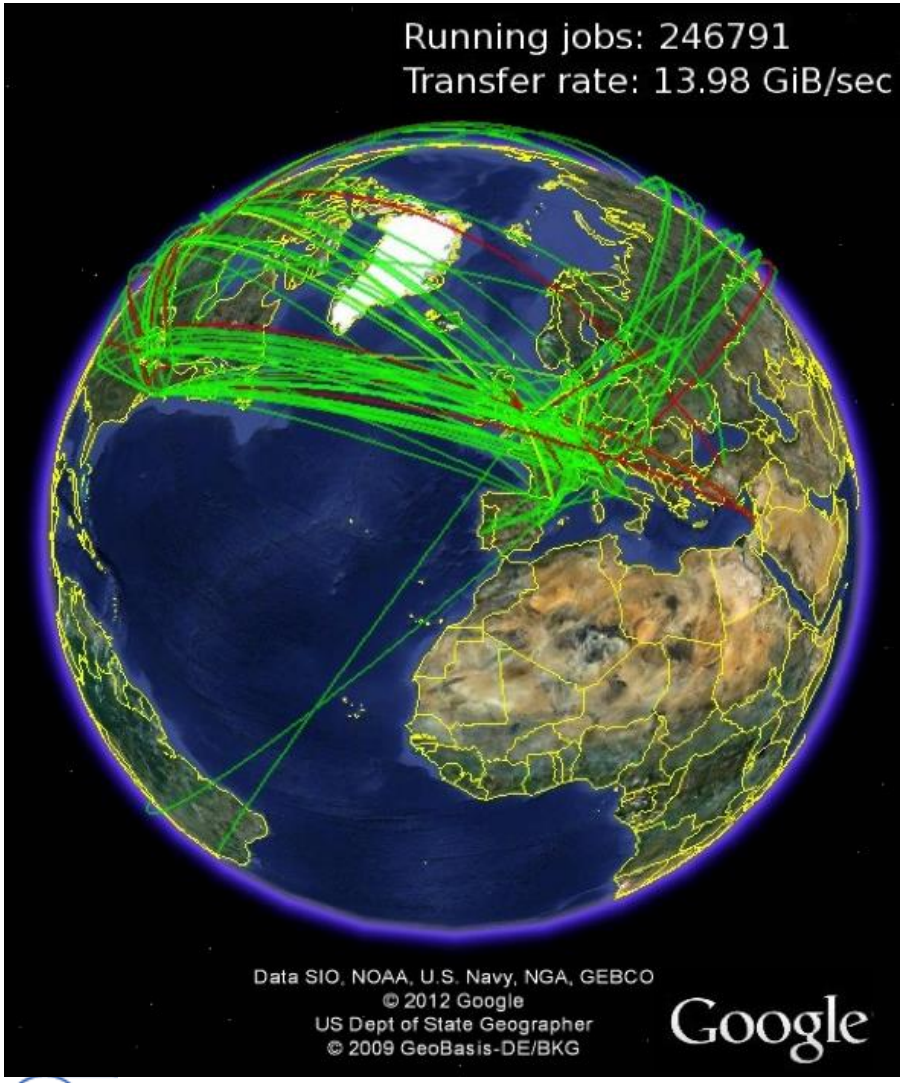
Helix Nebula The Science Cloud

Workshop on Best Practices for Data Management & Sharing

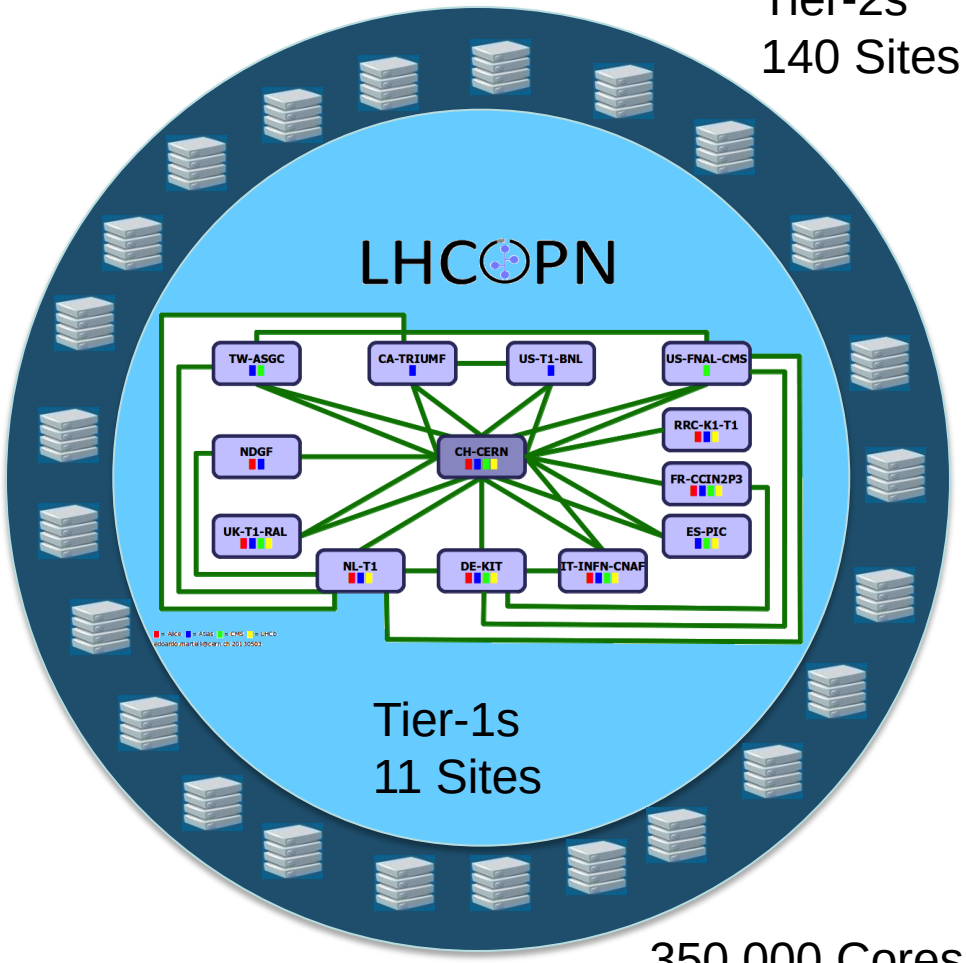
Ispra – 15 April 2014

Bob Jones (CERN)

Worldwide LHC Computing Grid



Tier-2s
140 Sites

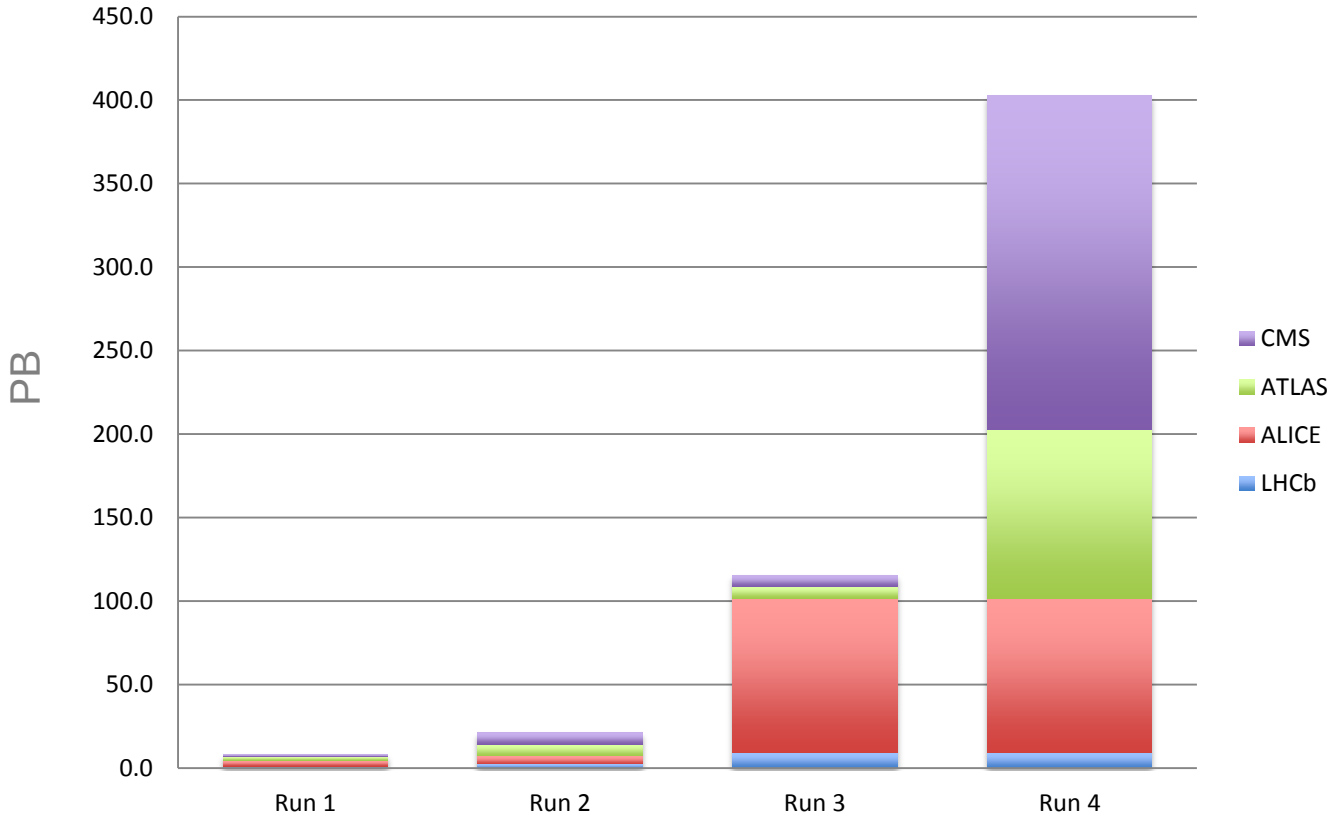


350,000 Cores

LHC schedule beyond LS1

Run 1 – which led to the discovery of the Higgs boson – is just the beginning. There will be further data taking – possibly for another 2 decades or more – at increasing data rates, with further possibilities for discovery!

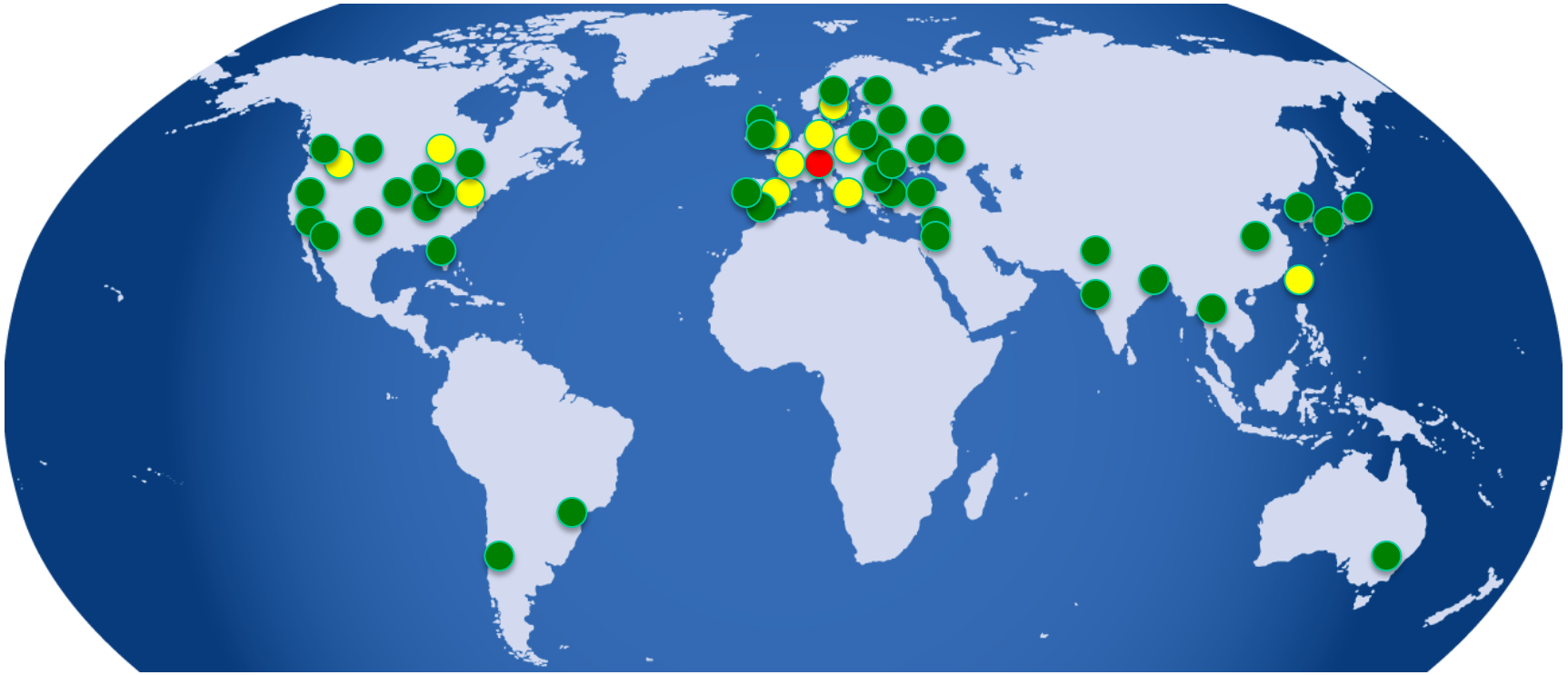
Data: Outlook for HL-LHC



- Very rough estimate of a new RAW data per year of running using a simple extrapolation of current data volume scaled by the output rates.
  - To be added: derived data (ESD, AOD), simulation, user data…

➢ **0.5 EB / year is probably an under estimate!**

WLCG



Strategic Plan for a Scientific Cloud Computing infrastructure for Europe

- **Establish a sustainable multi-tenant cloud computing infrastructure in Europe**
- **Initially based on the needs for the European Research Area & space agencies**
- **Based on commercial services from multiple IT industry providers**
- **Adhere to internationally recognised policies and quality standards**
- **Governance structure involving all stakeholders**

CERN-OPEN-2011-036
08/08/2011

Contacts

Dr. Maryline Lengert
ESA - European Space Agency
Senior Advisor
Maryline.Lengert@esa.int
Tel +39 06 941 80430

Dr. Bob Jones
CERN – European Organization for Nuclear Research
IT department
Bob.Jones@cern.ch
Tel. +41 22 767 14 82

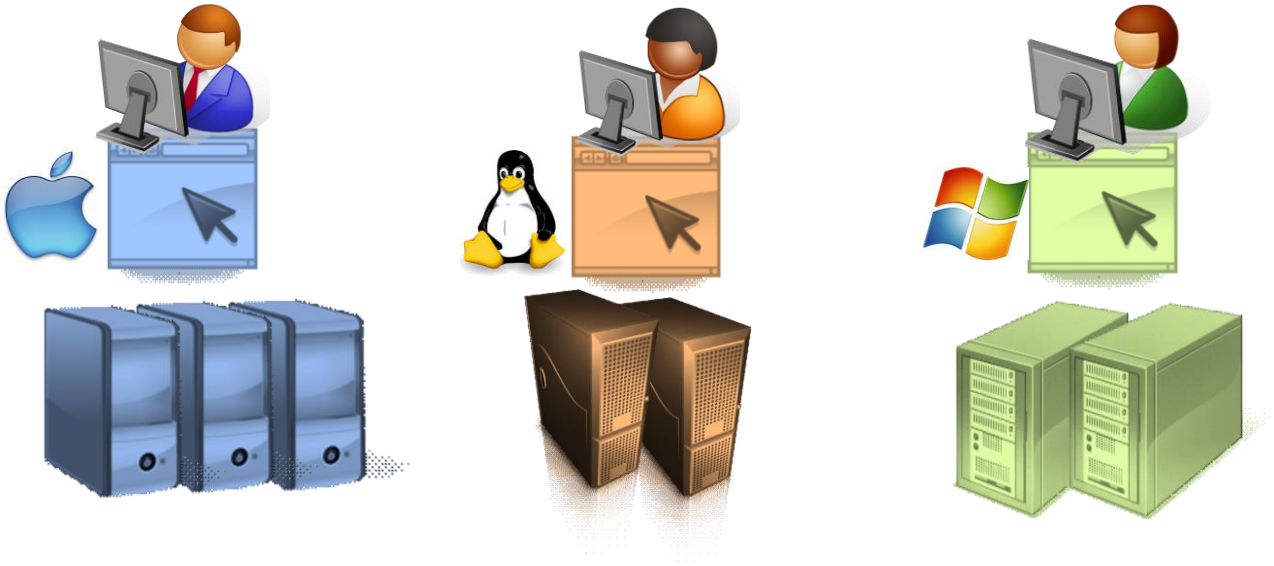
Timeline

- 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**

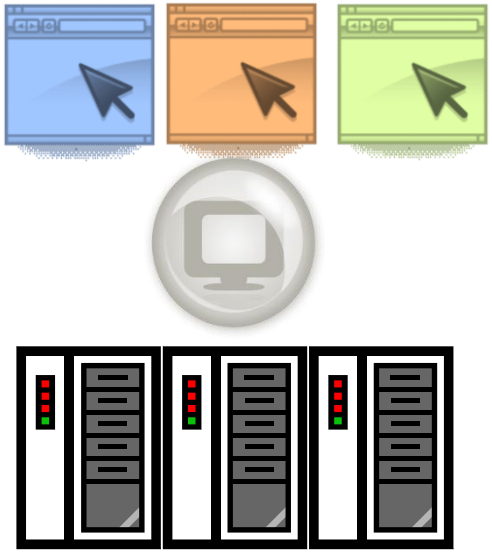
Cloud Services: *virtualisation*



User

Application

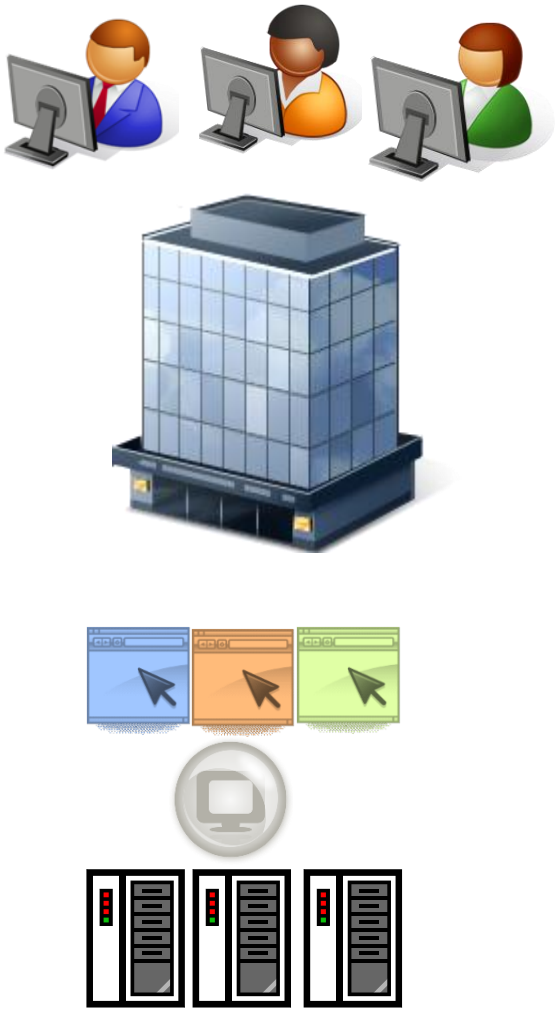
Server



Hypervisor

Lower Cost
Faster Maintenance
More Flexibility

Cloud Services: *outsourcing & offshoring*



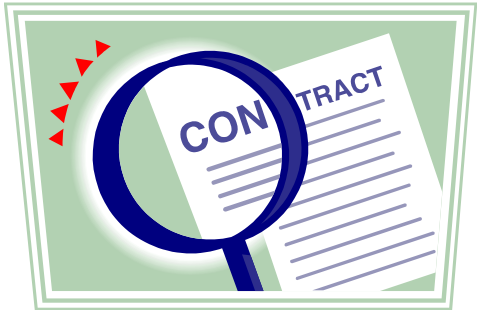
From CapEx to OpEx
Procurement Process
Terms and Conditions
Jurisdiction and Legislation

Cloud Services: *contracts*

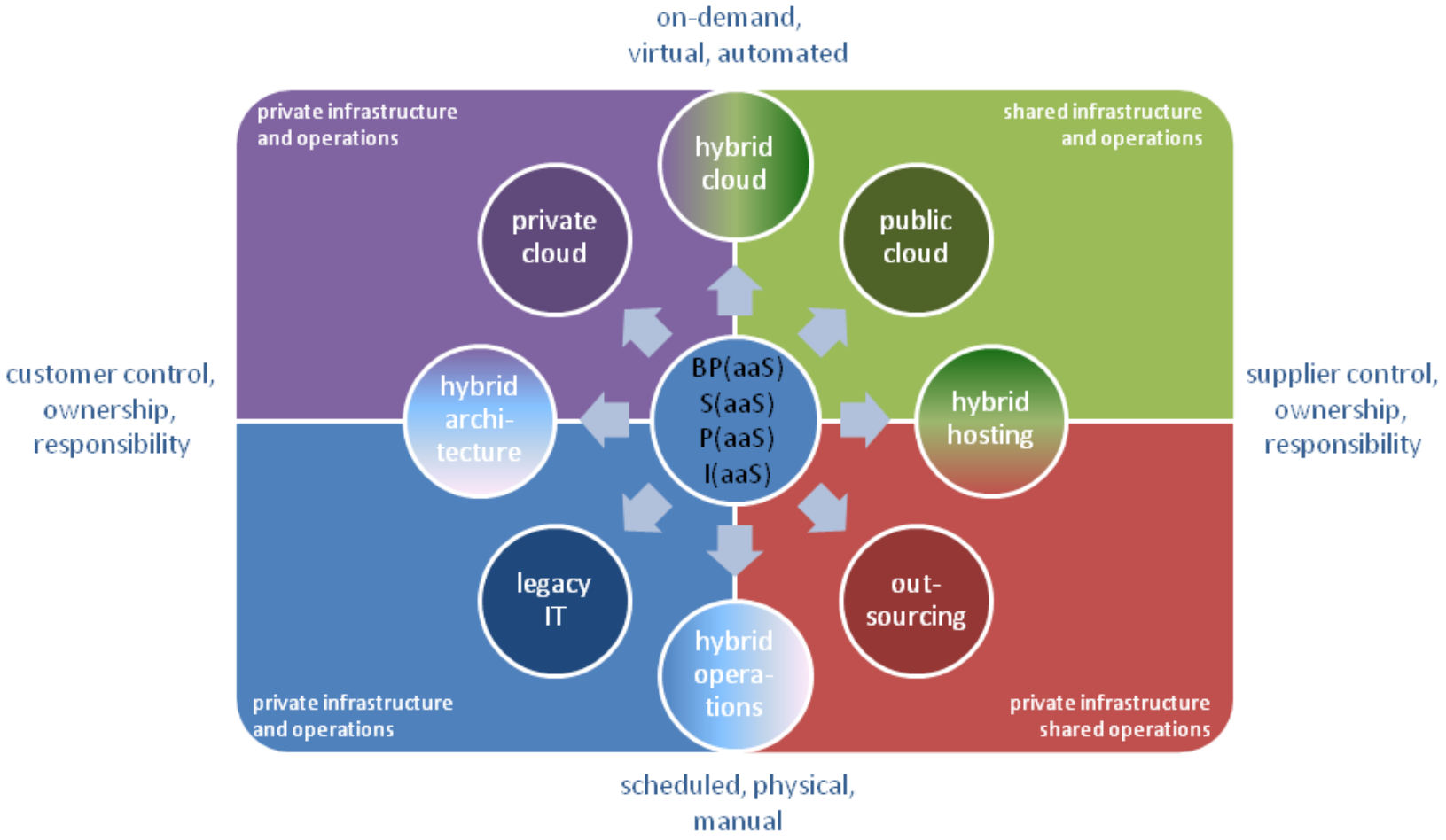
- **Liability** - *Exclusion, Limits and Remedies for Breach of Warranties & Indemnities*
- **Service Levels** - *Data integrity, resilience and business continuity, transparency*
- **Regulatory Issues** - *Data location and data export*
- **Confidentiality** - *Rights to Monitor, Access, Disclose or Use Customer Data*
- **Security** - *Requirements, Audit Rights, Security Incidents and Response*
- **Vendor Lock-In** - *Data retention, deletion and data portability*
- **Term and Termination**
- **Changing Service Description or Features**
- **Intellectual Property Rights**

Source: Cloud legal project, Queen Mary College London
http://www.cloudlegal.ccls.qmul.ac.uk/

Webcast lecture from CERN on 13 May:
https://indico.cern.ch/event/306750/



Hybrid clouds



Our preferred model is a hybrid cloud that combines commercial cloud services with resources managed by public organisations

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

HELIX NEBULA THESCIENCECLOUD

Strategic Plan

- Establish 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 LHC experiments

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

PIC port d'informació científica

To improve the speed and quality of research for finding surrogate biomarkers based on brain images

Suppliers

Atos CGI CloudSigma CSA cloud security alliance DANTE www.dante.net egi indra interoute from the ground to the cloud MEMSET HOSTING SAP sixsq SWITCH T-Systems terradue 2.0 the SERVER LABS the IT architects ultimum technologies Yandex

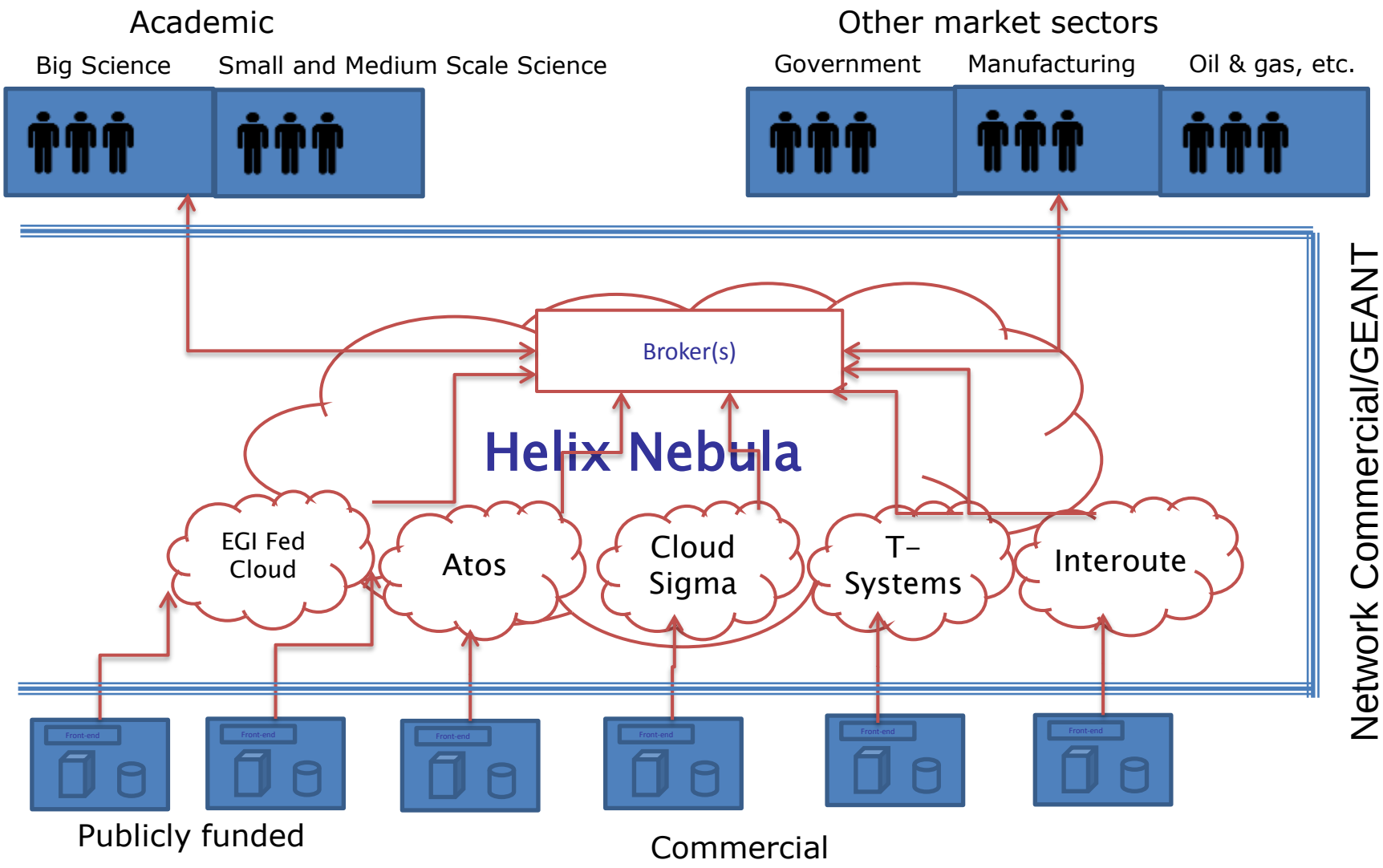
Adopters

AWST Capgemini CONSULTING.TECHNOLOGY.OUTSOURCING cnrs ECMWF Ifremer orange Business Services Telefónica THALES Trust-IT Communicating ICT to Markets www.trust-itservices.com

Flagship use cases

	ATLAS H.E.P. Cloud Use (CERN)	Genomic Assembly in the Cloud (EMBL)	SuperSites Exploitation Platform (ESA/CNES/DLR)
Scientific goal/society impact/photogenic	•	•	•
Scale of resources used	•	•	
Federation/Aggregation of datasets		•	•
Long-term archiving of data			•
On-demand processing	•	•	•
Impact on community & benefits	•	•	•
Potential increase of users	•	•	•
Interoperability	•	•	•
Data security	•	•	•
Maturity	•	•	•
Access to license-controlled sw			•

Hybrid Public-Private Cloud Model



Helix Nebula Marketplace (HNX)

- **Builds upon the work of the Helix Nebula Initiative and EC support action**
- **Supported by European cloud providers**
- **Integrates with existing e-Infrastructures to form a hybrid cloud Market Place and reach out to Europe's research communities**
- **Trusted cloud services through compliance with EU regulations and legislation**
- **Simplifies procurement process across multiple services providers**

sixsq Atos CloudSigma

the SERVER LABS the IT architects · T-Systems · interoute · CGI · egi

Relevance of HNX to H2020

HORIZON 2020 WORK PROGRAMME 2014 – 2015

4. European research infrastructures (including e-Infrastructures)

- Introduction:
  - "production-level e-infrastructures are able to serve the computing and data needs of any project in the framework programme fostering economies of scale in the use of ICT systems by projects supported by Horizon 2020."
  - "A related new element in Horizon 2020 is the use of Data Management Plans (DMPs) detailing what data the project will generate, whether and how it will **be exploited or made accessible for verification and re-use**, and how it will be **curated and preserved**."
- INFRADEV:
  - "Proposals should build upon the state of the art in ICT and e-infrastructures for data, computing and networking and work in cooperation with e-infrastructure service providers."

What have we learnt?

Public organisations see value and opportunities in making use of commercial cloud services

The production usage of commercial cloud services by public organisations has already started

The public sector is a potentially profitable market for commercial cloud service providers

The procurement and use of commercial cloud services poses a number of legal questions

A coordinated approach by public organisations will help structure the market and reduce the burden on individual organisations

E-Infrastructure for the 21st Century

- The goal is to transform existing Distributed Computing Infrastructures (DCIs) based on a range of technologies into a *service-oriented platform* for the *global research community* that can be *sustained* through *innovative business models*

- Prepared by CERN on behalf of the EIROforum IT Working Group

DOI:10.5281/zenodo.7592

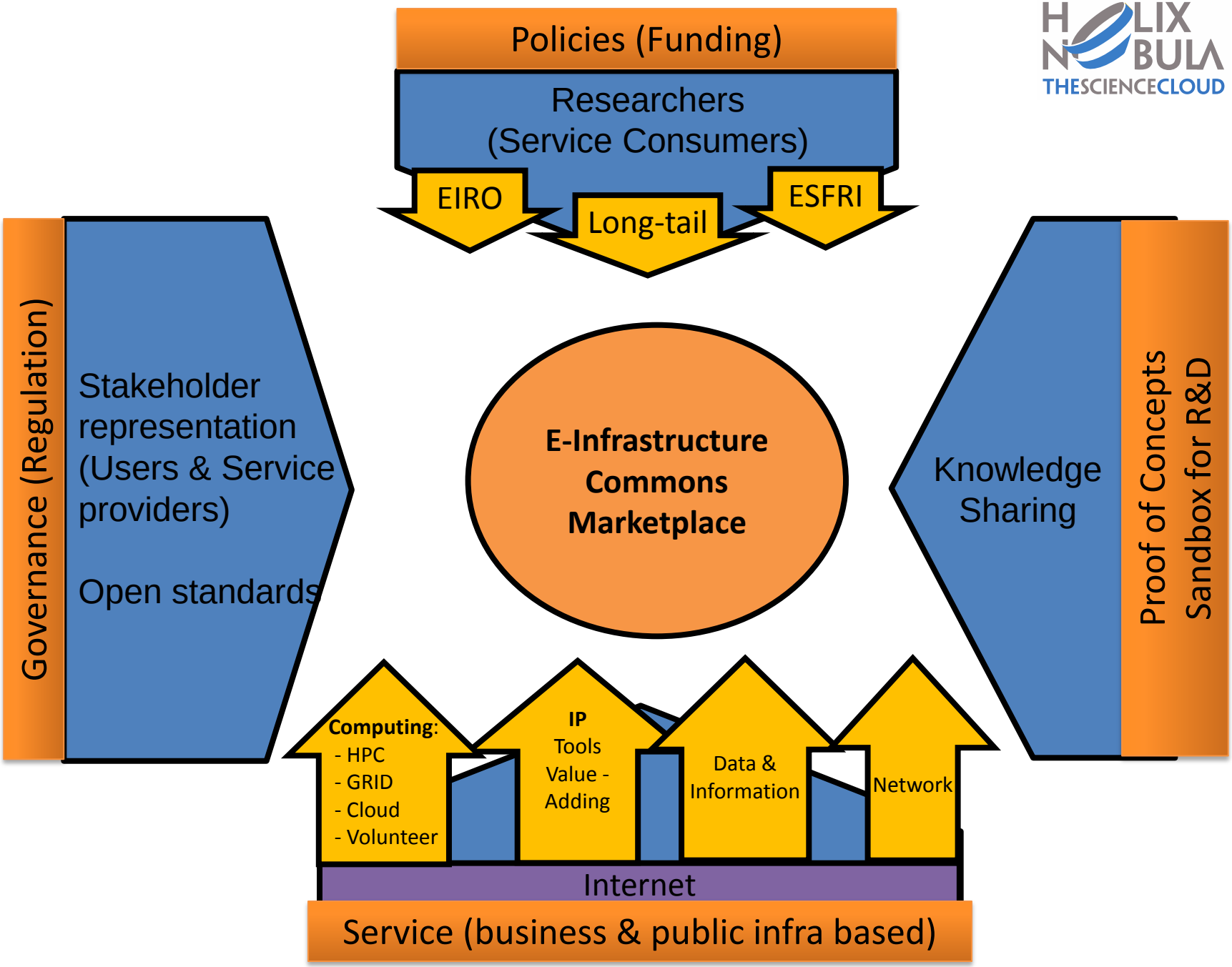
Vision for the future

The e-Infrastructure commons marketplace will

- Provide access to world class resources through a **dynamic and sustainable marketplace**
- Take a hybrid approach building on public and commercial assets to cover the **entire scientific workflow**
- Offer a broad range of services
- Use **open standards** to ensure **interoperability** of service providers and adhere to **European policies, norms and requirements**

To get there, we need to

- Driving convergence between existing e-Infrastructure service providers
- Use market mechanisms to attract new service providers & consumers to the Marketplace


Policies (Funding)
Researchers
(Service Consumers)
EIRO
Long-tail
ESFRI
Governance (Regulation)
Stakeholder representation (Users & Service providers)
Open standards
E-Infrastructure Commons Marketplace
Knowledge Sharing
Proof of Concepts Sandbox for R&D
Computing:
- HPC
- GRID
- Cloud
- Volunteer
IP
Tools
Value - Adding
Data & Information
Network
Internet
Service (business & public infra based)

Make it happen in H2020

- Make it possible to **trade** services
  - Services from commercial and public providers offered on a pay-per-usage model should be considered eligible costs for EC projects

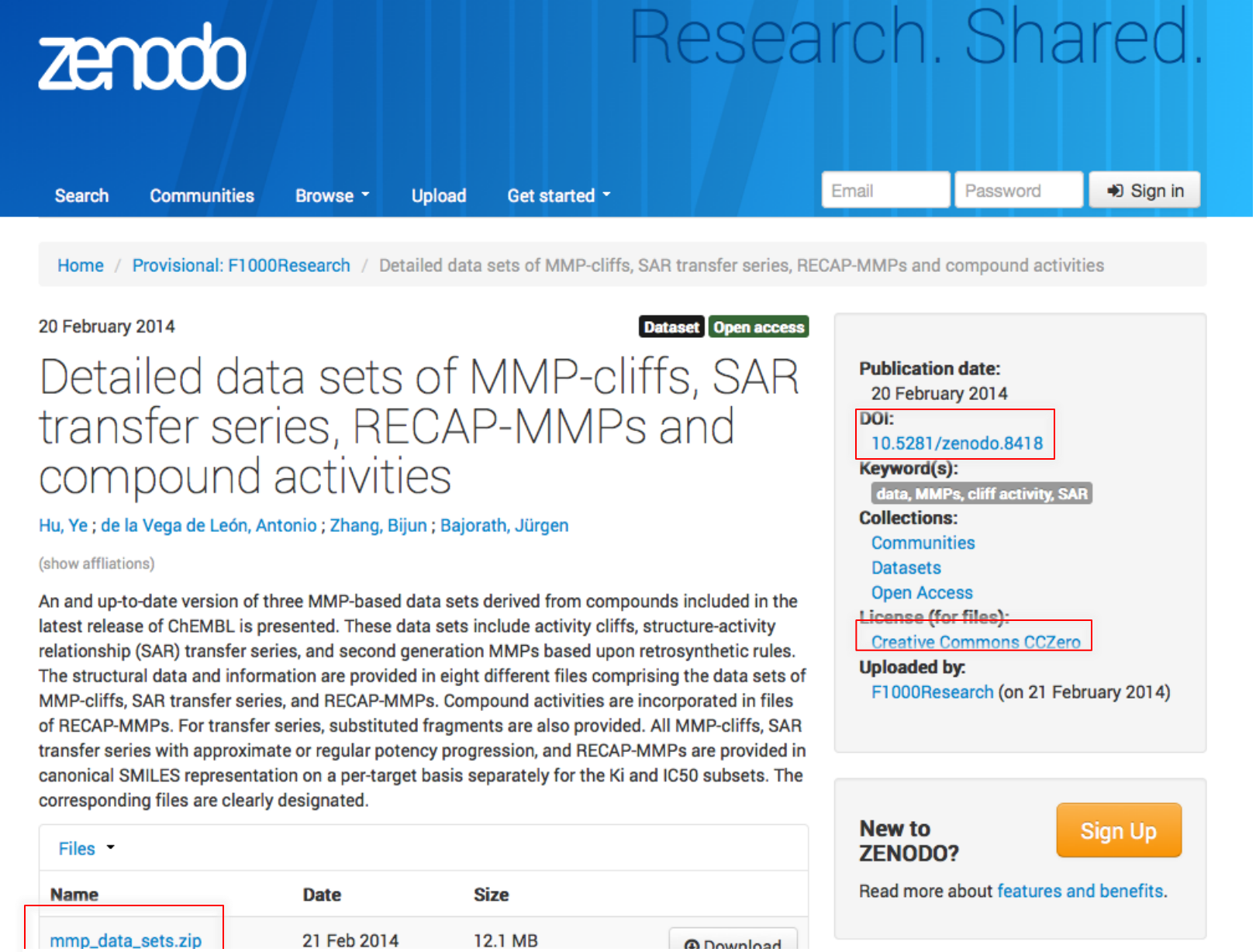
robert madelin @eurohumph · Feb 17

Data solutions. Help EU clouds and computation networks. Make use of them an allowable cost in EU grant agreements. Cloud/Grid as a service.

Expand Reply Retweet Favorite More

Model GA article13 — Implementation of Action Tasks by Subcontractors

The beneficiaries must ensure that the majority of the research and development work done by the subcontractor(s) (including the work of the main researchers) is located in the EU Member States or associated countries ('place of performance obligation').



DOI

Licence

Data

Being extended to include citation of software:
http://www.isgtw.org/spotlight/tool-developed-cern-makes-software-citation-easier

isgtw international science grid this week

Expected impact

- Researchers, supported by large scale long-term research infrastructures, drive the evolution of services for their research needs
- Funding agencies benefit from market forces
- Create a fertile environment that nurtures new scientific ideas and challenges
- Service providers are able to sustain services
- Grow an ecosystem that benefits downstream industries
- Assemble a dynamic market place, building on Information as a Service, based on federation meeting European requirements
- Provide visibility and incentives to industry to invest in new assets (*as a business case but also to use the science communities for testing cutting-edge technology and new services*)


HELIX NEBULA
THESCIENCECLOUD
60 YEARS / ANS CERN
Helix Nebula *Cloud Productive* event
CERN, Geneva, 14 May 2014
http://indico.cern.ch/e/Helix_Nebula_Cloud_Productive
e-infrastructure