LHC Computing, CERN, & Federated Identities

Ian Bird; WLCG Project Leader, CERN

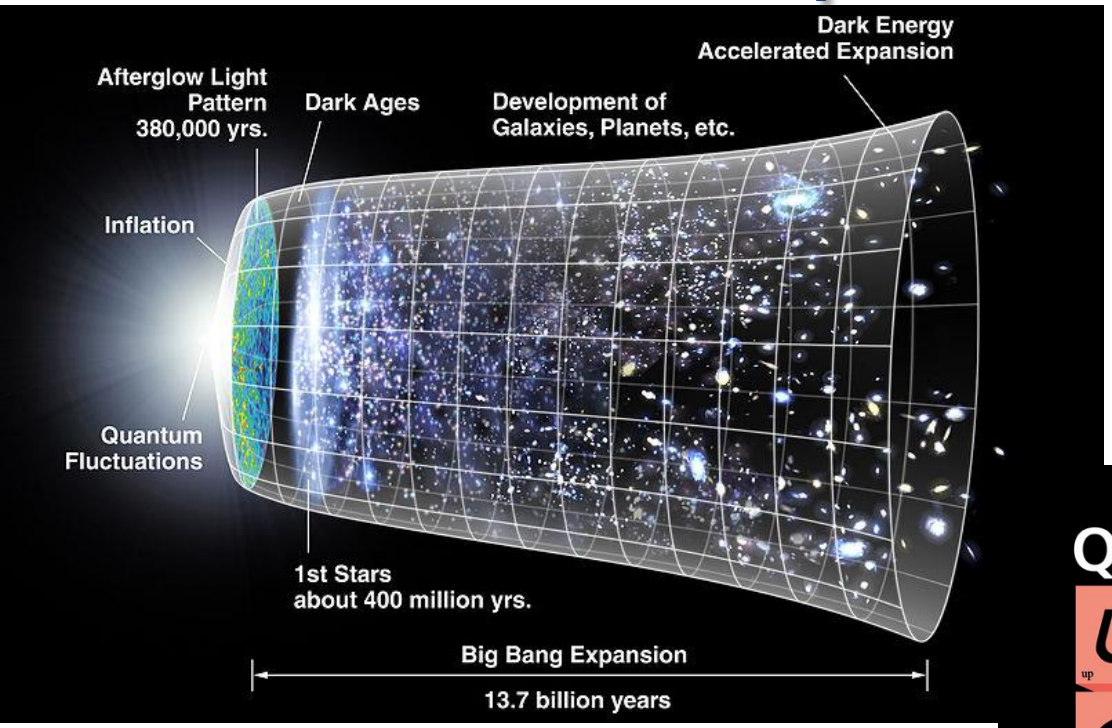
FIM4R Workshop, CERN

3rd February 2015

About CERN

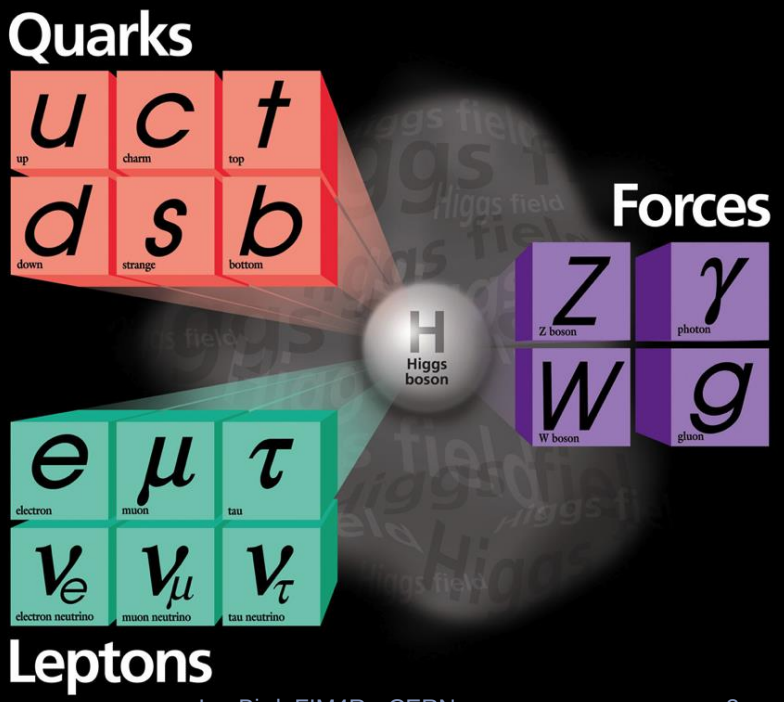
- **CERN is the European Organization for Nuclear Research in Geneva**
  - Particle accelerators and other infrastructure for high energy physics (HEP) research
  - Worldwide community
    - 21 member states (+ 2 incoming members)
    - Observers: Turkey, Russia, Japan, USA, India
    - About 2300 staff
    - >10’000 users (about 5’000 on-site)
    - Budget (2012) ~1000 MCHF
- **Birthplace of the World Wide Web**

Fundamental questions…



- How to explain that particles have mass?
  - Theories and accumulating experimental data…getting close
- What is 96% of the Universe made of?
  - We only observe 4% of the apparent mass

- Where is all the anti-matter?
  - Why is Nature not symmetric?
- What was the state of matter just after the Big Bang?
  - "Soup" of quarks and gluons before they condensed into matter?



Global Effort → Global Success

Results today only possible due to extraordinary performance of accelerators – experiments – Grid computing

Observation of a new particle consistent with a Higgs Boson (but which one…?)

Historic Milestone but only the beginning

Global Implications for the future

R-D Heuer

The moment when Cern director Rolf Heuer confirmed the Higgs results

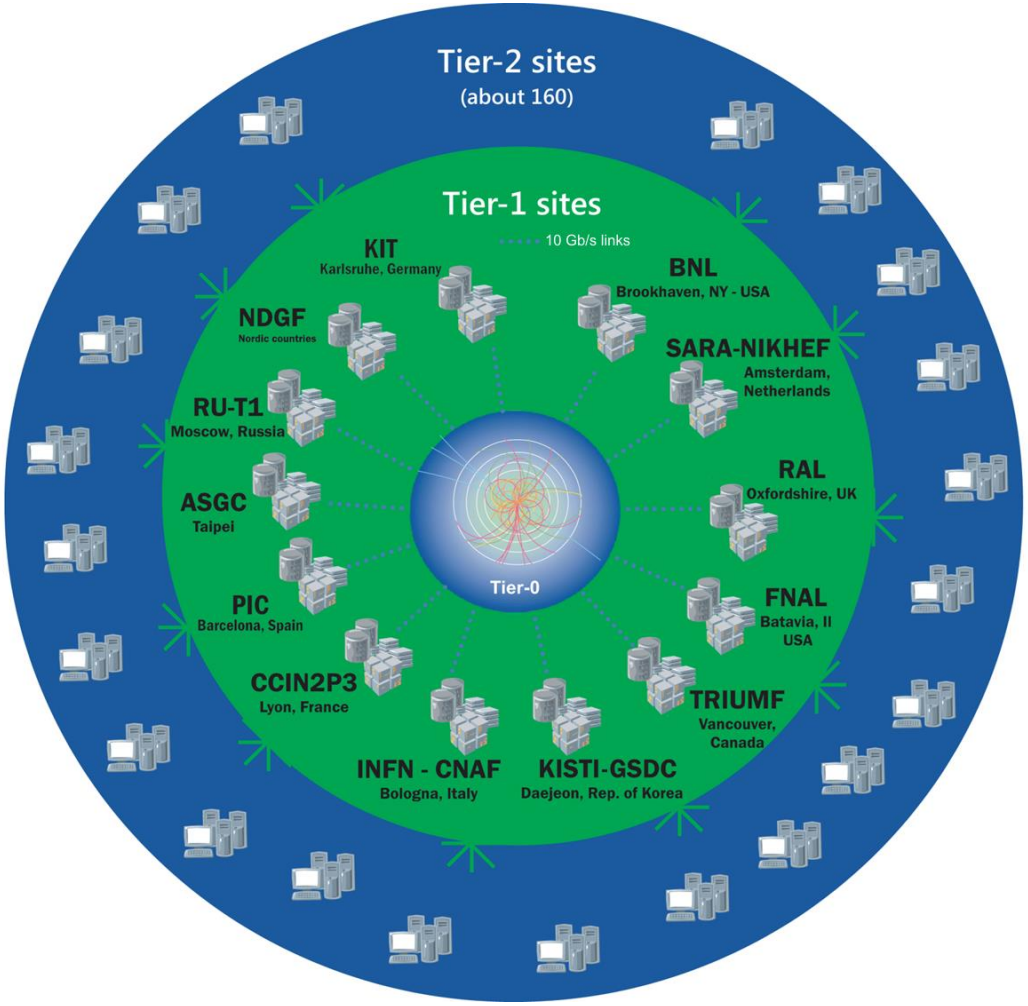
Cern scientists reporting from the Large Hadron Collider (LHC) have claimed the discovery of a new particle consistent with the Higgs boson.

The Worldwide LHC Computing Grid

Tier-0 (CERN): data recording, reconstruction and distribution

Tier-1: permanent storage, re-processing, analysis

Tier-2: Simulation, end-user analysis



nearly 170 sites, 40 countries

~350’000 cores

500 PB of storage

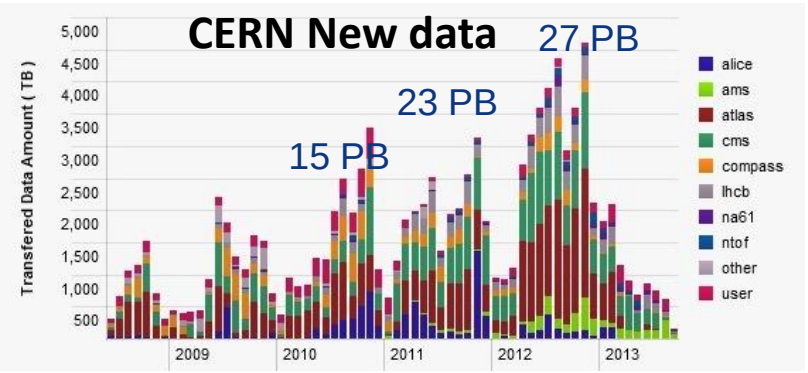
> 2 million jobs/day

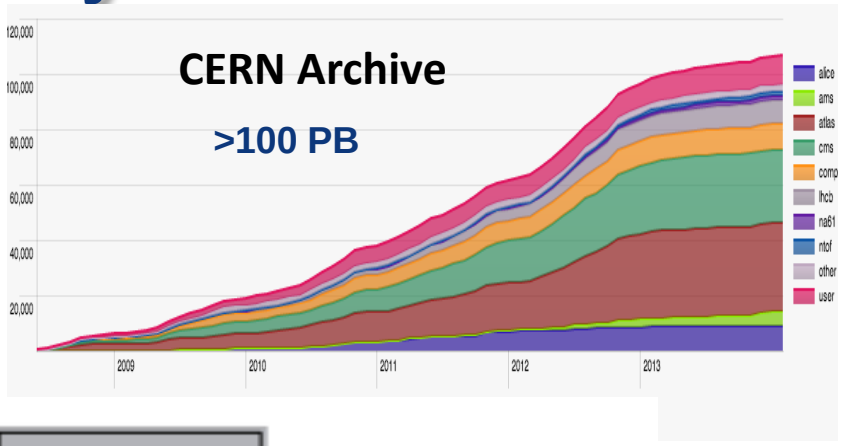
10-100 Gb links

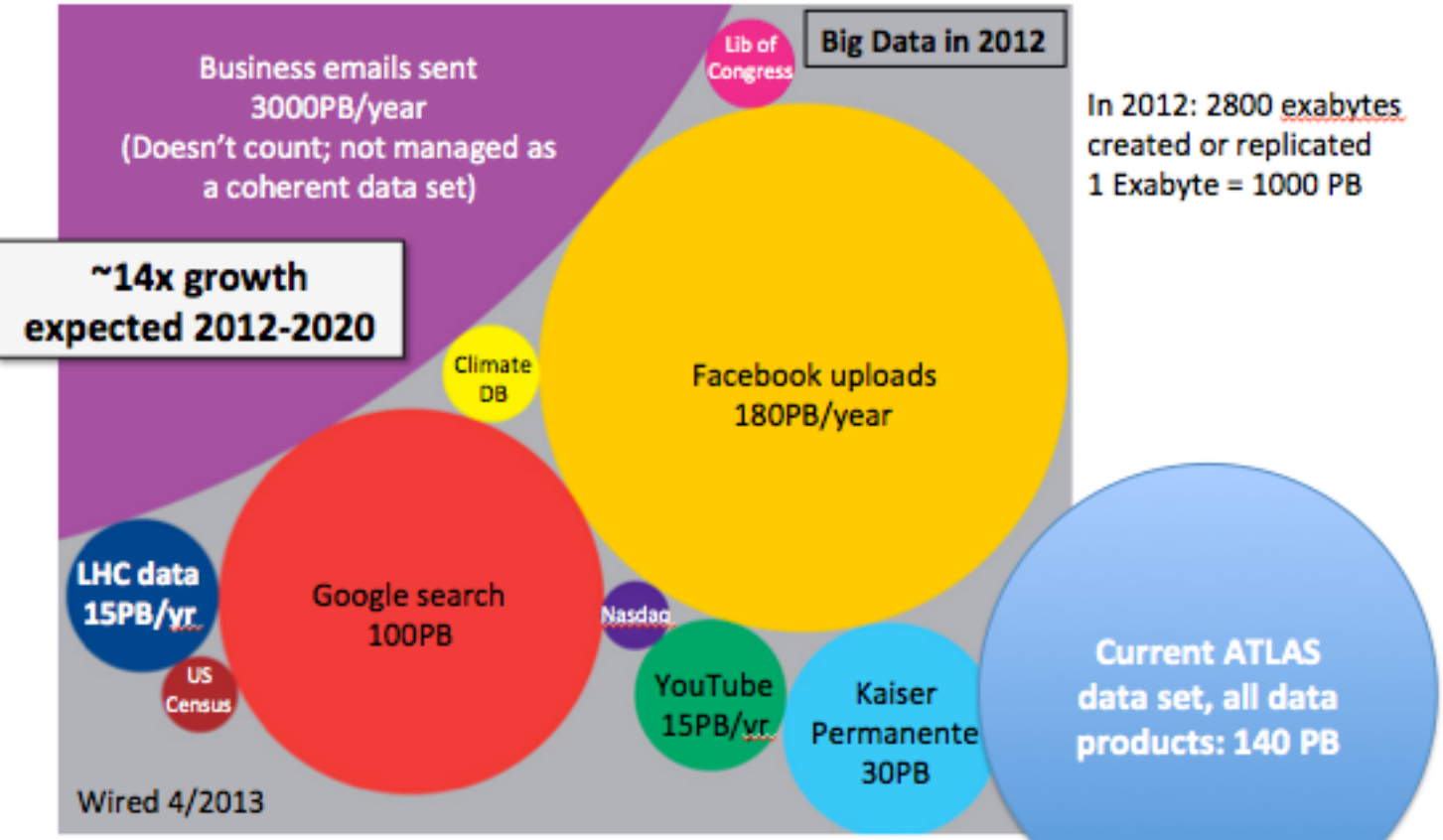
WLCG:
An International collaboration to distribute and analyse LHC data

Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists

Scale of data today …







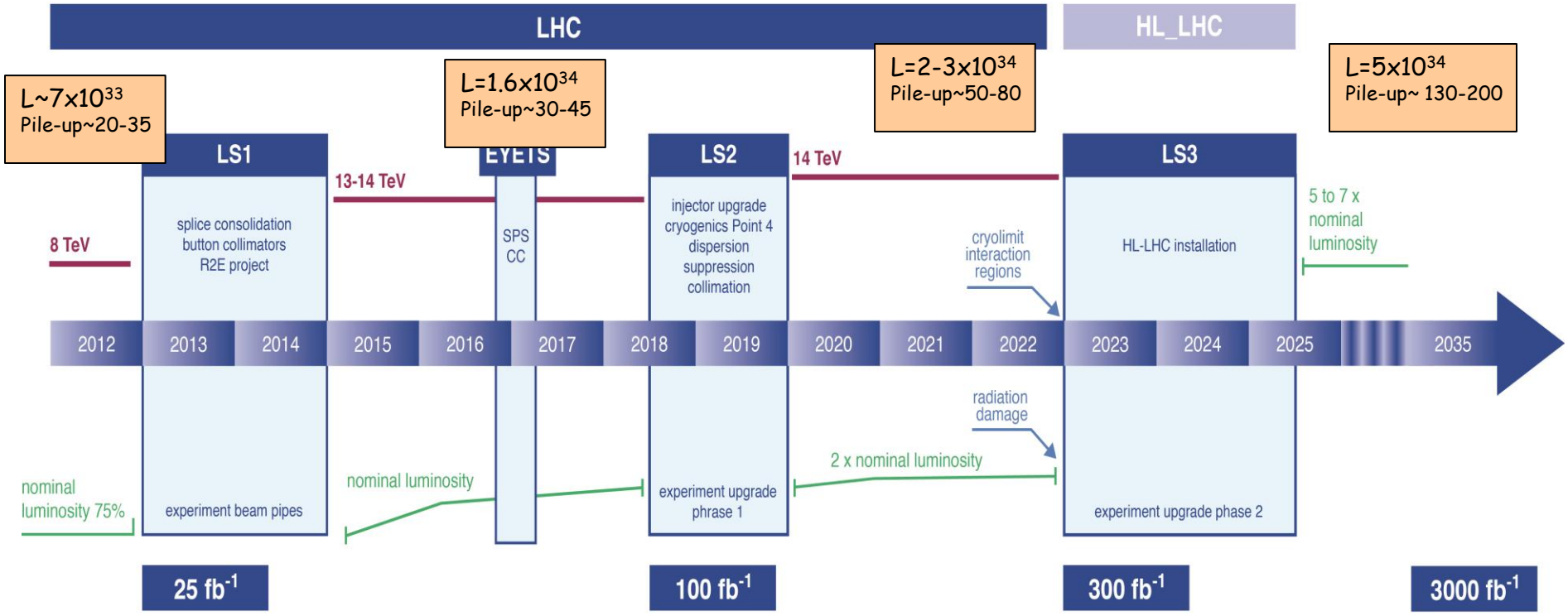
In 2012: 2800 exabytes created or replicated
1 Exabyte = 1000 PB

http://www.wired.com/magazine/2013/04/bigdata/

The LHC timeline

L.Rossi

New LHC / HL-LHC Plan



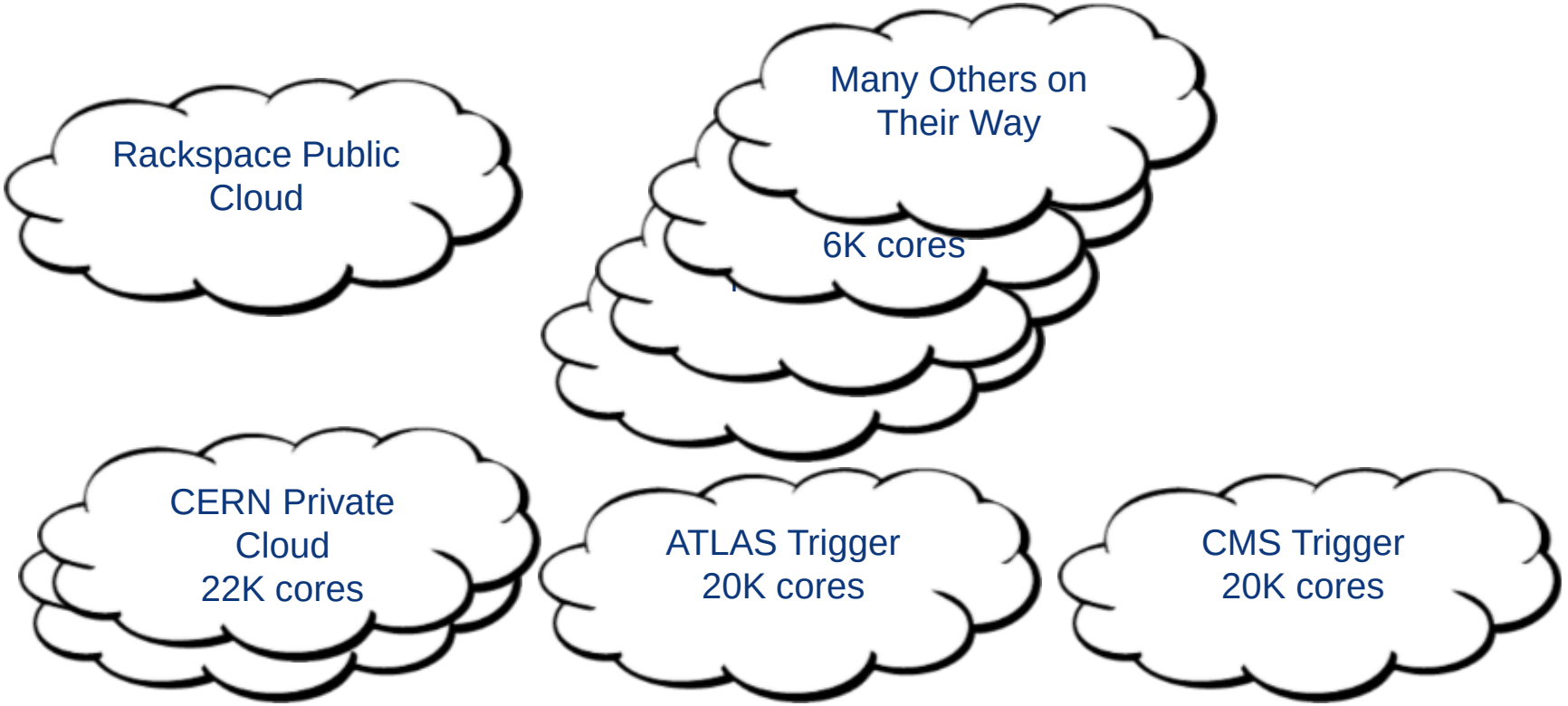
Computing Model update

- Need to keep evolving the computing models
- e.g. Evolution of grid model: use of new technologies
  - Cloud/virtualisation
  - Data federations, intelligent data placement/caching, data popularity service
  - Federated identities

Broad LHC Computing "strategy"

- Live within ~fixed budgets
- Remove complexity – reduce costs
  - In services, In operations; In support
- Gain x10 - 100?? in event processing throughput
  - Find additional resources (additional budgets)
  - Make (much!) better use of existing resources
    - Optimisation of cost of [CPU, storage, network, power]
    - Invest (limited) available effort where it is important
      - Software, data management
- Collaborate with other science communities
  - Share expertise, experience

Federation of resources



- Share resources, images, accounts between clouds ?
- In collaboration with Rackspace in CERN-openlab
- All contributions are to OpenStack upstream so will appear in all OpenStack clouds at all the sites

Federation of data

Access any data from any site without the need to first copy it

Network access to facilities and data will be cheap
Moving data around is expensive (needs disk!)

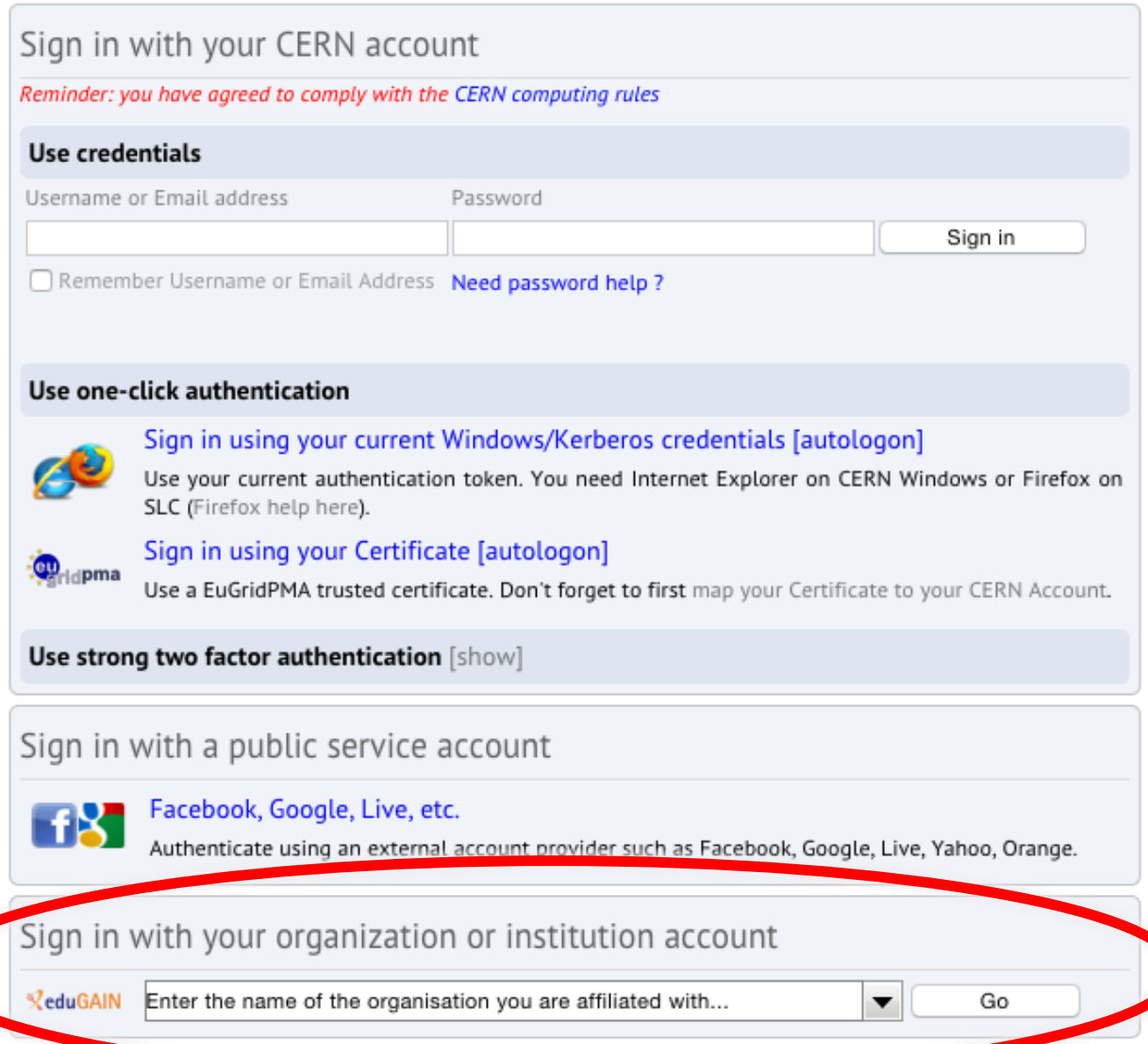
This sharing and interoperation of resources relies on federation of identities, etc

Federations

- WLCG is a “grid” …
  - “Federated distributed computing infrastructure”
  - Very little “grid” middleware left
- Initially “grid” came with a security model – X509
  - Never particularly user friendly, or reliable
  - Lived with it – workarounds
  - Always seen as a problem for other grid uses
- BUT: it has a global trust federation that works
  - We know how to do this

Federat

- CERN wou
  community
  - And toda
- WLCG wou
  certificates,
- CERN and
  infrastructu
  assumption
  - Essential
  - http://zen
- Essential th
  covers all u
- We see Edu
  - Broad wo
- Pilots have



YEARS/ANS CERN

Priorities

- In Europe the AARC project has just been approved to help develop EduGain
  - Essential that research communities communicate their needs clearly
- By definition, Federated identities for global communities must be a global effort
  - We must agree common levels of service and trust
    - Interoperation, attributes (definition, release), policies, etc.
  - Between all major parties – EU, US, etc
- These agreements are essential and urgent for progress and uptake.
- Large science projects are global collaborations now
  - It would be a complete failure if we cannot agree a way to harmonize basic identity management