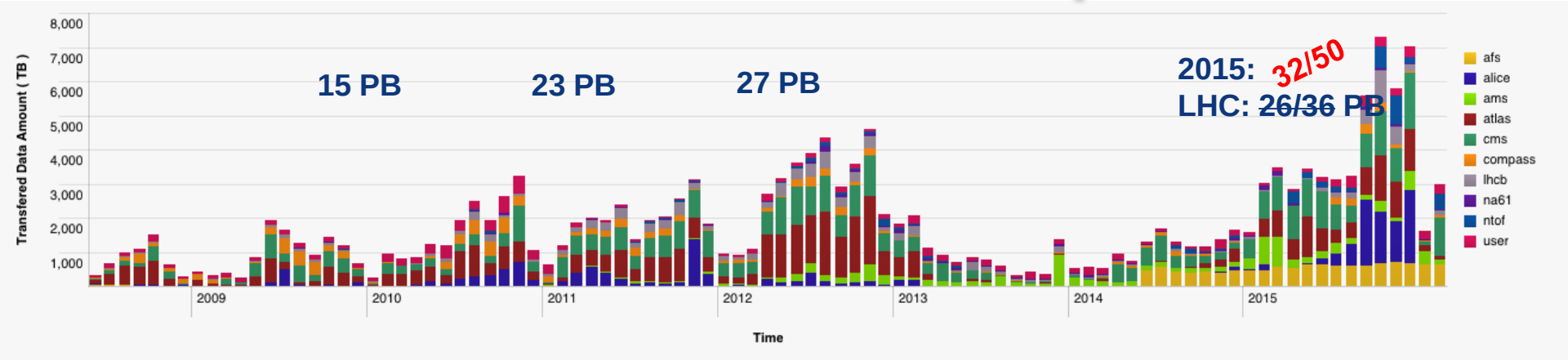
Ian Bird, CERN

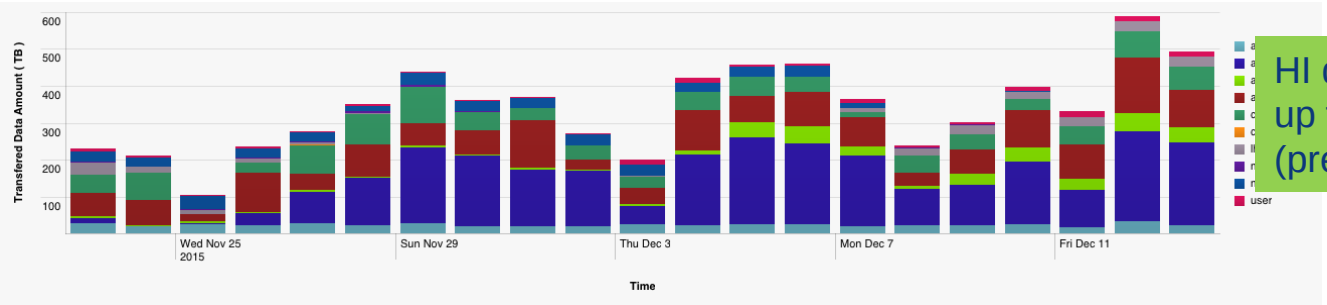
WLCG LHCC Referee Meeting

1st March 2016

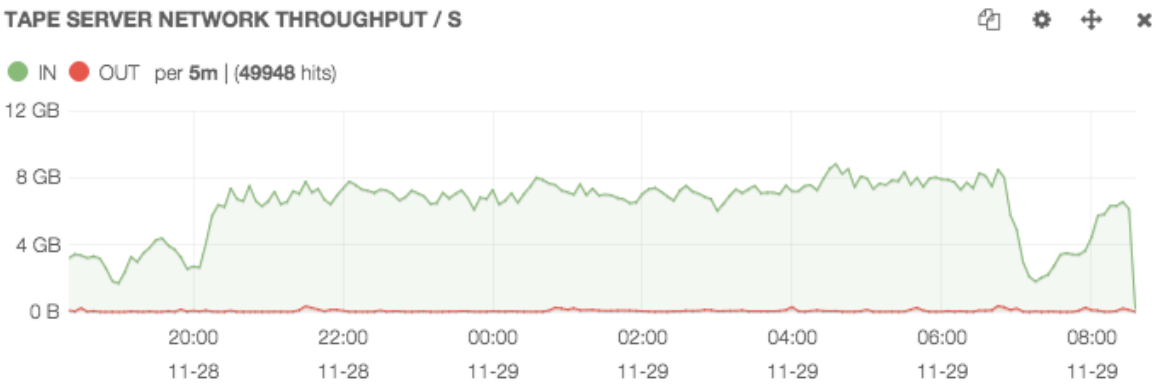
# WLCG Status Report

# 2015 data in Tier 0 - update



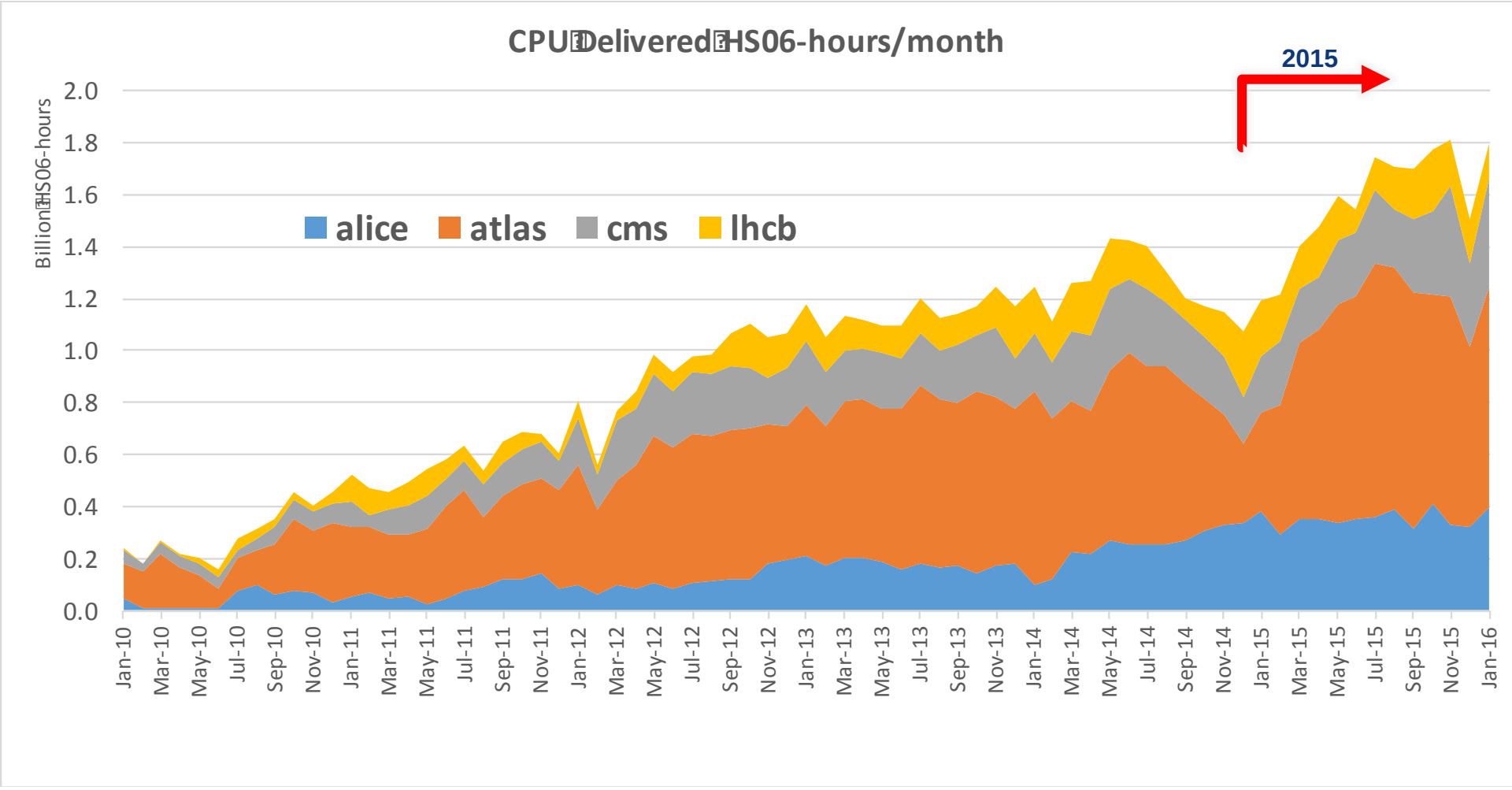




HI data: 600
up to 450 TB/day this weekend
(previous max was ~220 TB/day)

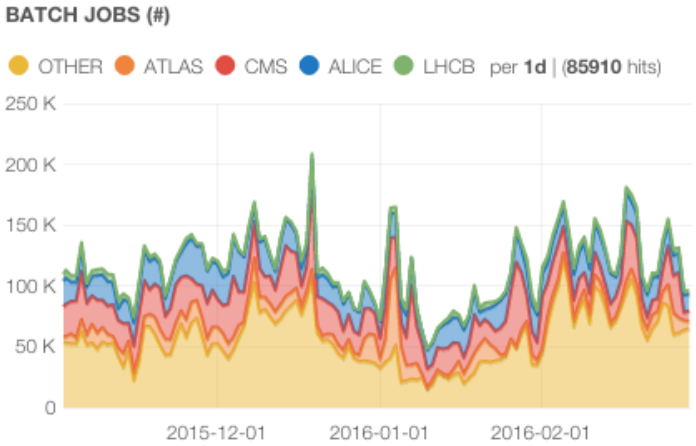


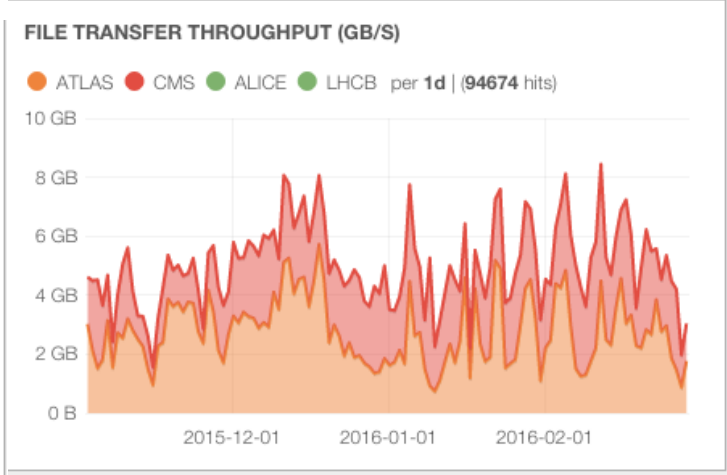

HI data:
Data rates >10 GB/s sustained for 8 h
(peaks ~10.5 GB/s)

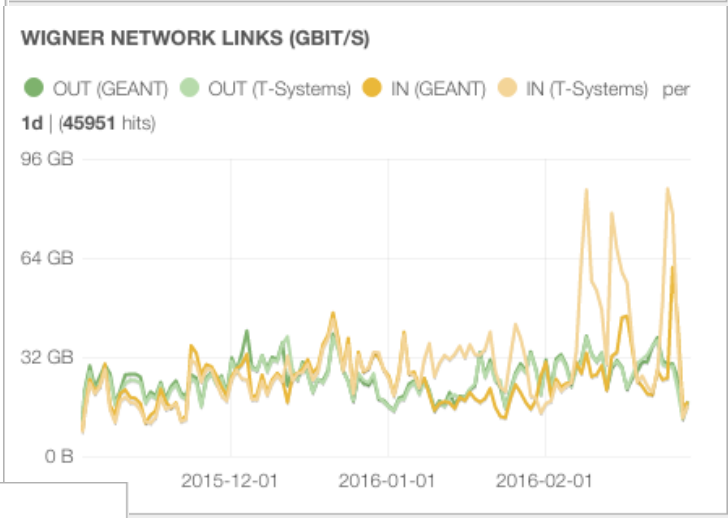
# Ramp-up of CPU - update



# Tier 0




**MEYRIN DATA CENTRE**

| | last_value |
|---|---|
| Number of Cores in Meyrin | 124,667 |
| Number of Drives in Meyrin | 70,994 |
| Number of 10G NIC in Meyrin | 6,007 |
| Number of 1G NIC in Meyrin | 21,651 |
| Number of Processors in Meyrin | 21,901 |
| Number of Servers in Meyrin | 11,720 |
| Total Disk Space in Meyrin (TB) | 122,859 |
| Total Memory Capacity in Meyrin (TB) | 509 |

**WIGNER DATA CENTRE**

| | last_value |
|---|---|
| Number of Cores in Wigner | 43,328 |
| Number of Drives in Wigner | 23,180 |
| Number of 10G NIC in Wigner | 1,399 |
| Numer of 1G NIC in Wigner | 5,067 |
| Number of Processors in Wigner | 5,418 |
| Number of Servers in Wigner | 2,712 |
| Total Disk Space in Wigner (TB) | 71,738 |
| Total Memory Capacity in Wigner (TB) | 172 |

# Pledge installation for 2016

- On track, no particular concerns have been flagged
  - KIT have additional funding and will be able to increase pledge ~October
- Triumf Tier 1 is moving to SFU, but is planning seamless transition with dual resources
- NL Tier 1 is also moving locations, but will have a 2 week down time in Autumn 2016

# Computing upgrades
# Summary from Lisbon workshop

# Introduction

- Two days devoted to medium term (Run 2-3) and longer term (Run 4) concerns
- ~140 people registered
- Aimed for more of a discussion format rather than presentations
  - (Informal) feedback from many said this was useful
    - Some aspects probably needed a bit more preparation to be more successful

# Observations

- Probably a lack of clarity over what the situation for Phase 2 upgrades will be:
  - In terms of requirements – what is the real scale of the problem – need better estimates
  - What we can really expect from technology
  - An understanding of the real limitations of the system we have today
- We should also bear in mind that while we potentially need to instigate *revolutionary* changes in computing models, nevertheless we will have to face an *evolutionary* deployment
- Concerns over software and efficiency (in all aspects) will be a significant area of work
- Commonalities may be possible in new tools/services or next generation of existing
- Propose a number of activities to address some of these aspects

# 1) Definition of the upgrade problem

Set up a study group to:

- Firstly:
  - Establish and update estimates of actual computing requirements for HL-LHC, more realistic than previous estimates:
    - what are the baseline numbers for data volumes/rates, CPU needs, etc.?
  - Build a realistic cost model of LHC computing, help to evaluate various models and proposals – this will be a key to guiding direction of solutions
- Secondly:
  - Look at the long term evolution of computing models and large scale infrastructure
    - Need both visionary “revolutionary” model(s) that challenge assumptions, and “evolutionary” alternatives
  - Explore possible models that address (propose strawman models)
    - Today’s shortcomings
    - Try to use best of evolving technologies
    - Address expectations of how the environment may evolve
      - Large scale joint procurements, clouds, interaction with other HEP/Astro-P/other sciences
    - Possible convergence of (the next generation of) main toolsets

# 2) Software-related activities

- Strengthen the HSF:
  - "Improve software performance" –
    - Need to define what the goals and to define metrics for performance:
      - E.g. time to completion vs throughput vs cost
    - Continue concurrency forum/HSF activities – but try and promote more
    - And other initiatives like reconstruction algorithms etc
  - Techlab
    - expand as a larger scale facility under HSF umbrella
    - Include support tools (profilers, compilers, memory etc)
      - Including support, training, etc
      - openlab can also help here
    - Should be collaborative – CERN + other labs
  - Technology review
    - "PASTA" – reform the activity – make into an ongoing activity, updating report every ~2 years
      - Broad group of interested experts
    - Also under HSF umbrella – strongly related to the above activities
  - What can be done about long term careers and recognition of software development

# 3) Performance evaluation/"modelling"

- Investigate real-world performance of today's systems:
  - Why is performance so far from simple estimates of what it should be?
  - Different granularities/scales:
    - Application on a machine
    - Site level: bottlenecks, large-scale performance
      - Different scale sites, different workflows
    - Overall distributed system
      - At which level?
      - Are data models and workflows appropriate?
- Once we have a better handle of actual performance – can we derive some useful models/parameterisations etc?
  - Useful enough to guide choices of computing models – don't have to be perfect or complete
  - This feeds into any cost models
- Small team in IT starting to work on this and consolidate existing efforts
  - Define a programme of work to look at current performance and concerns; define initial goals

# 4) Prototyping (demonstrators)

- Some specific prototyping of some of the ideas that arise from the above activities
- For example:
  - Data or storage management
    - Storage federations, caches rather than “SE”
    - Etc.
  - Optimisation of sites with little effort or expertise
    - “Site in a box” appliance,
    - What about cache, stage-out, etc
  - Others as ideas arise
- Common activity here would help to evolve into common solutions in production eventually

# Summary

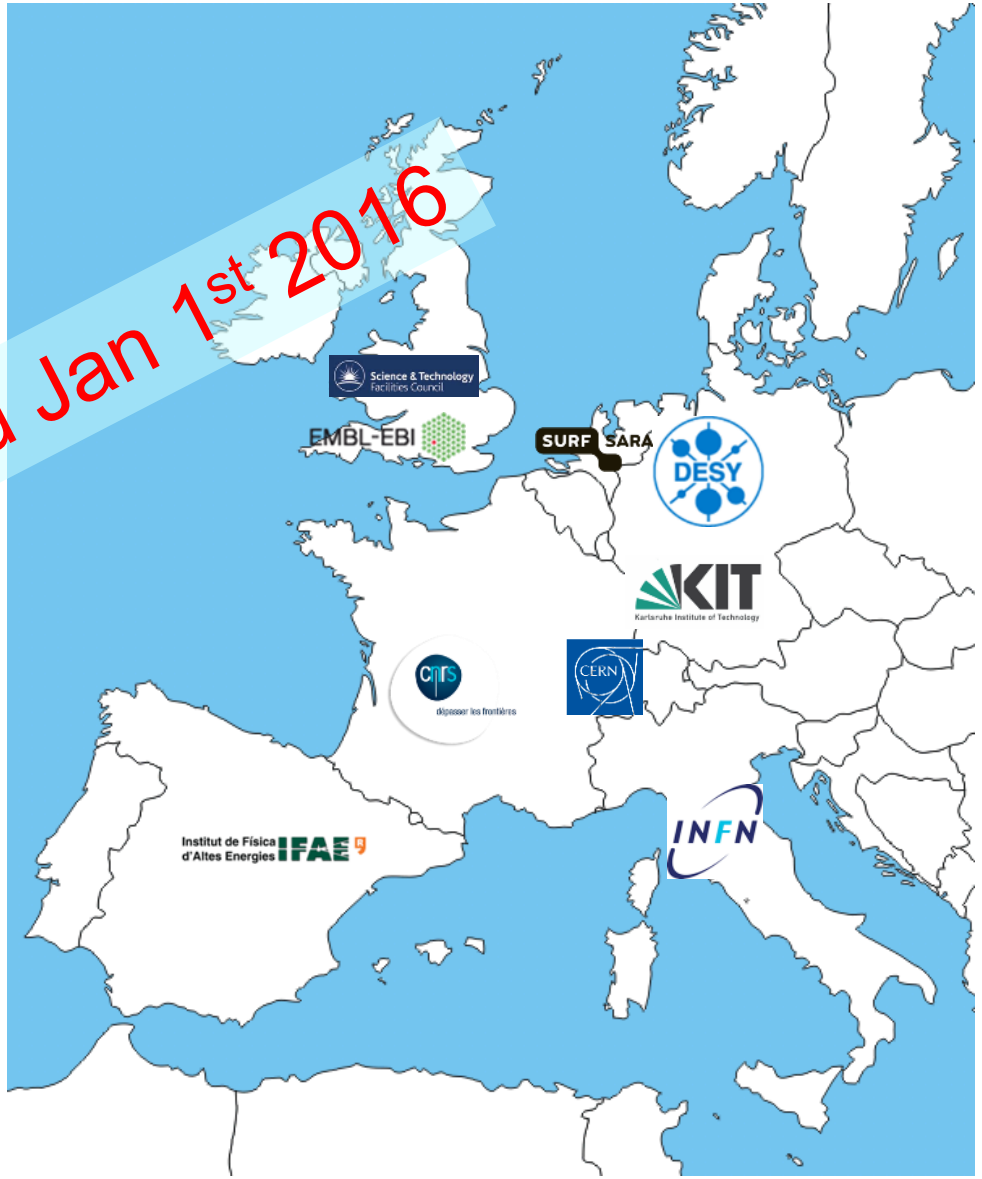
- Medium term
  - A lot of work ongoing
    - Including other aspects not discussed in Lisbon (e.g. cost of operations)
- Longer term
  - 3 areas of work proposed
  - MB will oversee, and define a more concrete plan
- Prototypes/demonstrators
  - A useful way to explore ideas and eventually converge on common solutions?

# HNSciCloud H2020 PCP Project

The group of buyers have committed
- ~1.6M€ of funds (generating ~6M€ total funds)
- Manpower
- Applications & Data
- In-house IT resources

To procure innovative IaaS cloud services integrated into a hybrid cloud model
- Commercial cloud services
- European e-Infrastructures
- In-house IT resources

Procured services will be made available to end-users from many research communities



# European Open Science Cloud

- The European Open Science Cloud is promoted in the context of the movement towards open science bringing ever greater transparency, accessibility and accountability,
  - stakeholders in the research process increasingly expect to be able to access and reuse the outputs of taxpayer funded research.
- The action INFRADEV-04-2016 (10M€ deadline 22 June 2016) foresees the evolution of existing e-Infrastructures into a 'European Open Science Cloud' (EOSC).
  - A pilot action to demonstrate how to make scientific data and data-analysis services more widely available enabling greater data sharing and re-use.
  - EOSC should deliver trusted access to services & systems in a federated environment by leveraging existing services, across Member States and disciplinary, social and geographical borders, where data complies with the "FAIR" principles (Findable, Accessible, Interoperable, Re-usable)
- The EC has set-up a High Level Expert Group on OSC which has consulted widely and will shortly produce a report with recommendations which will influence the direction of the EOSC.
- The scope of EOSC is very broad with ambitious goals
  - this first pilot funding call is very modest. As such it should really been seen as a preparatory phase and what is more important (for CERN) is to ensure that the direction the EOSC takes can serve the physics community.

# EOSC …

- There are a range of opinions about what should be the focus of pilot EOSC within the EC directorates, across the ESFRI research infrastructures and the role of European level e-infrastructures as well as the Commercial sector (both as service providers and users).
- CERN has been discussing these questions with its partners in the context of EU-T0, EIROforum and Helix Nebula
- Clearly, if the EOSC is to achieve the goals outlined by the High Level Expert Group then it will require a significant increase in funding compared to that invested by the stakeholders today.
  - Improving cost-effectiveness through new technology, better governance and innovative business models will certainly help but will not offset all the increased costs of making a wider range of interoperable services available to far more users.
- This is a message that should be reinforced by funding agencies when they meet with the EC in a dedicated meeting on the 15th March.
- Upcoming important events include the conference on open science being organised as part of the Dutch presidency of the EC (April 2016).
- The EC will also issue a communication on the Cloud in the same timeframe.