Identity Management for Research Collaborations: *from Pilots to Production*

Bob Jones
IT dept
CERN

VAMP

Vo Architectural Middleware Planning

WORKSHOP

30 Sept & 1 Oct 2013
Helsinki, Finland

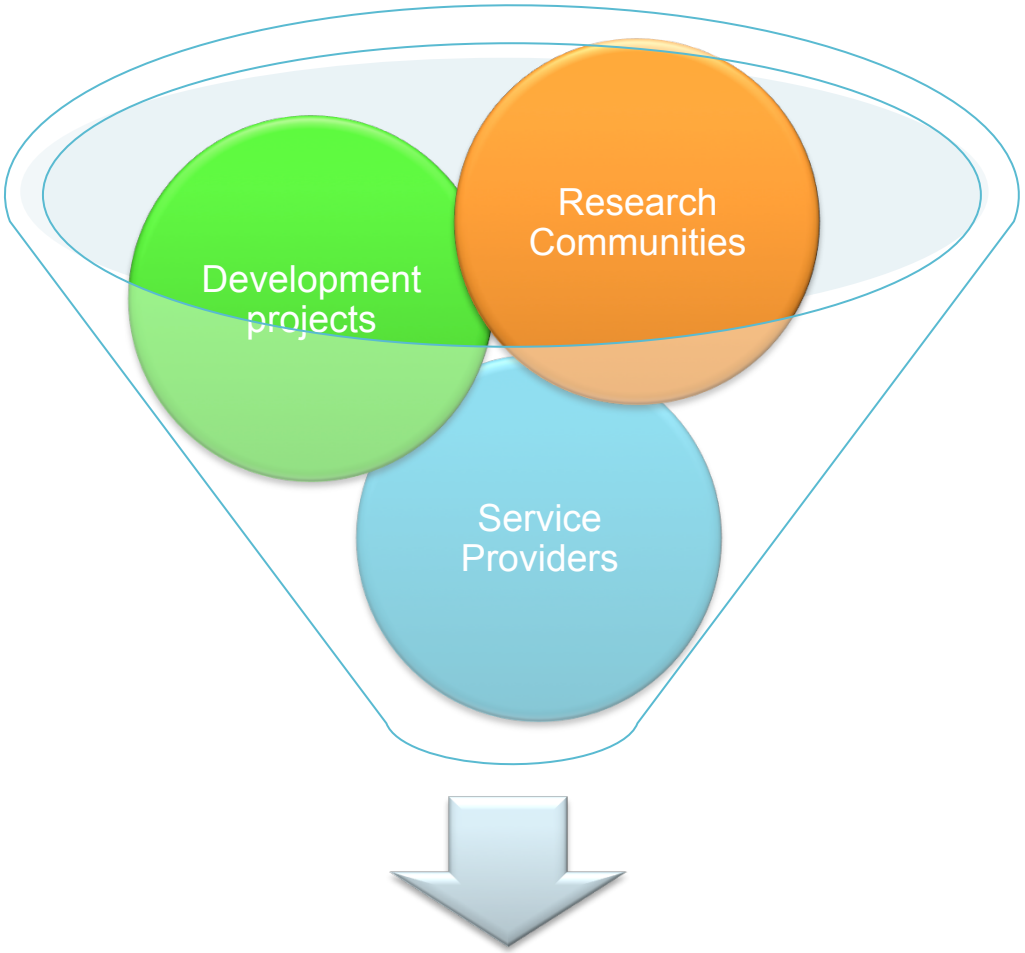
TERENA

CSC

SURF NET

CERN

Outline



Production Identity Management Services

Bob Jones (CERN) – October 2013


CERN
ILL
EFDA
E.
XFEL
EIROforum
EMBL
ESRF
ESA
ESO

A Vision for a European e-Infrastructure for the 21st Century

Sustainable - RIs currently in construction (FAIR, XFEL, ELIXIR, EPOS, ESS, SKA, ITER and upgrades to ILL and ESRF etc.), need to be convinced that e-Infrastructure will exist and continue to evolve throughout their construction and operation phases if they are to take the risk and invest in its creation & exploitation

Inclusive - Need an e-Infrastructure that supports the needs of the whole European research community, including the *"long tail of science",* and interoperate with other regions

Flexible - Cannot be a one-size-fits-all solution

Integrated - Coherent set of services and tools must be available to meet the specific needs of each community

Innovative - Essential that European industry engages with the scientific community to build and provide such services

User driven - The user community should have a strong voice in the governance of the e-Infrastructure

See https://cds.cern.ch/record/1550136/files/CERN-OPEN-2013-018.pdf

What do we have already?

- Existing European e-infrastructure *long-term* projects
  - GEANT, EGI, PRACE
- Many “pathfinder” initiatives have prototyped aspects of what will be needed in the future
  - Includes much of the work in the existing e-Infrastructure projects but also projects such as EUDAT, Helix Nebula, OpenAIRE+, etc
  - Thematic projects such as BioMedBridges/ CRISP/ DASISH/ ENVRI, as well as Transplant, VERCE, Genesi-DEC and many others

How can we create e-infrastructures that overcome fragmentation?

- Fragmentation of users (big science vs. long tail)
- Fragmentation of infrastructure (not integrated services)
- Common platform (*e-infrastructure commons*) with 3 integrated areas
  - **International network, authorization & authentication, persistent digital identifiers**
  - **small number of facilities to provide cloud and data services of general and widespread usage**
  - **Software services and tools to provide value-added abilities to the research communities, in a managed repository**
- Need a *data continuum* - linking the different stages of the data lifecycle, from raw data to publication, and compute services to process this data

The Business of Research

- Publicly funded research communities make significant investments in E-infrastructure that must be justified
  - To justify these investments the e-infrastructures must show a clear impact for the research communities
  - To gauge the impact, this market of end-users must be well understood by funding agencies and e-infrastructure services providers
- So the user communities must have a strong voice in the governance of the e-infrastructures to ensure they remain relevant and upto-date

User Forum

- A pan-European forum for organisations and projects that operate at an international level
- Present to the policy makers and the infrastructure providers where there are common needs and opinions and where there is divergence
- Independent of any supplier and engage across research domains
- Supplements but does not replace existing e-infrastructure user engagement channel

See https://cds.cern.ch/record/1545615/files/CERN-OPEN-2013-017.pdf

Inspired by structure and results of FIM$_4$R

Federated Identity Management for Research Collaborations

Paper Type: Research paper

Date of this version: 28 August 2013

- **Requirements from the research communities**
- **Status of the activities & use cases**
- **Common vision across these communities**
- **Key stages of a roadmap**
- **Set of recommendations**

Abs

Feder lets
subsc the
group ers.
For ex n of
resour

A nur of a
delug ross
organisational and national boundaries.

Driven by these needs, representatives from a variety of research communities, including photon/neutron facilities, social science & humanities, high-energy physics, atmospheric science, bioinformatics and fusion energy, have come together to discuss how to address these issues with the objective to define a common policy and trust framework for Identity Management based on existing structures, federations and technologies.

This paper will describe the needs of the research communities, the status of the activities in the FIM domain and highlight specific use cases. The common vision for FIM across these communities will be presented as well the key stages of the roadmap and a set of recommendations intended to ensure its implementation.

Keywords
federated identity management, security, authentication, authorization, collaboration, community

CERN-OPEN-2012-006
28/08/2013

Authors: Daan Broeder, Bob Jones, David Kelsey, Philip Kershaw, Stefan Lüders, Andrew Lyall, Tommi Nyrönen, Romain Wartel, Heinz J Weyer

The FIM_4R Vision

A common policy and trust framework for Identity Management based on existing structures and federations either presently in use by or available to the communities.

This framework must provide researchers with unique electronic identities authenticated in multiple administrative domains and across national boundaries that can be used together with community defined attributes to authorize access to digital resources.

FIM_4R workshop here tomorrow

Prioritisation of FIM$_4$R requirements

- **User friendliness** (high)
  - Support for citizen scientists and researchers without formal association to research labs or univ
- **Browser & non-browser federated access** (high)
- **Bridging communities** (medium)
  - Bridging is a central issue with an efficient mapping of the respective attributes
- **Multiple technologies with translators including dynamic issue of credentials** (medium)
- **Implementations based on open stds and sustainable with compatible licenses** (high)
- **Different Levels of Assurance with provenance** (high)
  - Credentials need to include the provenance of the level under which it was issued
- **Authorisation under community and/or facility control** (high)
- **Well defined semantically harmonised attributes** (medium)
- **Flexible and scalable IdP attribute release policy** (medium)
  - Bi-lateral negotiations between all SPs and all IdPs is not a scalable solution
- **Attributes must be able to cross national borders** (high)
  - Data protection considerations must allow this to happen.
- **Attribute aggregation for authorisation** (medium)
  - Attributes need to be aggregated from different sources of authority including federated IdPs and community-based attribute authorities.
- **Privacy and data protection** addressed with community-wide individual ids (medium)

Technologies being piloted by resource communities

	Technology	Inter-federation	Status	Plans
ELIXIR	SAML2	Joined to the HAKA federation.	Development	Join eduGAIN and Kalmar union.
ESA	Shibboleth		Production	Join NASA and EUMETSAT
WLCG	Web based and non-web based. Pilot project on non-web based. SAML2.		CLI solution achieved, but looking into alternatives.	Refocus on web-based use case.
DARIAH	Shibboleth		Development	Implement DARIAH-EU
SWITCH	Shibboleth	Open to inter-federation with eduGAIN.	Production (for the majority of Swiss Universities)	Moonshot and Interfederation with eduGAIN
CLARIN	Shibboleth		Implementation	Interconnection through Service Provider Federation (SPF)
Umbrella	Shibboleth	Bridging concept being developed.	Implementation	Affiliation Database, Sync with other programs iCAT, Moonshot. Bridging. Implementation up to Sept. 2013.
WeNMR	Drupal based WeNMR VRC, Shibboleth authN, phpCAS authN, robot certificates.	Flexibility to connect to a wide set of federations (from Drupal).	Production	
SyBIT	Azure	Any federation could be easily integrated.	Microsoft product.	Self-federation model.

ESFRI Cluster projects

ENVRI

Common Operations of Environmental Research Infrastructures

32 Research Infrastructures

CRISP

The Cluster of Research Infrastructures for Synergies in Physics

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Cross-Disciplinary Challenges

A matrix showing the interest in common topics for the four cluster initiatives

	CRISP	ENVRI	DASISH	BioMed
Data identity	■	■	■	■
Data identity continuum	■	■	■	■
Software identity	■	■	■	■
Concept identity		■	■	■
User identity management	■	■	■	■
Common data standards and formats	■	■	■	■
Service discovery	■	■	■	■
Service market places	■	■	■	■
Integrated data access and discovery	■	■	■	■
Data storage facilities	■	■	■	■
Data curation	■	■	■	■
Privacy and security	■	■	■	■
Volatile data management		■	■	
User Community Body	■	■	■	■
Semantic annotations and bridging		■	■	■
Reference models	■	■	■	■
Education & training	■	■	■	■

Research Infrastructures need a **service**

- **Risk Analysis -** *implications of having a malicious SP in a federation*
- **Traceability -** *identifying the cause of any security incident*
- **Security Incident Response –** *including all IdPs and SPs*
- **Transparency -** *essential to gain the trust of the users and service providers*
- **Reliability and Resilience -** *of the framework services*
- **Smooth Transition -** *of the existing production systems to a federated identity management model*
- **Easy integration with local SP environment -** *SPs are likely to want to support multiple means of authentication*
- **Specific requirements -** *from some communities*

A European cloud computing partnership: big science teams up with big business

HELIX NEBULA THE SCIENCE CLOUD

Strategic Plan

- Establish a federated multi-tenant, multi-provider cloud infrastructure
- Identify and adopt policies for trust, security and privacy
- Create governance structure
- Define funding schemes

CERN

To support the computing capacity needs for the ATLAS experiment

EMBL

Setting up a new service to simplify analysis of large genomes, for a deeper insight into evolution and biodiversity

esa

DLR cnes

To create an Earth Observation platform, focusing on earthquake and volcano research

Atos CloudSigma CSA cloud security alliance DANTE www.dante.net DLR egi interoute from the ground to the cloud

logica be brilliant together SAP sixsq SWITCH T-Systems terradue 2.0 the SERVER LABS the IT architects

Adopters

AWST Capgemini CONSULTING. TECHNOLOGY. OUTSOURCING cnrs ECMWF Ifremer OpenNebula.org The Open Source Toolkit for Cloud Computing

orange Business Services THALES Telefónica Trust-IT Communicating ICT to Markets www.trust-itservices.com

Timeline

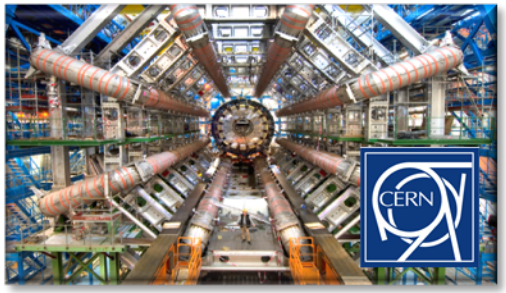
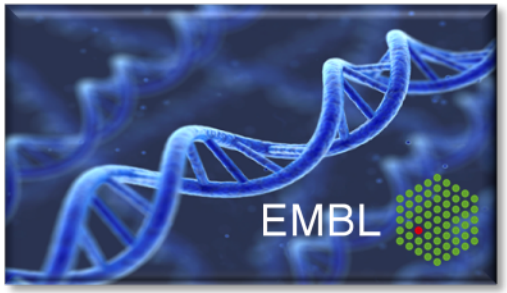
- Endorse the Common **Strategy**
- Agree on the **Partnership**
- Select **flagships** use cases
- Define **governance** model

- **Pilot** Phase
- **Deploy** flagships,
- **Analysis** of functionality, performance & financial model

Towards an **open market for Science**

See A Catalyst for Change:
https://cds.cern.ch/record/1537032/files/HelixNebula-NOTE-2013-003.pdf

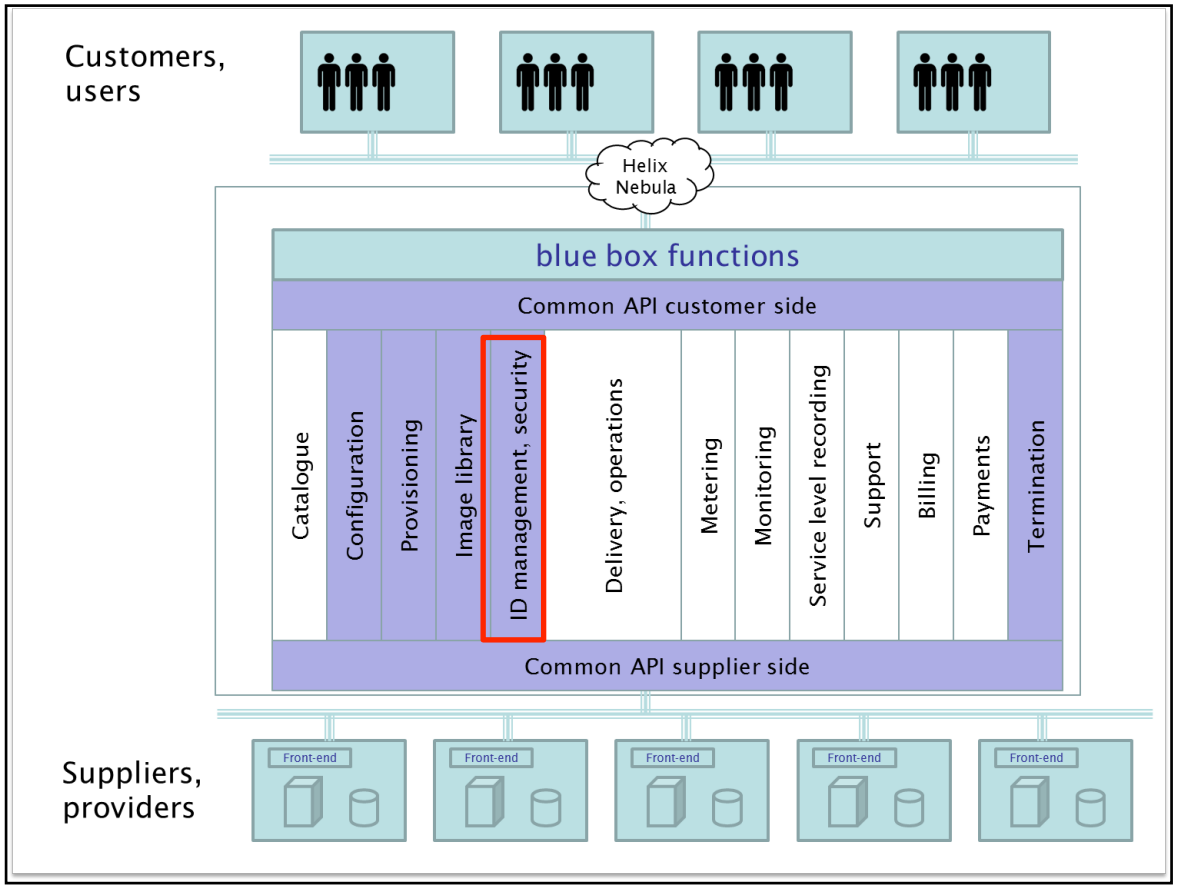
Initial Flagship Use Cases

ATLAS High Energy Physics Cloud Use	Genomic Assembly in the Cloud	SuperSites Exploitation Platform
		
To support the computing capacity needs for the ATLAS experiment	A new service to simplify large scale genome analysis; for a deeper insight into evolution and biodiversity	To create an Earth Observation platform, focusing on earthquake and volcano research

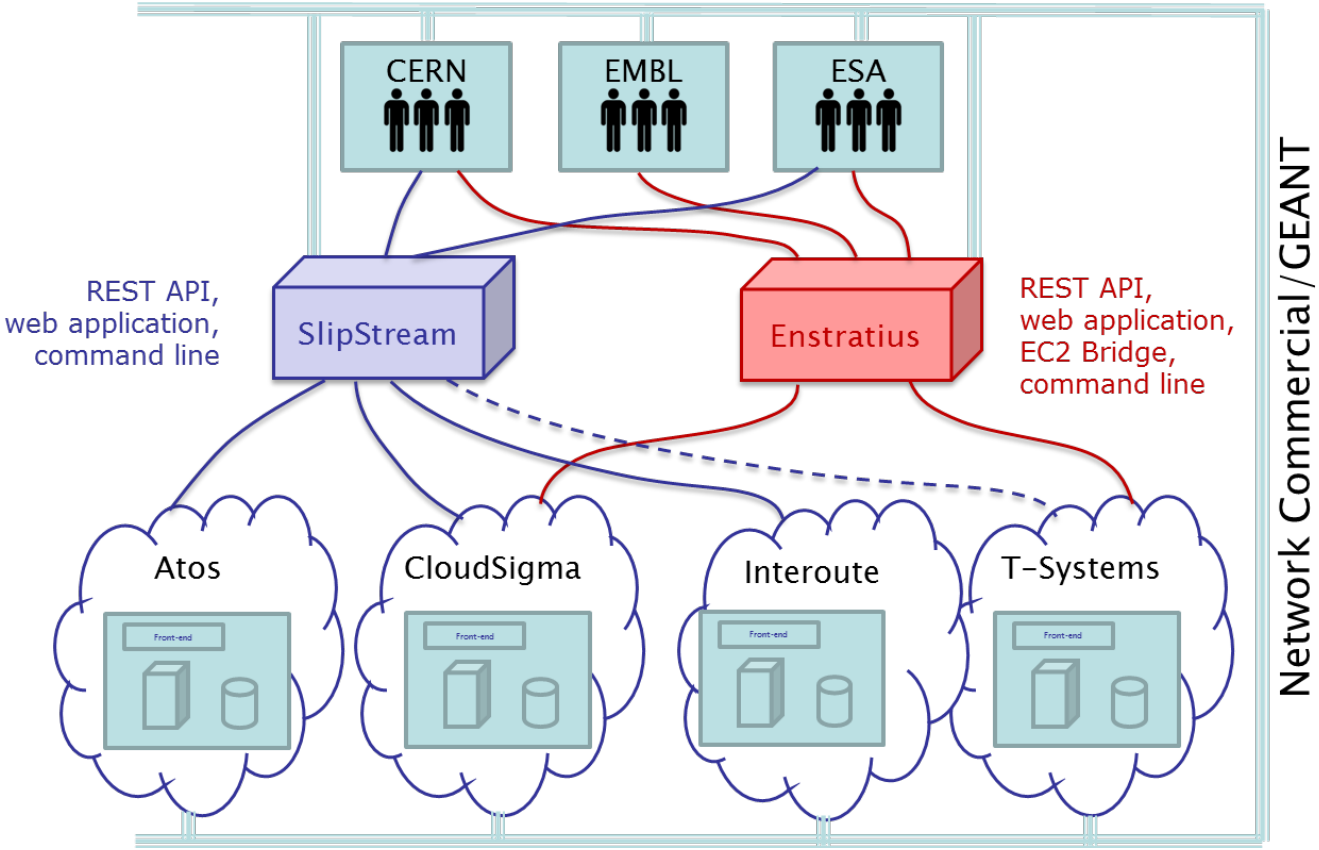
- Scientific challenges with societal impact
- Sponsored by user organisations
- Stretch what is possible with the cloud today

Blue Box brokerage functions

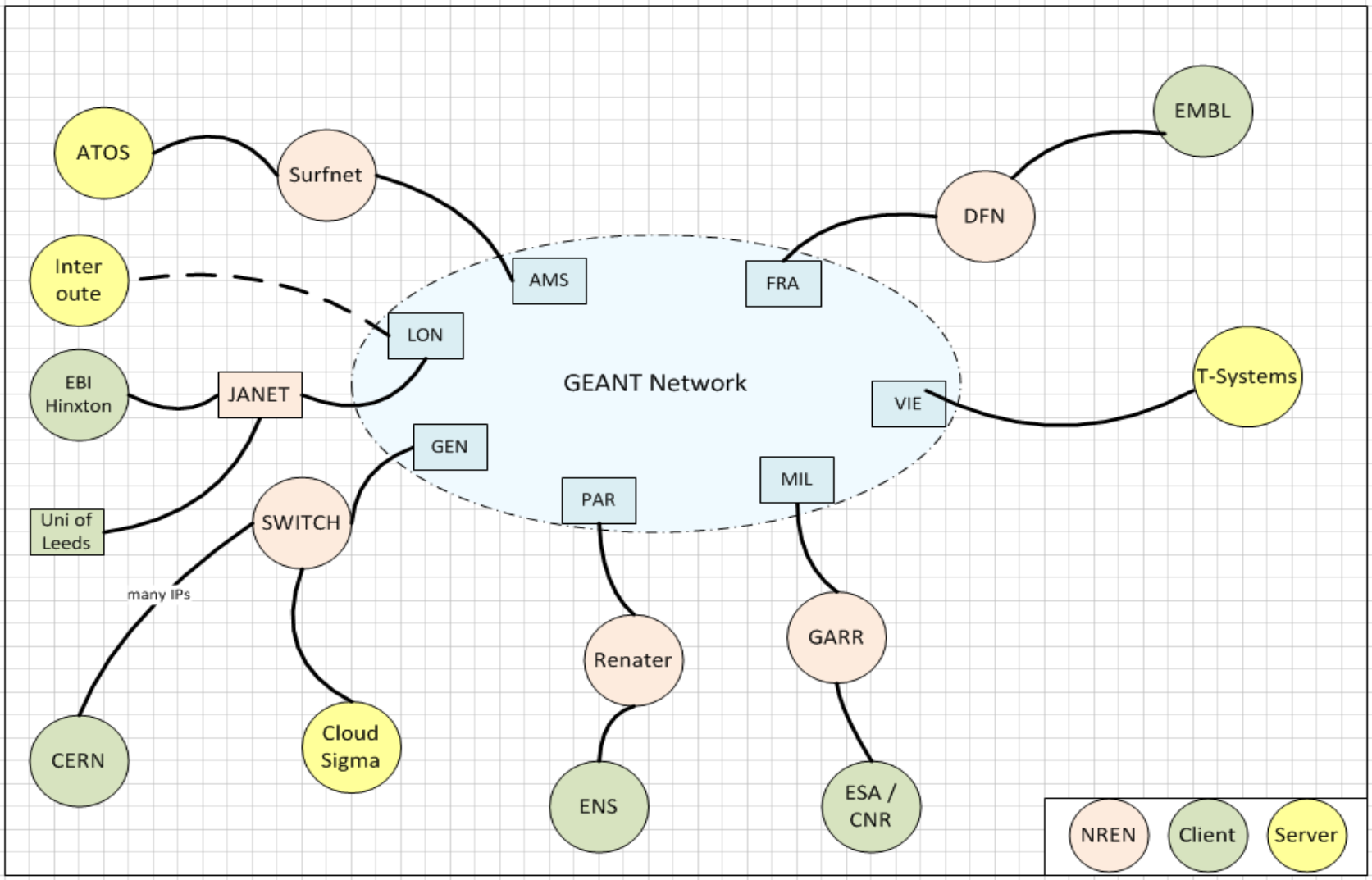
Each customer and supplier have a single connection to the Blue Box resulting in M + N relationships



Blue Box Broker use in pilot phase



Topology


ATOS
Surfnet
EMBL
DFN
Inter oute
AMS
FRA
LON
GEANT Network
EBI Hinxton
JANET
VIE
T-Systems
GEN
MIL
PAR
Uni of Leeds
SWITCH
many IPs
Renater
GARR
CERN
Cloud Sigma
ENS
ESA / CNR
NREN
Client
Server

Building the hybrid cloud

Testing the public-commercial cloud interoperability

- Deployed the ESA/CNES/DLR SuperSites Exploitation Platform on the EGI Fed Cloud
- Will deploy the CERN CMS/ ATLAS flagship use case across commercial suppliers and EGI Federated Cloud via a Blue Box broker

EGI Federated Cloud

Task Force

- Launched in Sep 2011
- 70 members from 40 institutions and 13 countries

Pre-production test-bed:

- 14 resource centres actively providing resources (900 cores, 16 TB storage)
- 30 active users from structural biology, linguistics, ecology, space science, software engineering

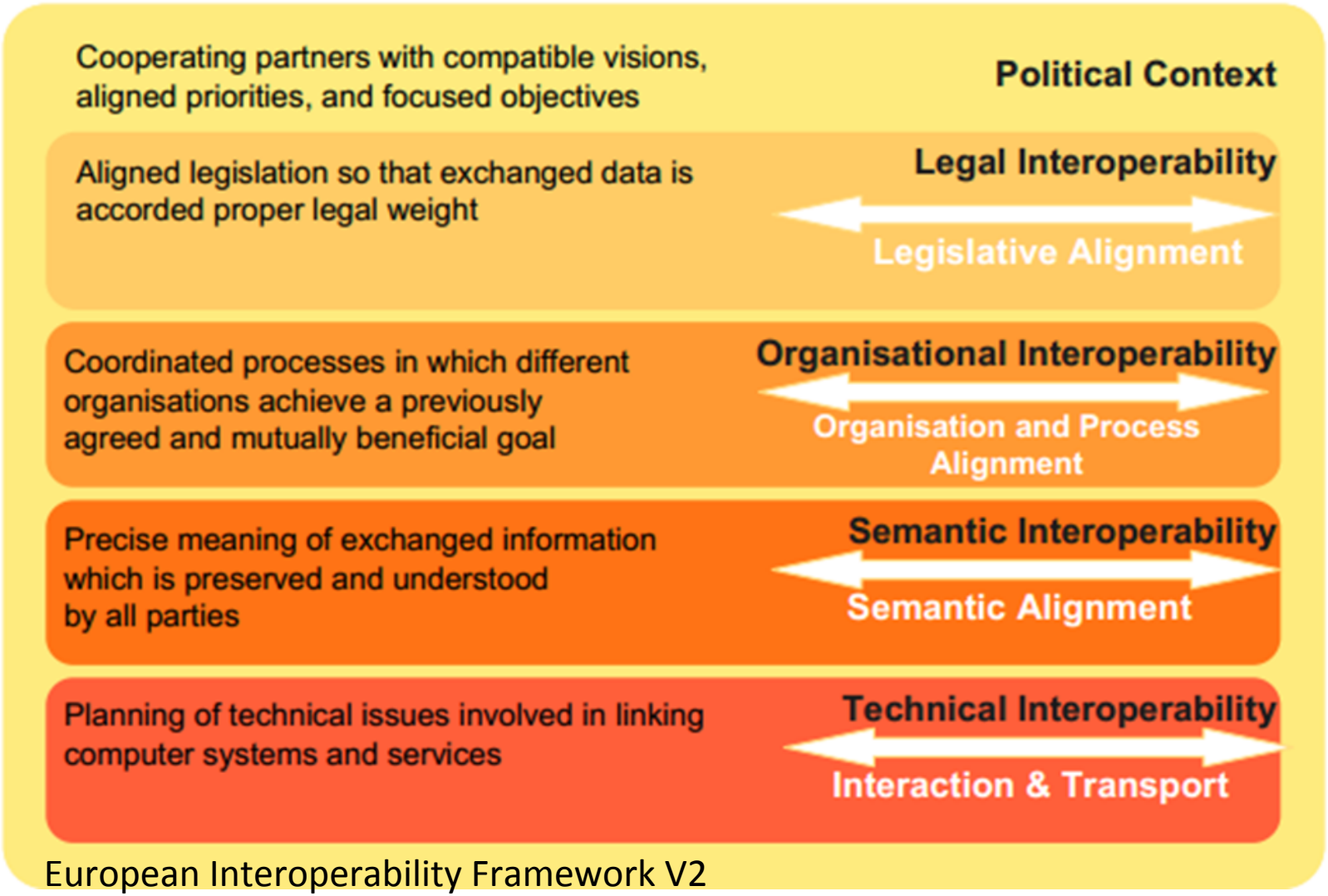
http://go.egi.eu/cloud

New flagship use cases

- 3 selected from 15 proposals:
  - European Center for Medium Range Weather Forecasts (ECMWF)
    - Weather Data Information Supersite (WDIS) with 100 years of weather data
  - UNESCO (Intergov. Oceanographic Commission)
    - Ocean and Coastal Information Supersite (OCIS)
  - Port d'Informació Científica (PIC), Barcelona
    - Reduce costs and improve speed of delivery, increase volume and accuracy for Neuroimaging
- Expect to deploy the new flagships by end 2013

Building the hybrid cloud

How to approach interoperability



European Interoperability Framework V2

Useful not only for public-commercial hybrid model but also between public services

Implementing the e-infrastructure vision

- Build a hybrid model of public and commercial service suppliers into a network of *Centres of Excellence*
- Make use of existing European e-infrastructures to jointly offer integrated services to the end-user
- *Centres of Excellence* can be owned and operated by a mixture of commercial companies and public organisations offering a portfolio of services
  - Services made available under a set of terms & conditions compliant with European jurisdiction & legislation and service definitions implementing recognised policies for trust, security and privacy notably for data protection
- A management board where the *Centres of Excellence* operators are represented to provide strategic and financial oversight - coupled with the user forum
- A pilot service (2014) initially offering a limited set of services at prototype *Centres of Excellence*

See https://cds.cern.ch/record/1562865/files/CERN-OPEN-2013-019.pdf

Prototype Centres of Excellence – Example from CERN

- This *Centre of Excellence* will focus on data-centric services representing a platform on which more sophisticated services can be developed
- Use the resources installed by CERN at the Wigner Research Centre for Physics in Budapest, Hungary
- Services will be accessible via single sign-on through a fed id. mgmt system
  - Multi-tenant compute environment to provision/manage networks of VMs on-demand
  - 'dropbox' style service for secure file sharing over the internet
  - Point-to-point reliable, automated file transfer service for bulk data transfers
  - Open access repository for publications and supporting data allowing users to create and control their own digital libraries (see www.zenodo.org)
  - Long-term archiving service
  - Integrated Digital Conferencing tools allowing users to manage their conferences, workshops and meetings
  - Online training material for the services

Sustainability of CERN's Centre of Excellence - role of partners

- **Partners will**
  - curate their data-sets
  - connect their identity federations
  - deploy their community specific services & portals
  - manage the interaction with their registered users and associated support activities

- Beyond this first year, partners engage to fund the cost of the services their users consume according to a pay-per-usage model (to be jointly-developed with CERN during the first year)

Beyond the initial prototype Centres of Excellence

- Learn from the prototype Centres of Excellence to build similar structures around Europe
  - Not identical: each has its own portfolio of services and funding model
  - All interconnected: to offer a continuum of services
  - All integrated with public e-infrastructures:
    - GEANT network (commercial networks are not excluded!)
    - PRACE capability HPC centres
    - EGI fed cloud

Summary

The Research Communities have

- **Highlighted identity mgmt as a key service**
- **Aligned their basic requirements**
- **Undertaken a series of prototypes/pilots with providers**
- **Engaged with industry to develop hybrid public-commercial models**

From 2014 onwards they will start to deploy services for their users

This is an of opportunity to propose identity management services to the resource communities

Think ***service***