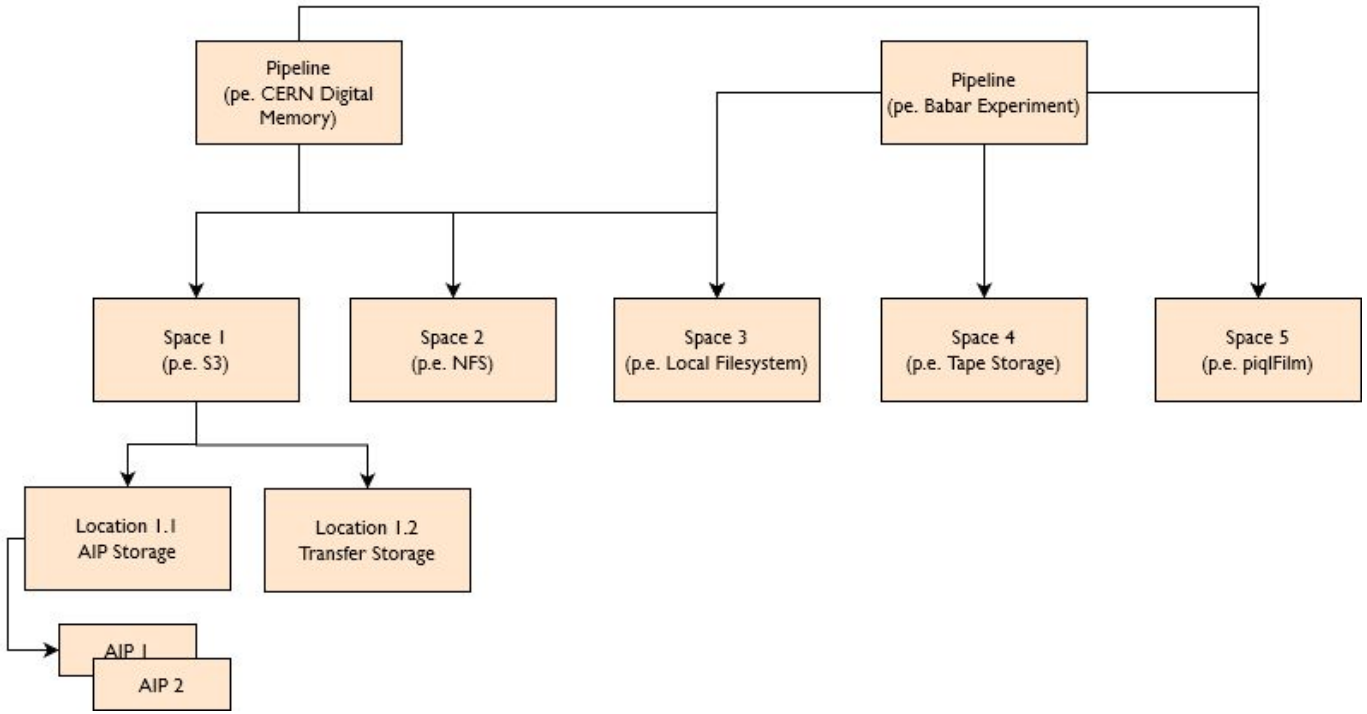




Services on top of open source developments

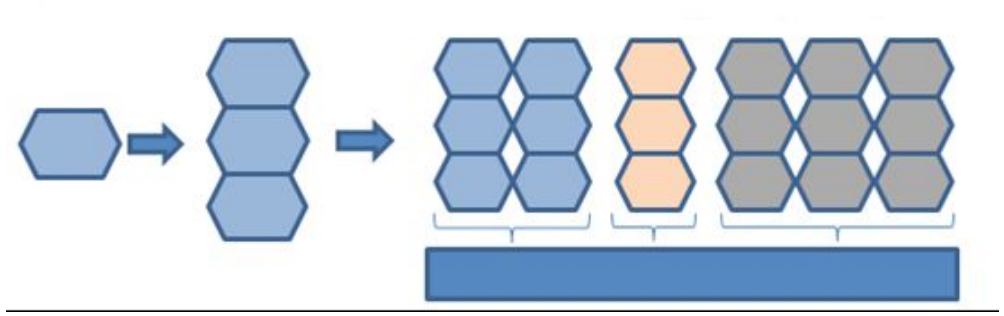
Storage Service Data Integrity

- Logical Scalable Storage Management
- Checksum on ingest
- Periodic checksum validation
- Cloud Independent
- Different Storage types for different uses
  - Fast
  - Large
  - Simple
  - Eternal



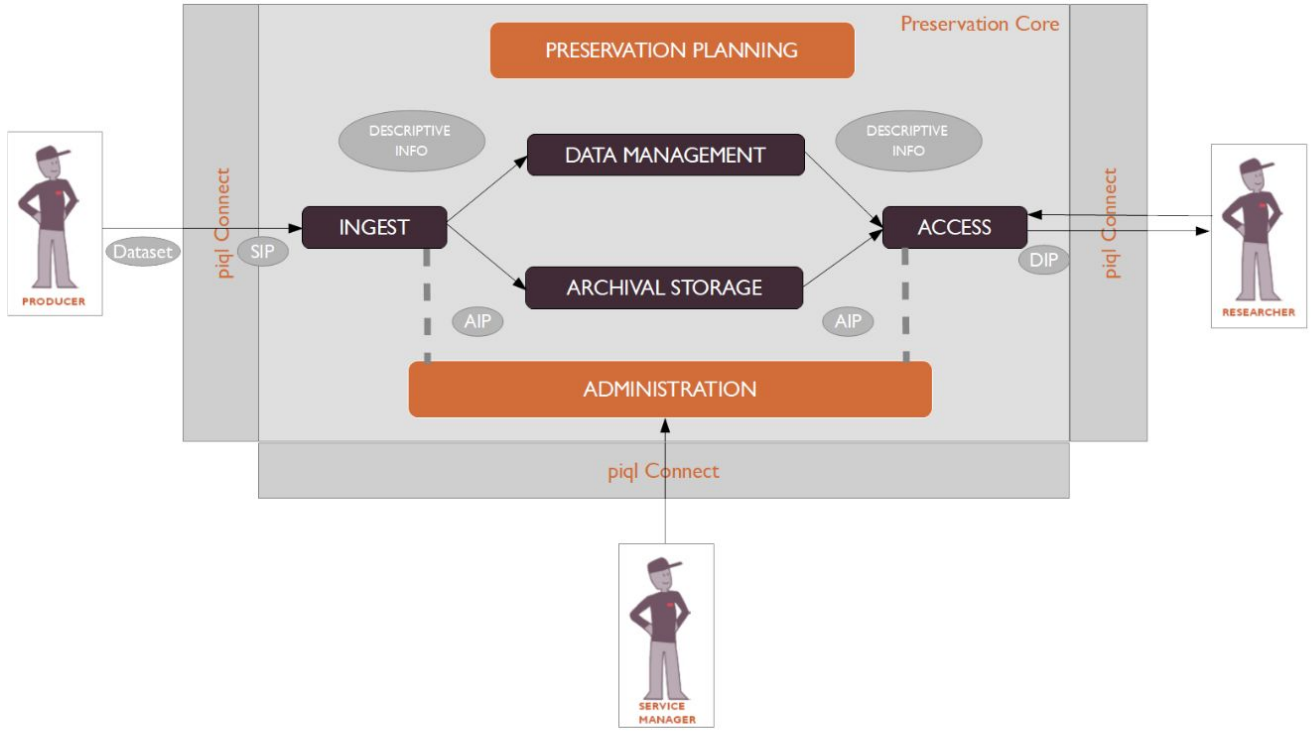
Solution scalability

- Modular Microservices Architecture
- Containering and Orchestration
- Multidimensional resource management
- Ingestion Mechanisms to maximize rates
- No limit on search and access data
- Adaptable indexes



OAIS Conformance and CoreTrustSeal certification

- Micro service approach
- Reduce complexity for the user
- Microservice based
- Ready to start certification paths on ISO16363 or others
- ISO27001 security certified

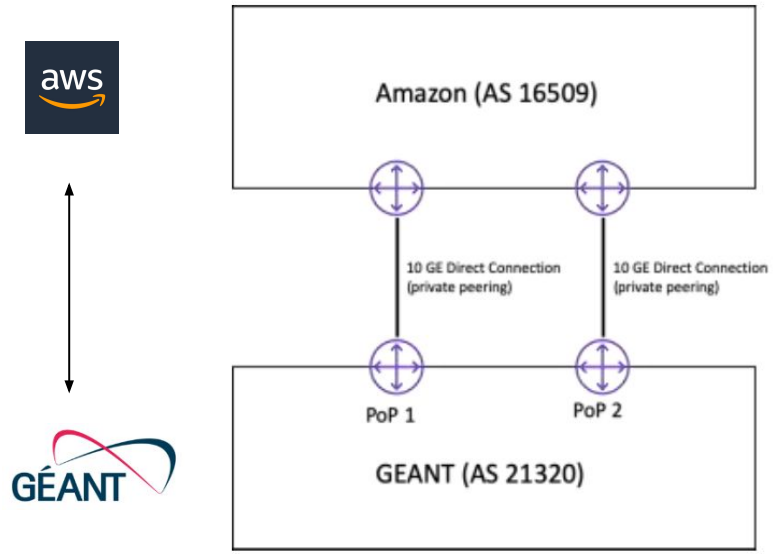


FAIR Guiding Principles

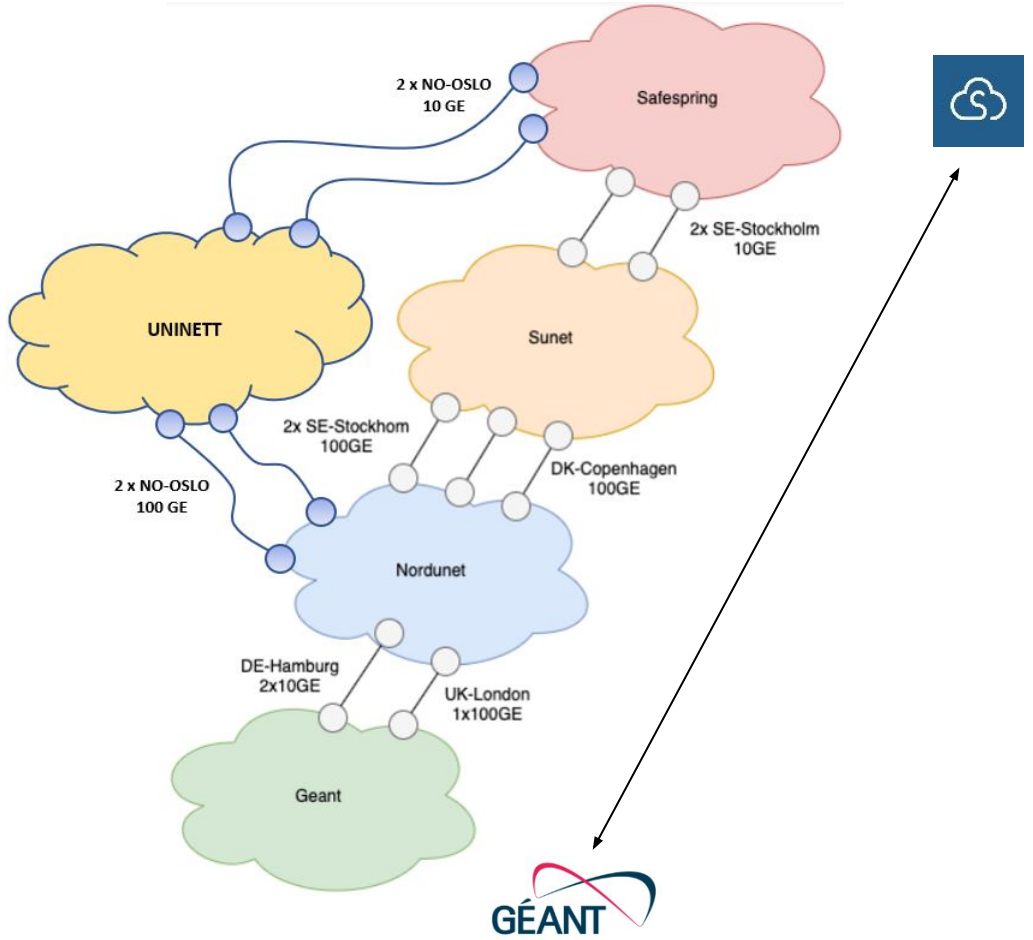
- Ensure Access in the long term

- Use of open technologies and standards
  - Powerful search engine
  - Open protocols
  - Industry Standards for metadata, data and packages
  - Context together with the data
  - Fairifiring tools for the researcher

Network Peering



- **AWS** has currently two direct connections to GEANT using 10 GE ports.
- If more bandwidth is needed between AWS and GEANT is possible to upgrade the bandwidth and use 100 GE ports.



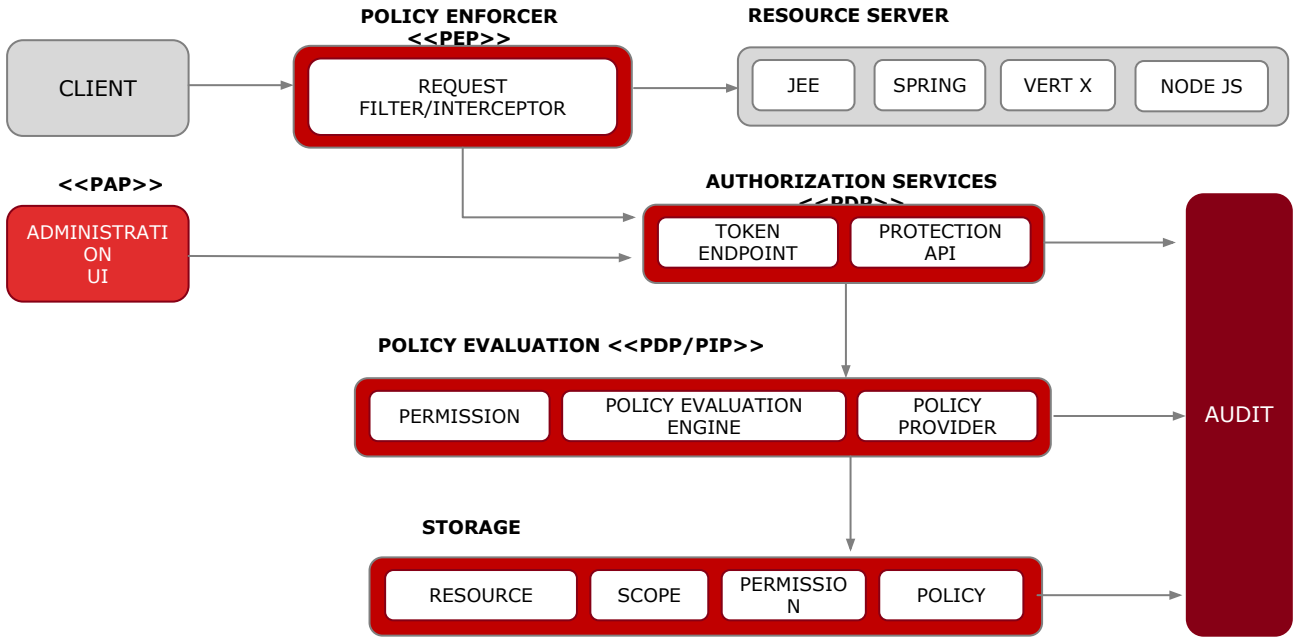
- **SAFESPRING** is directly connected to the NREN network in Stockholm (SUNET) and Oslo (UNINETT) with 2-way redundant 10 Gbps connections per site.

Support for identity and access management services

Inter-federation services will be based on:

- Open source single sign-on solution (keycloak)
- Standard authentication protocols for web will be supported, including:
  - Open ID Connect
  - OAuth 2.0
  - SAML 2.0
- Authorization policies able to combine:
  - ABAC
  - RBAC
  - UBAC
  - CBAC
  - Rule-BAC
  - Time-BAC
  - Other customised mechanism

Other interesting features are clustering, 2-factor authentication, social login, brokers (with Kerberos), etc.



Cybersecurity

ISO 27001 approach

- A governance framework.
- Confidentiality, Integrity and Availability.
- Risk Assessment Process.
- Set of policies, procedures and controls.
- Evaluation of Implemented Controls.
- Detection of Security Breaches.
- Compliance.
- Monitoring.

Single Sign On

dP

PKI

AntiMalware

HSM

Firewall

VPN

G/ACL

Data Privacy

SUBJECTS | BUYERS | CONTROLLERS | PROCESSORS | CUSTOMERS

Obtain consent

Exercise their rights

Give access to Data

Processing activities

Purpose and lawful basis

DPIAs activities

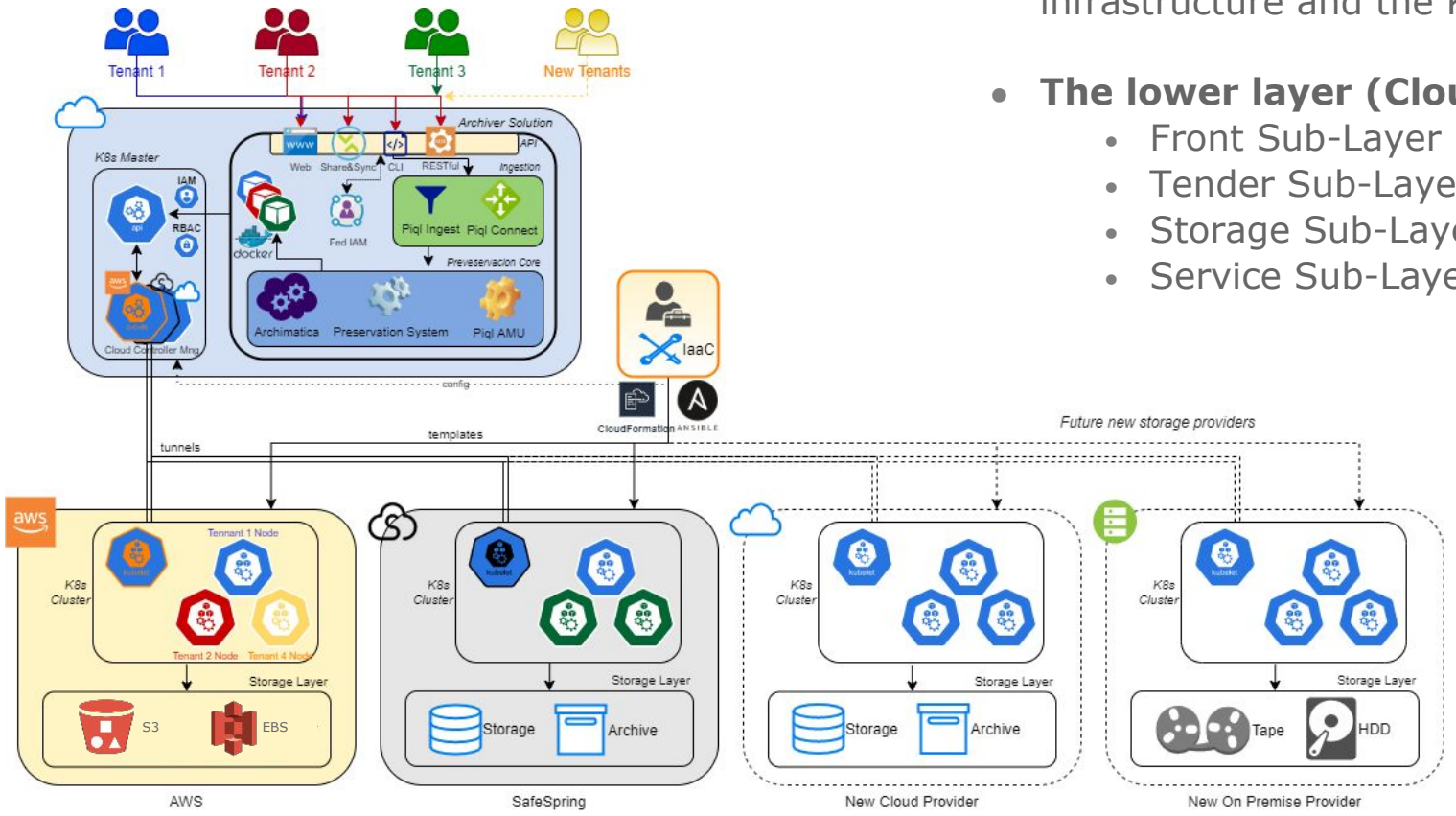
Retention & storage periods

Privacy by design & by default

Consent management

Subjects' rights management

Deployment Model



- **The upper layer (GMV layer)** contains the archiver solution infrastructure and the Kubernetes Master node
- **The lower layer (Cloud Layer)** split in four functional sub-layers:
  - Front Sub-Layer
  - Tender Sub-Layer
  - Storage Sub-Layer
  - Service Sub-Layer

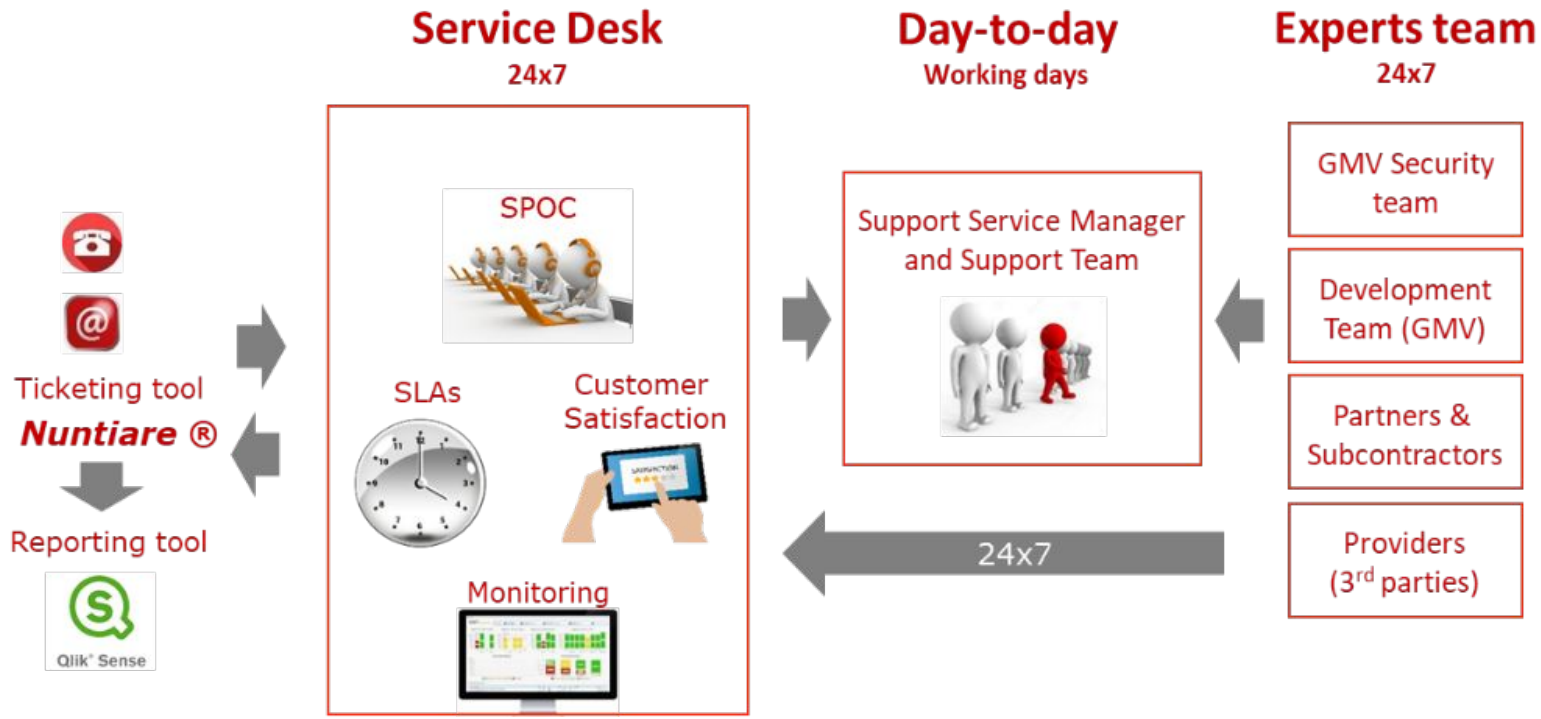
Cost-effectiveness of resulting services

- Savings on commercial licenses
- Multi cloud approach from design
- Customer empowerment
- Smart Costs control

Escalation process

- **Service Desk** as single point of contact.
- **2nd level:** (Only working hours) A Support Service Manager will be appointed to lead, coordinate and manage all support and maintenance activities.
- **3rd level:(Experts)**

  GMV Development.
  GMV Security.
  Piql and cloud providers.
  Other Providers.
- **4th level: Administrative escalation for special situations** (complaints, contingency situations, provide additional resources).

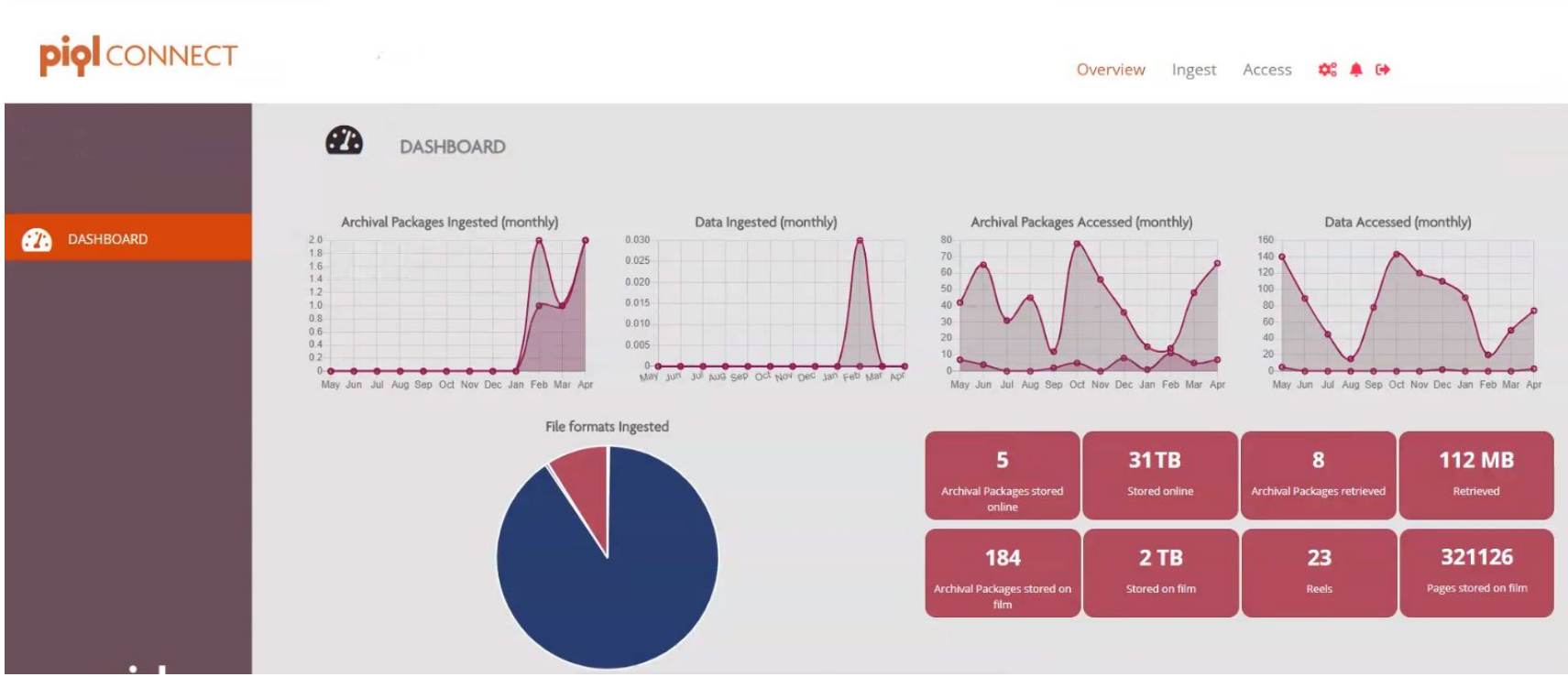


Licensing Model

- Service Oriented
- Flexibility on deployment and components
- Open Source tools
- Promoting End User independence

Merit of the Reporting, Accounting and Management portal

- Full Multidimensional Control Dashboard
  - Ingestion, Archive, Retention, Validation
  - Speed, Size, Cost, Users, ….
- Inference for future evolution



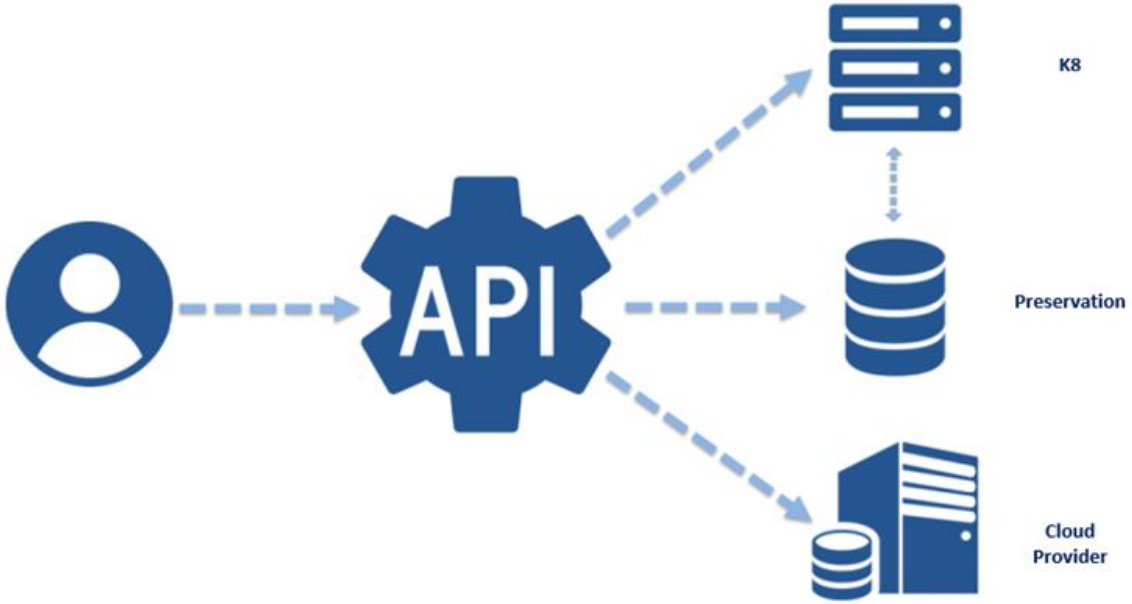
Resource management and service configurations

- Preservation Plan Design and Monitoring
- Retention Periods per dataset type
- Retention workflows and alerts
- Compliance features
- Resource quotas
- Templates
- QoS limits

Merit of the proposed API capabilities

Using API calls to generate, operate or execute different tasks related with the application or the infrastructure is a general rule in the solution. This highlights the need to define different API domains, to get a more controllable, accurate and secure solution. Three API domains shall be defined to interact using these capabilities with the environment:

- Preservation API Domain
- K8 API Domain
- Cloud Provider APIs Domain



Commercialization Plan and Impact

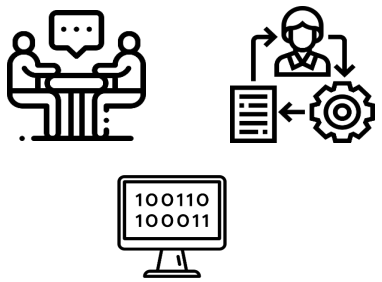
- Services on top of open source developments.
- Customers:

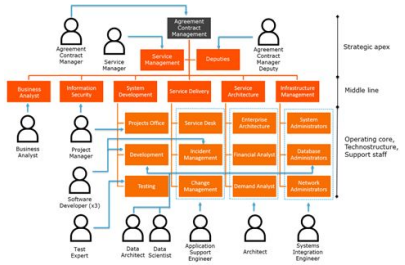
Buyers Group	Space Market	Legal & Notaries
Early Adopters	Audio-visual Market	Bank & Finance
Other Research Institutes	Health & Medical	Telco

- Sales networks defined by GMV and Piql

Governance, risk and compliance model

- Collaborative model
- Risk informed decissions
- Demostrate standards adherence





PEOPLE PROCEDURES & TECHNOLOGIES

GAPS, CONTROLS & MONITORING ACTIVITIES

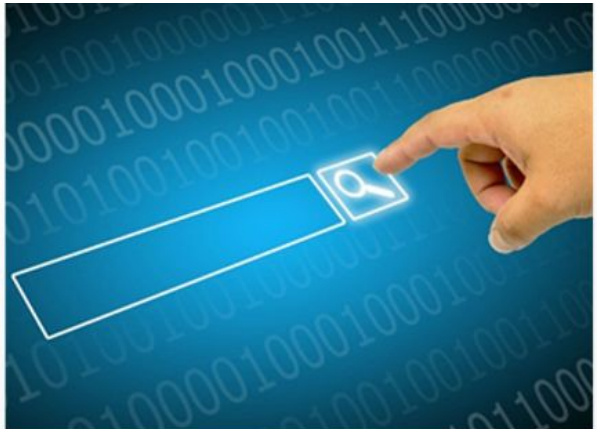
COMPLIANCE

Advanced services

Layer 4 of the Archiver solution covers advanced services oriented to replicate computational experiments and gain deeper insights into archived data. Diverse use cases can be addressed using modern AI techniques.

REPRODUCIBILITY SERVICES

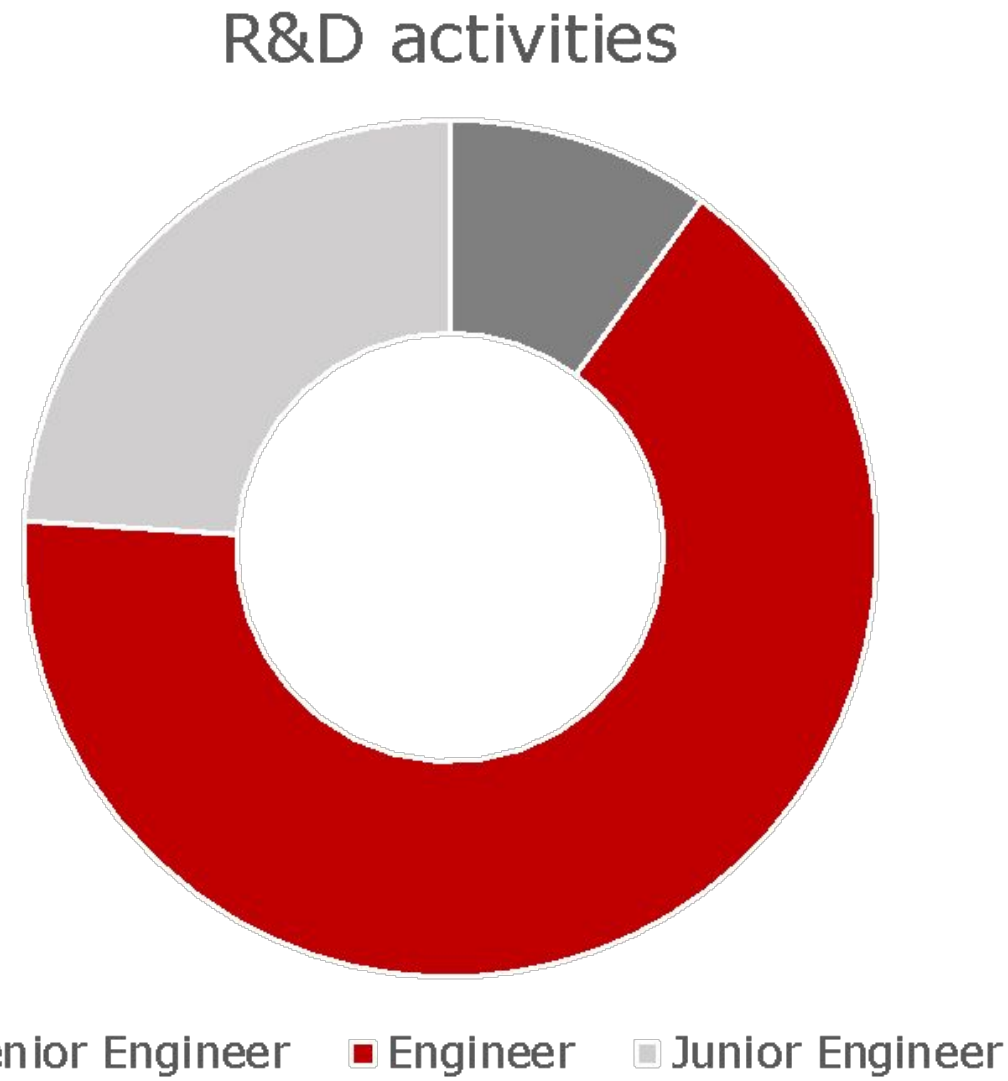
ARTIFICIAL INTELLIGENCE



SEARCH RECOMMENDATION ALGORITHMS

AUTOMATIC DOCUMENT CLASSIFICATION

Qualifications and Experience of Key Personnel



Nuria Gómez Rojo
nngomez@gmv.com

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

Libnova – CSIC – University of Barcelona – Giaretta Associates

libnova

CSIC
Consejo Superior de Investigaciones Científicas

Universitat de Barcelona

Giaretta Associates

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

CONSORTIUM

LIBNOVA – CSIC – UB – Giaretta Associates

libnova

the most advanced digital preservation platform

UNIVERSITAT DE BARCELONA

CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Giaretta Associates

Consortium members:

- LIBNOVA
- Spanish National Research Council
- Universitat de Barcelona
- Giaretta Associates

About the consortium

libnova
the most advanced digital preservation platform

- **LIBNOVA mission is to safeguard the world's research and cultural heritage. Forever.**
- We do that by working to **have the most advanced digital preservation platform.** Year after year, LIBNOVA has been pushing the boundaries of what is possible in digital preservation, researching and incorporating innovations that **empower the organizations to preserve their content in an easier and more efficient way**.
- LIBNOVA was founded in 2009, has offices in the US and Europe and is now present in 12 countries with activity in the **academic, cultural heritage and research communities. LIBNOVA Research Labs** (2017) manages all research initiatives for the company.
- Customers like the *British Library*, *Stanford University*, the *EPFL* and many more already trust us.

About the consortium: LIBNOVA

CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

The Spanish National Research Council is the main agent of the **Spanish System for Science, Technology and Innovation**; and in order to carry out its mission, it has competences aimed at:

- Generation of knowledge through **scientific and technical research**.
- **Transfer of results** from research, especially to boost and create technology-based enterprises.
- **Expert advice** provided to public and private institutions.
- Highly-qualified **pre-doctoral** and **post-doctoral** training.
- Promotion of **scientific culture** in society.
- Management of large facilities and **unique scientific and technical infrastructures**.
- Presence and **representation** in international bodies.
- Development of targeted research.

About the consortium: CISC

UNIVERSITAT DE BARCELONA

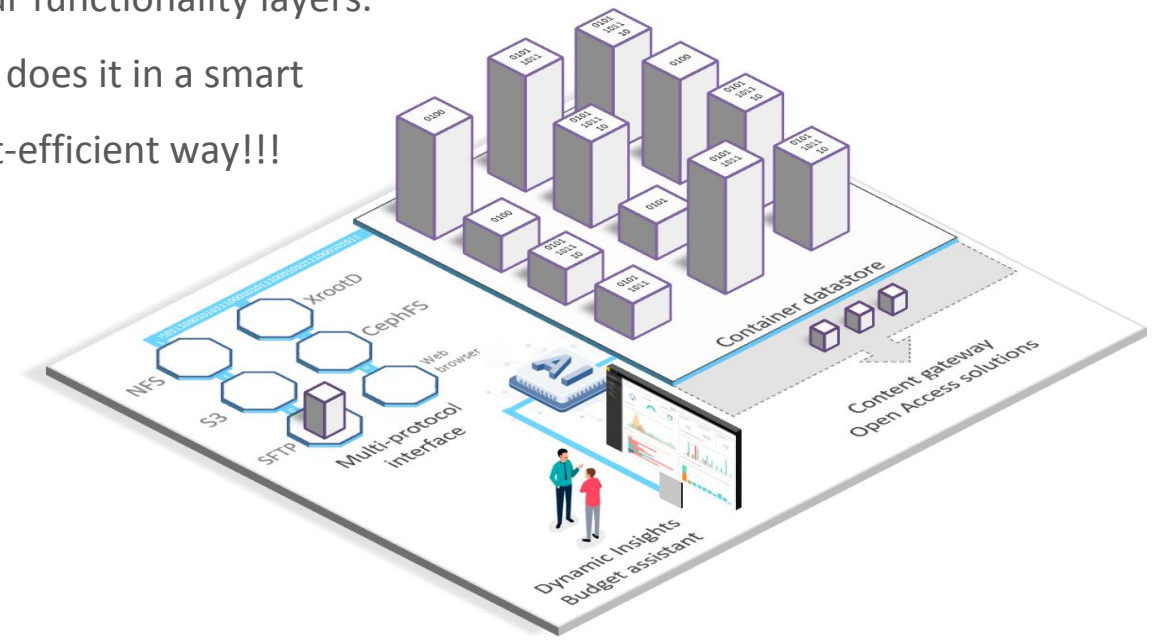
- The University of Barcelona is the **foremost public institution of higher education in Catalonia**, catering to the needs of the greatest number of students and delivering the broadest and most comprehensive offering in higher educational courses.
- The **University of Barcelona is also the principal centre of university research in Spain and has become a European benchmark for research activity**, both in terms of the number of research programmes it conducts and the excellence these have achieved.
- Its own history is closely tied to the history of Barcelona and of Catalonia.
- The university combines the values of tradition with its position as an institution dedicated to **innovation and teaching excellence**: a university that is as outward-looking and cosmopolitan as the city from which it takes its name.

About the consortium: Universitat de Barcelona

- David Giaretta has worked in digital preservation since 1990 and has led many of the most important developments in this area.
- He chaired the panel which produced the OAIS Reference Model (**ISO 14721**), the “de facto” standard for building digital archives, and made fundamental contributions to that standard.
- He leads the group which produced the **ISO standard for audit and certification** of trustworthy digital repositories (**ISO 16363**), and ISO 16919.
- David Giaretta has led a number of **large digital preservation projects**, with funding from the EU and more than 50 partner organisations (*CASPAR*, *PARSE.Insight*, *APARSEN* and *SCIDIP-ES*).
- Involved with the **Alliance for Permanent Access** (APA) from its start to its establishment, he became the Director of the APA in July 2010.

About the consortium: Giaretta Asociattes

- We have been interviewing 50+ research-related organizations in the last years to understand what would be needed to properly preserve their research data in an efficient way.
- The solution we are proposing is built on pre-existing digital preservation platforms already in use by many leading organizations across the world.
- It proposes a solution for the whole organization and for the whole data life-cycle, completely aligned with **OAIS**, **ISO16363**, **FAIR** and **TRUST** principles, with powerful and really innovative capabilities in all four functionality layers.
- …And it does it in a smart and cost-efficient way!!!



About the planned solution

About the planned solution

- We are going to be building a multi-petabyte scale with the **CSIC**'s vast experience on supercomputing and large-scale infrastructures.

- We are going to be making it aligned with the EU legal requirements, GDPR, FAIR principles, TRUST principles and applying really advanced Artificial Intelligence techniques to gain unprecedented efficiency (classification, PII detection, etc) working with the **Universitat de Barcelona**.

- We are going to be making it completely aligned to the OAIS, ISO 16363 and CoreTrustSeal for the most demanding organizations, working with **David Giarietta**.

- And we are going to be building it over **LIBNOVA**'s rock-solid foundation, based on our extensive digital preservation experience and proven solutions.


libnova
the most advanced digital preservation platform


CSIC
Consejo Superior de Investigaciones Científicas


Universitat de Barcelona


Giaretta
Associates

Thank you!

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

RHEA System Spa – DEDAGROUP – GTT

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

ARCHIVER CONSORTIUM

Iolanda Maggio RHEA GROUP

Earth Observation Support Engineer and Long Term Data Preservation expert

Roberta Svanetti DEDAGROUP

Digital Knowledge Life Operations Manager and Long Term Data Preservation expert

OUTLINE

1. Who we are
2. Consortium Relevant Expertise
3. Consortium Solution

CONSORTIUM AND REPRESENTATIVES

RHEA GROUP

- WP: Project Management
- WP: Federated Identity Management
- WP: Help Desk

SUB

gtt

- WP: Hybrid Cloud & Storage capabilities

DEDAGROUP

- WP: Application Layer for Preservation and Exploitation
- WP: CoreTrustSeal and/or ISO 16363 Self Assessment

ABOUT RHEA GROUP

RHEA Group is a privately-owned professional engineering and solutions company, providing bespoke engineering solutions, system development and security services for space, military, government and other critical national infrastructure organizations. Since its creation in 1992, RHEA has built a reputation as a trusted partner, developing tailored solutions that help drive organizational and cultural initiatives, leading to sustainable added value for its customers.

Headquartered in Belgium, RHEA Group counts nearly 500 people and has offices in Belgium, UK, Czech Republic, Italy, France, Germany, Spain, Switzerland, the Netherlands and Canada, and works at clients' premises throughout Europe and North America. RHEA is ISO 9001 and ISO 27001 certified.

RHEA GROUP

SECURITY SERVICES
Governance, physical and cyber security services

SYSTEM ENGINEERING & DIGITAL ENGINEERING
Concurrent Design ; Advancing your complex engineering projects

GROUND SYSTEMS ENGINEERING
Shaping the next generation of European satellite ground segment

PROFESSIONAL ENGINEERING SERVICES
The experts for your engineering programs

EARTH OBSERVATION & DATA ANALYTICS
Solutions to global environmental challenges

AVIATION
Aircraft ground icing solutions

https://www.rheagroup.com/
https://twitter.com/rheagroup
https://www.linkedin.com/company/rheagroup/

ABOUT DEDAGROUP

With revenues of €247 million in 2018, a staff of over 1,700 and more than 3,600 customers, Dedagroup is a major aggregator of Italian excellence in software and solutions as a service and a natural partner to companies, financial institutions and public services in developing their IT and digital strategies. Founded in 2000 and based in Turin, the Group has enjoyed constant growth. In addition to its over 20 offices in Italy, it also operates in Switzerland, France, Germany, the UK, the USA, Mexico and China.

www.dedagroup.it
http://www.linkedin.com/company/dedagroup-spa
https://twitter.com/DEDAGROUP_ICT

Deda.Cloud

Deda.Cloud is the cloud managed service provider to companies and organisations that use innovative technologies to develop products and services and constantly improve their processes.
A division of Dedagroup S.p.A., it specialises in cloud strategy and is organised to work in synergy with the Group's other companies and business units: Dedagroup Business Solutions, Dedagroup Public Services, Dedagroup Stealth, Dedagroup Wiz, Derga Consulting and Piteco.

ABOUT GTT

From financial services trading firms to manufacturers and government, GTT is committed to providing our clients with the services, reach and capabilities that improve communication, productivity and efficiency across their organisations.
GTT connects people across organizations, around the world and to every application in the cloud. Our clients benefit from an outstanding service experience built on our core values of simplicity, speed and agility. GTT owns and operates a global Tier 1 internet network and provides a comprehensive suite of cloud networking services.

GTT offers wide area networking, internet, managed services, transport & infrastructure, and voice, all designed to meet our clients' unique needs. Take advantage of GTT's software-defined wide area networking, a managed service with the broadest range of access options, to gain visibility and control across your WAN. We deliver services in over 100 countries across six continents, ensuring that we are everywhere you do business.

https://www.gtt.net/gb-en/
https://twitter.com/gttcomm
https://www.facebook.com/GTTCommunications/
https://www.linkedin.com/company/gtt

OUTLINE

1. Who we are
2. Consortium Relevant Expertise
3. Consortium Solution

RHEA GROUP – RELEVANT EXPERIENCES

- PCP contract (HNSciCloud) and H2020 projects (OCRE) experience;
- Providing the European research community with access to commercial (IaaS, SaaS and PaaS) and Earth Observation (EO) digital services (OCRE project);
- Promoting the use of European resource and platform services to facilitate a simplified and efficient exploitation of EO data in cloud environments (EO Network of Resources initiative);
- Responsible for the Preservation service of EO datasets of ESA missions together with the evolution of standards and practices. The platform used for Long Term Data Preservation implements all OAIS reference model;
- Responsible for data stewardship (i.e. preservation, discovery, access and exploitation) of space science heritage data for an unlimited time;
- Responsible for ESA services including dataset and information preservation, Persistent Identifier assignment, Heritage Software exploitation, Metadata management (OGC standard, PREMIS and Dublin Core), Provenance and context management and a Data Management and Stewardship Maturity Matrix self-assessment. Best practices and guidelines on preservation are produced and maintained up-to-date.

DEDAGROUP – RELEVANT EXPERIENCES

- Digital Preservation System of the Historical Archives of the European Union;
- Service model transformation (from non-digital flows) into a Digital Archive Preservation OAIS service, in alignment with the OAIS functional model (ISO14721) and adhering to the metrics established by ISO16363;
- IT strategy for Long Term Data Preservation, preservation planning and data stewardship policies, procedures and processes;
- Checklist, use cases and functional technical definition, design and testing;
- API integration with catalogs of the Historical Archive (SIP automatic transfer and ingestion, DIP publication and access);
- Storage Management technologies integration with Data Archive and Preservation Open Sources platform (e.g. Archivematica, AtoM, MuleSB, JASPERsoft, …);
- Digital transmission of the Library service model, in a Cloud Management and digital resource archive service;
- Integration of IIIF open-source APIs for interoperability and image visualization in a scalable Cloud Digital Asset Management.

GTT – RELEVANT EXPERIENCE

- Provision of storage and archive of Satellite Multi-Mission Data, based on dedicated fully redundant storage clusters;
- Provision of a dedicated and Secure Network with Firewall and 24/7 Operation with NOC and SOC plus a dedicated private Cloud Infrastructure for storage up to 8 PB of satellite data to be made available to the science community in a simple, fast and secure way;
- Provision of disaster recovery, second copy of their primary site, offsite backup;
- Deployment of tape storage infrastructures for long term preservation;
- Provision of customised policies and setup to allow customer to meet their business requirements;
- Building of Hybrid Cloud service to serve a multinational client base and integrate with existing networks and services to provide best-in-class performance for users and customers;
- Management of several Data Centres in different EU countries able to host the infrastructure. One of them is located in Geneva and already connected via direct fibre to CERN.

OUTLINE

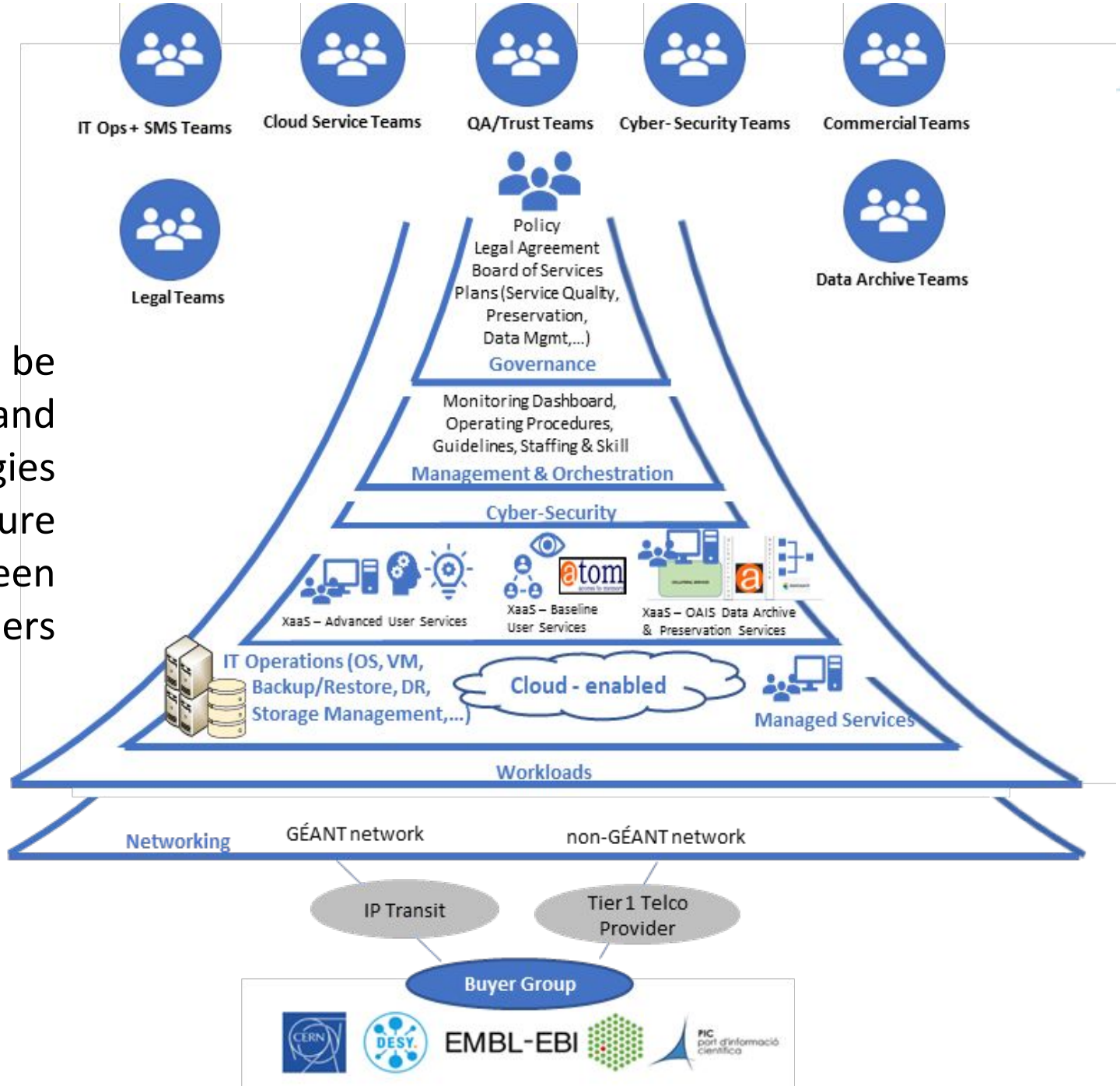
1. Who we are
2. Consortium Relevant Expertise
3. Consortium Solution

SOLUTION PILLARS

- FAIR Principles
- Relevant Standards, Guidelines and Regulations (OAIS-ISO 14721:2003, ISO16363, PREMIS, ISAD-G, ISAAR-CPF, EAD, METS, GDPR, …)
- Information Governance (policies, procedures and processes, data management, preservation, business continuity and service quality plans, risk management, …)
- Open Source (Dedicated Community, Scientific Scenarios, Brand Independent, …)
- Dedicated Hybrid Cloud (Buyers Use Cases DRIVEN)

PROPOSED SOLUTION

The proposed architecture will be based on open standards and robust and scalable technologies (the baseline), enabling a secure and efficient interaction between data producers, service providers and service consumers.



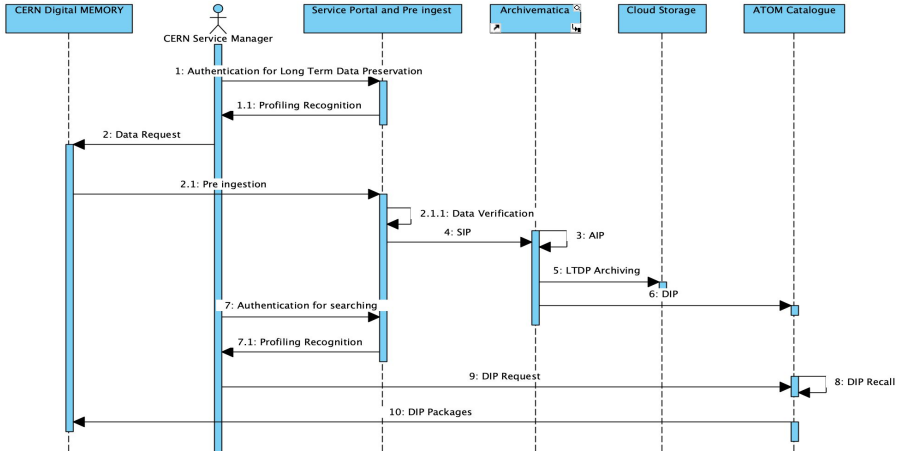
ARCHITECTURE COMPONENTS

The architecture components of the solution are the following:

- Secure Service Portal (Identity Access Management, Access Layer Interface, Validation and Pre-ingestion services);
- Existing and mature Open Source platforms for data archiving, preservation, reporting and access/discovery (Archivematica/AToM/JasperSoft);
- Readiness and large-scale XaaS services;
- Cloud connect product for integration with proposed robust and scalable managed Hybrid Cloud.

Governance Board for ensuring and monitoring that processes and services are aligned with data management, preservation and business continuity plans for granting data integrity.

ARCHIVER PROJECT STARTING POINT



			CERN Digital Memory
Id BURS	Title	Subtitle	Business User Requirement Description
BRCEDMINT001	Digital object formats	Submission agreement	We need to archive the CERN Digital Memory which consists of the digital production of the institution for the 21st century (including new types like web sites, social media, emails, etc) as well as the analog documents produced by the institution in the 20th century, composed of digitized papers (physical archive) and various multimedia: audio (e.g. recordings of meetings), still images and moving images.
BRCEDMINT002	Producer archive integration	Pre-ingest	The digitized institutional content is loaded and maintained into CERN Live information systems. These information systems use various underlying storage solutions (e.g. systems like DFS,EOS etc.) but none of them is OAIS compliant. The goal is to connect the active services with a dark archive where Archival Information Packages will keep comprehensive information for each 'document'. The goal is to have at our disposal at the end a standard trustworthy ISO16363-compliant digital archive where live systems can deposit content selected for long term preservation.
BRCEDMGEN001	Registry of formats	Registry of formats	By ensuring that data are kept in the same archiving solution, we can introduce specific standards and formats for long term preservation and, as a consequence, minimize fragility which is one of the highest risks for long term preservation in the organization. We want to ensure that we keep a globally accepted preservation processes and standard formats for the same type of content (e.g. video).
BRCEDMGEN002	Registry of formats	OAIS alignement	Aligning the CERN digital archive to best practices (OAIS) for the sake of long term digital preservation is one of the most important benefits for this use case. In addition to this, the existence of successful disaster recovery solution institution-wide could impact all individuals, as everyone produces the same
BRCEDMWKF001	Lifecycle - WF	Authentication to Archiving Service	The authentication needs for the basic use case are minimal, as only the Service Manager will need to access the Archiving Service and not the end user.
BRCEDMWKF002	Lifecycle - WF	Partial access to data	One important aspect of the use case is the ability to have partial access to the data, i.e. to recall just one file or even a chunk of a file out thousands of files in a specific dataset.
BRCEDMWKF003	Lifecycle - WF	Re-ingest	The system manager should be able not only to submit and download data from the service but also to update and re-ingest specific parts

Principle	Code	Indicator	Priority
Findable	F1	Metadata is identified by a persistent identifier	Essential
Findable	F1	Data is identified by a persistent identifier	Essential
Findable	F1	Metadata is identified by a globally unique identifier	Essential
Findable	F1	Data is identified by a globally unique identifier	Essential
Findable	F2	Rich metadata is provided to allow discovery	Essential
Findable	F3	Metadata includes the identifier for the data	Essential
Findable	F4	Metadata is offered in such a way that it can be harvested and indexed	Essential
Principle	**Code**	**Indicator**	**Priority**
Accessible	A1	Metadata contains information to enable the user to get access to the data	Important
Accessible	A1	Metadata can be accessed manually (i.e. with human intervention)	Essential
Accessible	A1	Data can be accessed manually (i.e. with human intervention)	Essential
Accessible	A1	Metadata identifier resolves to a metadata record	Essential
Accessible	A1	Data identifier resolves to a digital object	Essential
Accessible	A1	Metadata is accessed through standardised protocol	Essential
Accessible	A1	Data is accessible through standardised protocol	Essential
Accessible	A1	Data can be accessed automatically (i.e. by a computer program)	Important
Accessible	A1.1	Metadata is accessible through a free access protocol	Essential
Accessible	A1.1	Data is accessible through a free access protocol	Important
Accessible	A2	Metadata is guaranteed to remain available after data is no longer available	Essential
Principle	**Code**	**Indicator**	**Priority**
Interoperable	I1	Metadata uses knowledge representation expressed in standardised format	Important
Interoperable	I1	Data uses knowledge representation expressed in standardised format	Important
Interoperable	I1	Metadata uses machine-understandable knowledge representation	Important
Interoperable	I1	Data uses machine-understandable knowledge representation	Important
Interoperable	I2	Metadata uses FAIR-compliant vocabularies	Important
Interoperable	I3	Metadata includes references to other metadata	Important
Interoperable	I3	Metadata includes qualified references to other metadata	Important
Principle	**Code**	**Indicator**	**Priority**
Reusable	R1	Plurality of accurate and relevant attributes are provided to allow reuse	Essential
Reusable	R1.1	Metadata includes information about the licence under which the data can be reused	Essential
Reusable	R1.1	Metadata refers to a standard reuse licence	Important
Reusable	R1.1	Metadata refers to a machine-understandable reuse licence	Important
Reusable	R1.2	Metadata includes provenance information according to community-specific standards	Important
Reusable	R1.3	Metadata complies with a community standard	Essential
Reusable	R1.3	Data complies with a community standard	Essential
Reusable	R1.3	Metadata is expressed in compliance with a machine-understandable community stand	Essential
Reusable	R1.3	Data is expressed in compliance with a machine-understandable community standard	Important

www.rheagroup.com

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

T-Systems International – GWDG – Onedata

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

Archiving and Preserving to discover

vision, TEAM and approach

Jurry de la Mar

T · · Systems ·

Archiving and Preservation for research

Vision T-Systems

“we mobilize more know-how and create more discovery in Research by democratizing access to professional archiving and preservation for the cost of storing the information.”

Team T-Systems.

Innovate and Showing the WAY

T-Systems Team of Experts

Jurry de la Mar

Science and Research Expert,

Prof. Dr. Philipp Wieder

Research Data and Preservation Expert,

Lukasz Dutka

Research Data Expert, Onedat

Matthias Pink

Cloud Expert, T-Systems

Bartosz Kryza

Distributed Data Expert,

Prof. Dr. Ramin Yahyapour

GWDG

Archiver - starting point.

experience

HELIX NEBULA THESCIENCECLOUD

SENDATE

MIKELANGELO

INDIGO - DataCloud Better Software for Better Science

Various FP7 and H2020 projects. Founding member of Helix Nebula. HPC and Cloud provider for Science.

STATE-OF-THE-ART.

community

MAX-PLANCK-GESELLSCHAFT

EUROPEAN OPEN SCIENCE CLOUD

DFG Deutsche Forschungsgemeinschaft German Research Foundation

GAIA-X

openstack™

bitkom

Established partner in leading and thriving communities to collaborate with users and developers.

STRONGEST TEAM.

public cloud

Open Telekom Cloud is a leading public cloud service from Germany, scalable, secure and cost-effective.

SIMPLE, SECURE, AFFORDABLE.

networks

T · · GÉANT

Registered GÉANT IaaS provider and established nx10G network peering.

BEST ACCESS.

Computing

- **Elastic Cloud Server (10)**
  Elastic, scalable computing servers
- **Bare Metal Server (0)**
  Provides dedicated physical servers for tenants
- **Image Management Service (2)**
  Self-service image management
- **Cloud Container Engine (0)**
  container service that features high availability and elastic scalability
- **Auto Scaling (0)**
  Dynamically adjusts computing resources
- **Dedicated Host (0)**
  Dedicated physical servers

Database

- **Relational Database Service (0)**
  Highly reliable relational database service
- **Distributed Cache Service (0)**
  Provides secure, convenient, and high-speed cache service

Storage

- **Elastic Volume Service (12)**
  Elastic, scalable block storage
- **Cloud Server Backup Service (1)**
  Secure, reliable cloud server backup
- **Storage Disaster Recovery Service (0)**
  Storage disaster recovery service
- **Volume Backup Service (1)**
  Secure, reliable block storage backup
- **Object Storage Service**
  Scalable cloud storage
- **Scalable File Service (0)**
  Elastic, scalable file storage

Application

NEW Archiving and Preservation

- **Simple Message Notification**
  Provides simple and reliable message notification service
- **Software Repository for Container (0)**

Network

- **Virtual Private Cloud (3)**
  Provides securely isolated virtual networks
- **Elastic Load Balancing (0)**
  Distributes traffic across multiple ECSs
- **Direct Connect (0)**
  Provides high-speed, stable network access services
- **Private Link Access Service (0)**
  High-quality, secure and dedicated network access service
- **Domain Name Service**
  Stable, secure, fast domain name resolution
- **NAT Gateway (0)**
  provides source NAT service
- **Virtual Private Network (0)**
  Enables remote secure access to VPC networks
- **CDN (Akamai)**
  Easy-to-use, reliable, quick content distribution
- **Elastic IP (10)**
  Flexible public network access

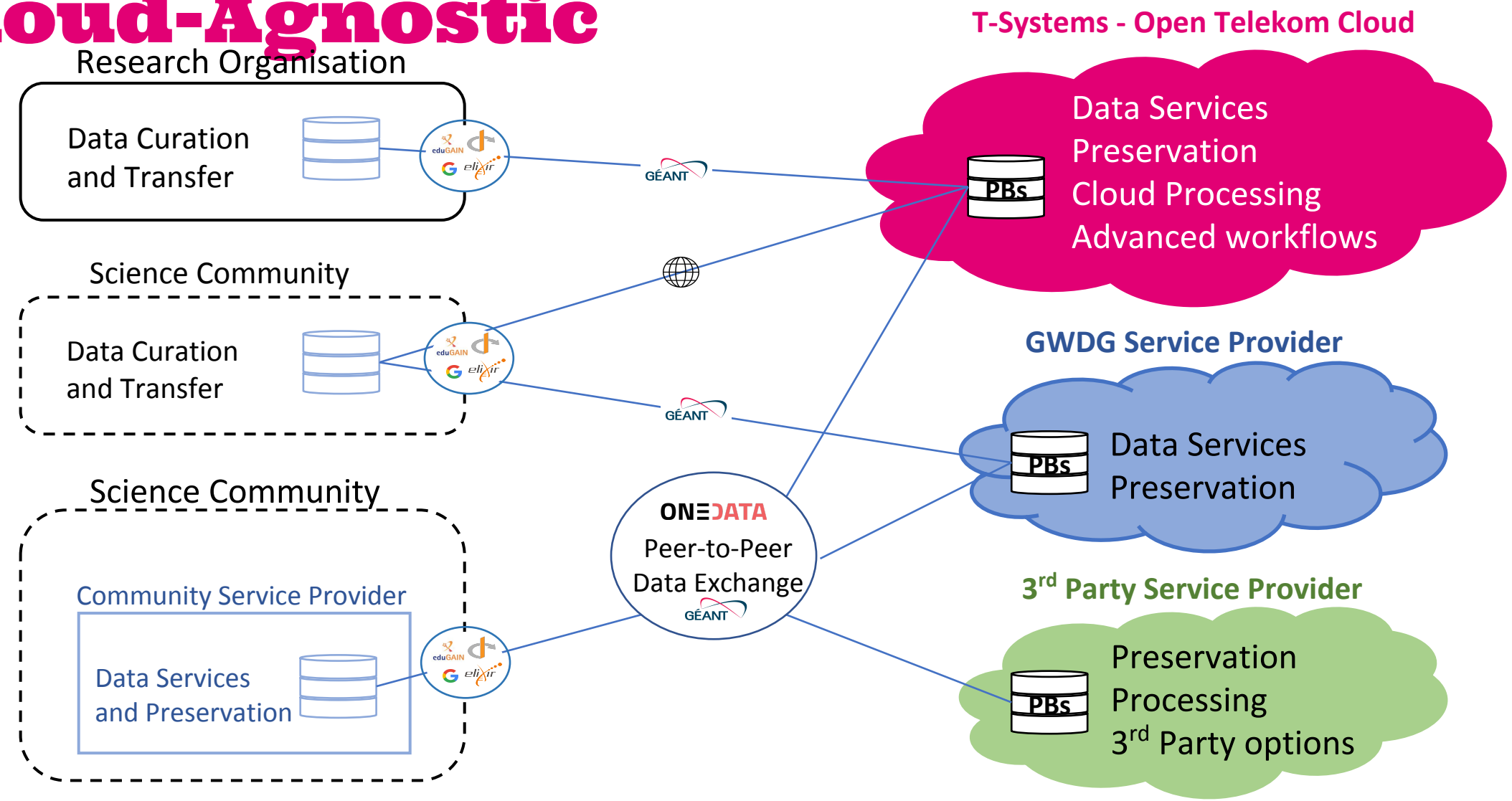
Security

- **Anti-DDoS**
  Provides Anti-DDoS protection
- **Web Application Firewall**
  Filters malicious web traffic
- **Key Management Service (0)**
  Easily manage the keys used to encrypt your data

Management & Deployment

- **Cloud Eye**
  Resource monitoring and alarm notification
- **Identity and Access Management**
  Manages user access and encryption keys
- **Resource Template Service**
  Provides orchestration for resources
- **Cloud Trace Service**
  Records operations performed on cloud rescources
- **Log Tank Service**
  Log collection, query, and storage
- **Tag Management Service**

The Approach: OPEN-Source and Cloud-Agnostic



THANK YOU

LIFE IS FOR SHARING.

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

Feedback Session

Marion Devouassoux
Project Analyst (CERN)

ARCHIVER - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

Questions

1. What is your role in this award ceremony ?
2. Did the event meet your expectations ?
3. This award ceremony helped me better understand the project. Do you agree ?
4. Did you receive sufficient information on the selected consortia's planned solutions ?
5. Do you find the Early Adopters Program interesting ?

Go to menti.com

- Grab your phone or open a new window
- Go to www.menti.com