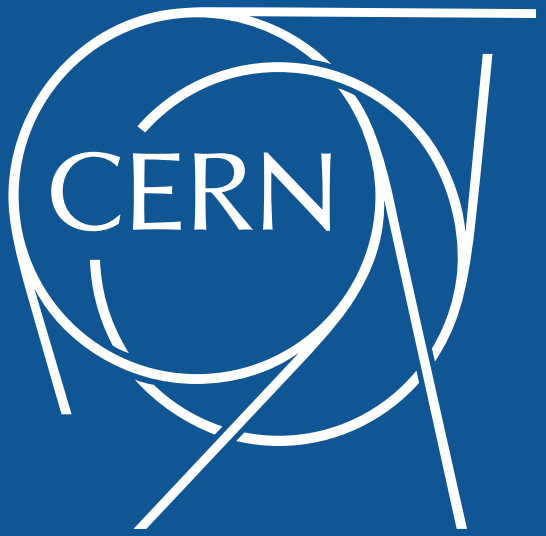

CERN

IT Applications and Data Processing at CERN

David Abdurachmanov (Fermi National Accelerator Laboratory, US)
Justas Balčas (Vilnius University, LT)

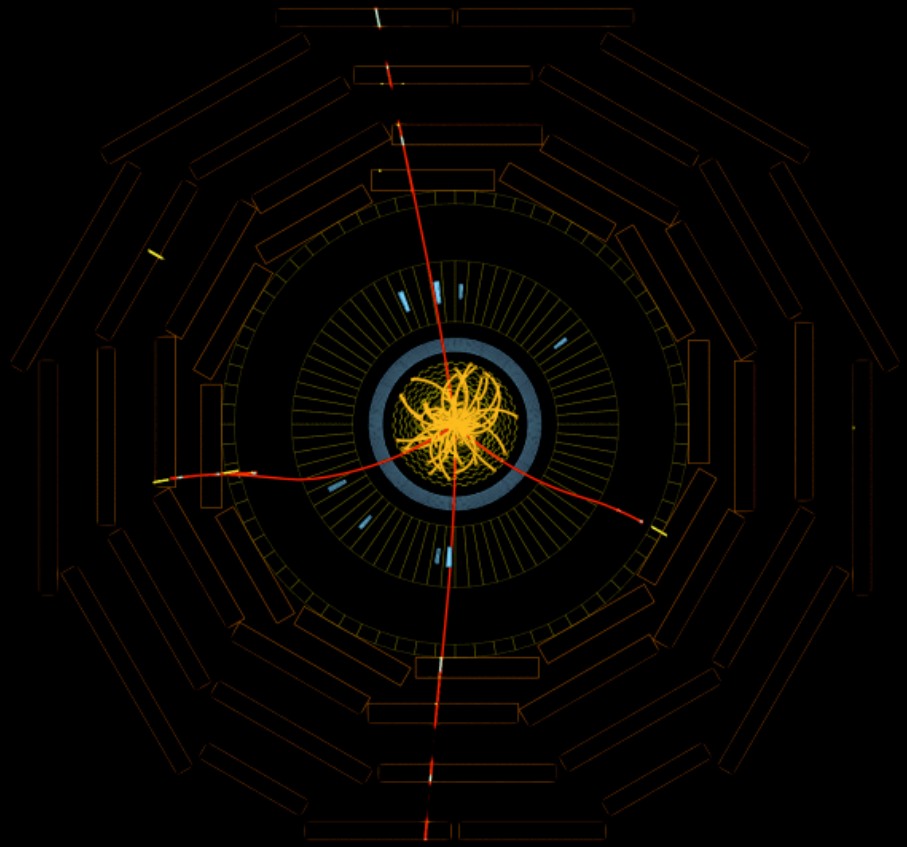
Our Universe is Yours
Notre Univers est le vôtre

An Early “Computer”

- Wim Klein
- Calculating the 73rd root of a 500 digit number took less than 3 minutes (1976)

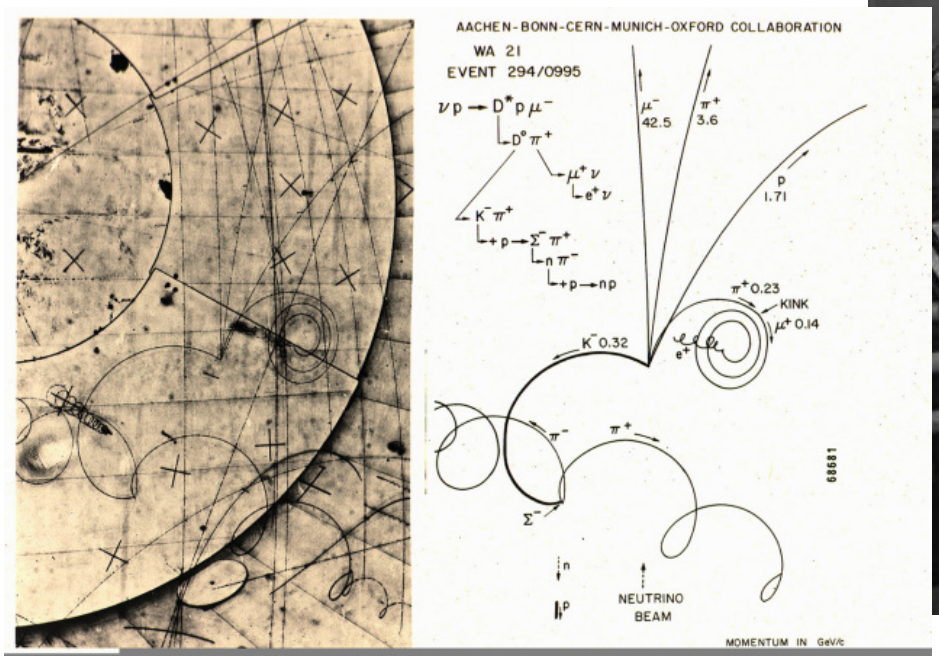
Computing

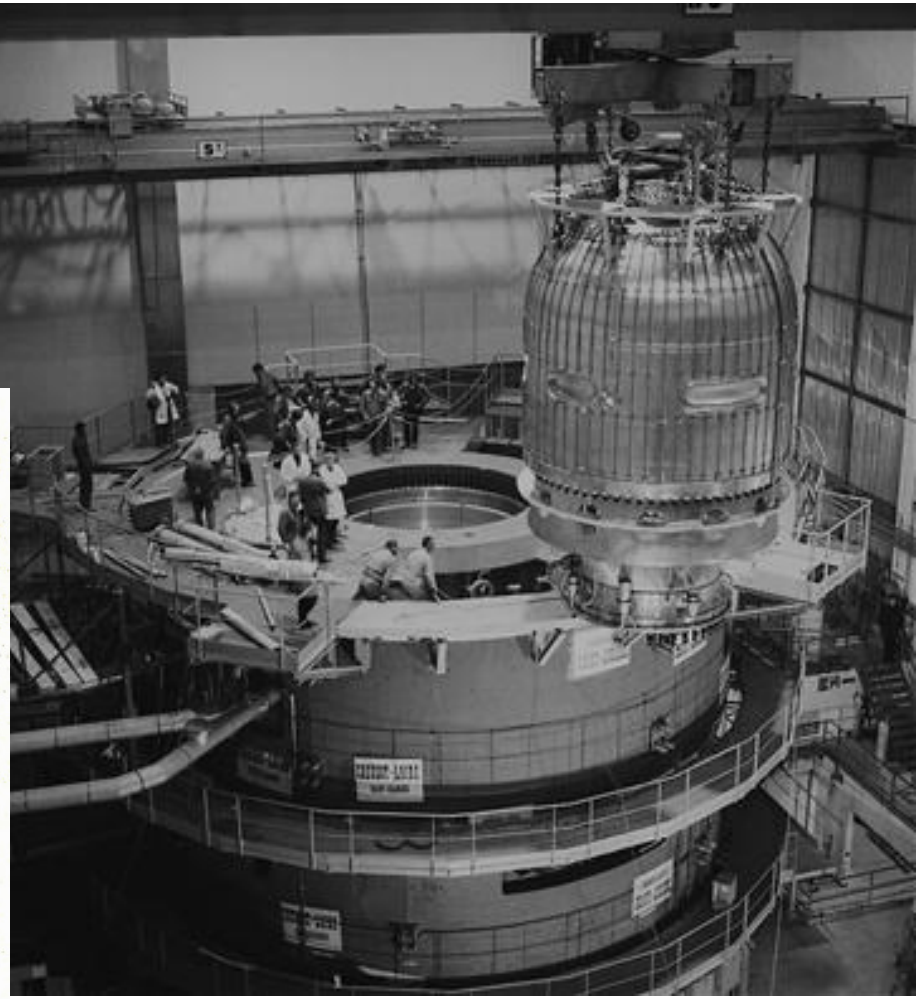
- Compute
- Network
- Storage
- Technical
- Collaboration



Bubble Chamber

- BEBC 1973-1984
- 6.3 million photos
- 3000 km of film





Momentous Events

1960: Vacuum tubes

1965: CDC 6600: #3

Mainframes

New Computer Centre for a New Computer

1983: 2 CDCs and an IBM 3081

Super Computers

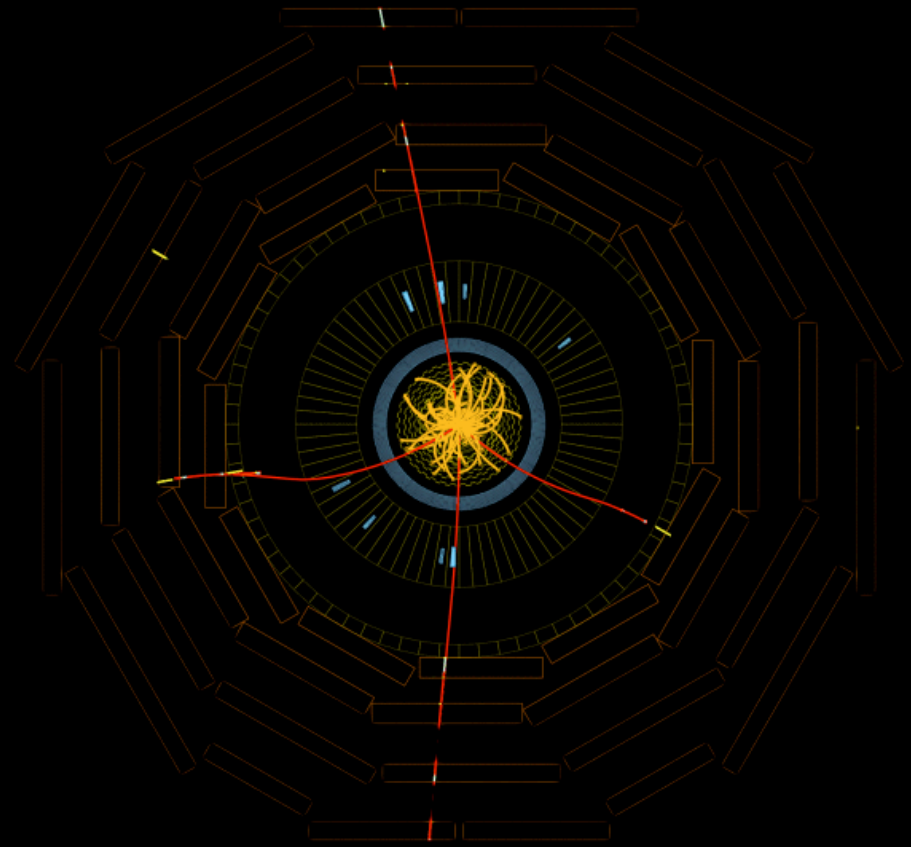
1988: Cray XMP

Comodity Computing

Farming in a Data Centre

Computing

- Networking



Networking

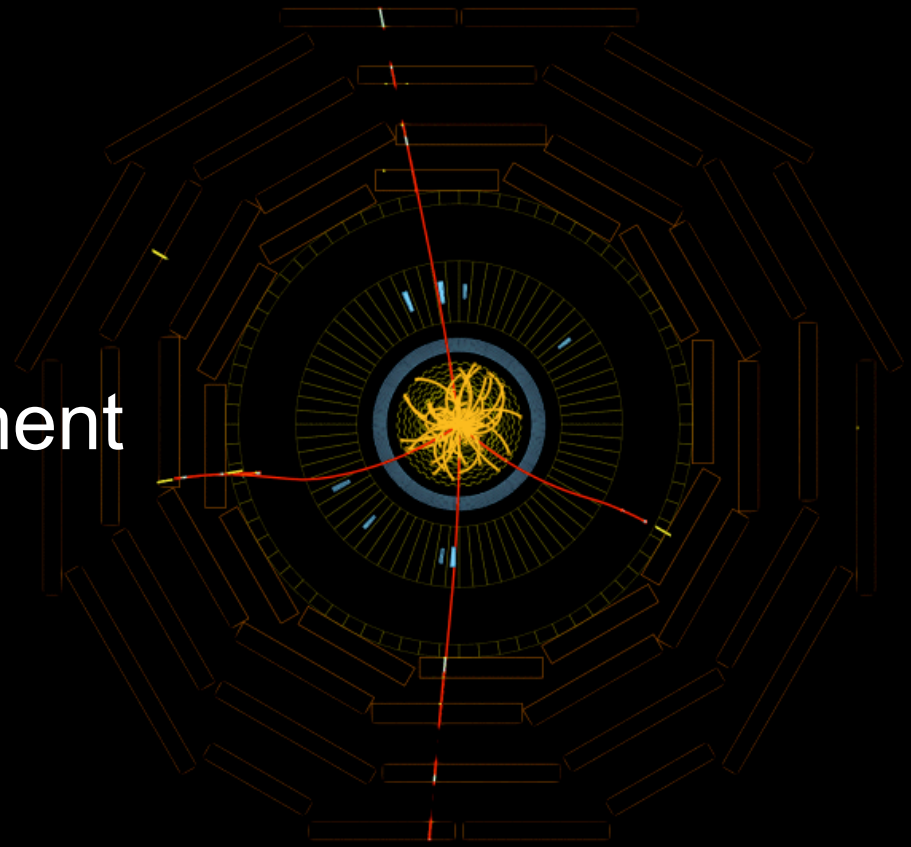
- Packet-switched network
- 1969 ARPANET
  - US DOD sponsored research at US Universities
- 70s and 80s proliferation
  - US: NASA Science Net, CSnet, Energy Sciences Net, NSFnet
  - FR: CYCLADES
  - UK: Mark I, SERCnet
  - Commercial: Tymnet, CompuServ, BITnet, DECnet
  - Protocols: NCP, X.25, TCP/IP
  - CERnet

Networking

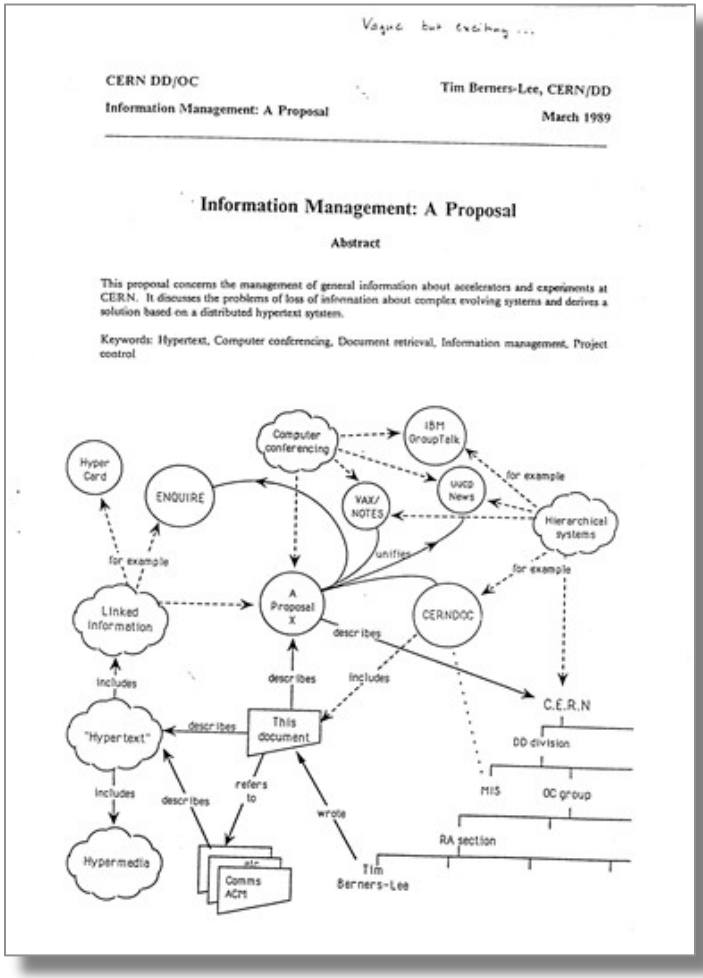
- Science without borders
  - Data exchange across the iron curtain
  - 1988 first data connection between China and scientific world – IHEP to CERN
- Truly international Internet
  - 1989 first external TCP/IP connection
  - 1990 principle link US-EU from CERN
    - (1.5Mb/s)
  - 1991 80% of the internet capacity installed in Europe for international traffic was terminated at CERN

Computing

- Information Management



Vague but Exciting …



Vague but exciting ...

CERN DD/OC

Tim Berners-Lee, CERN/DD

Information Management: A Proposal

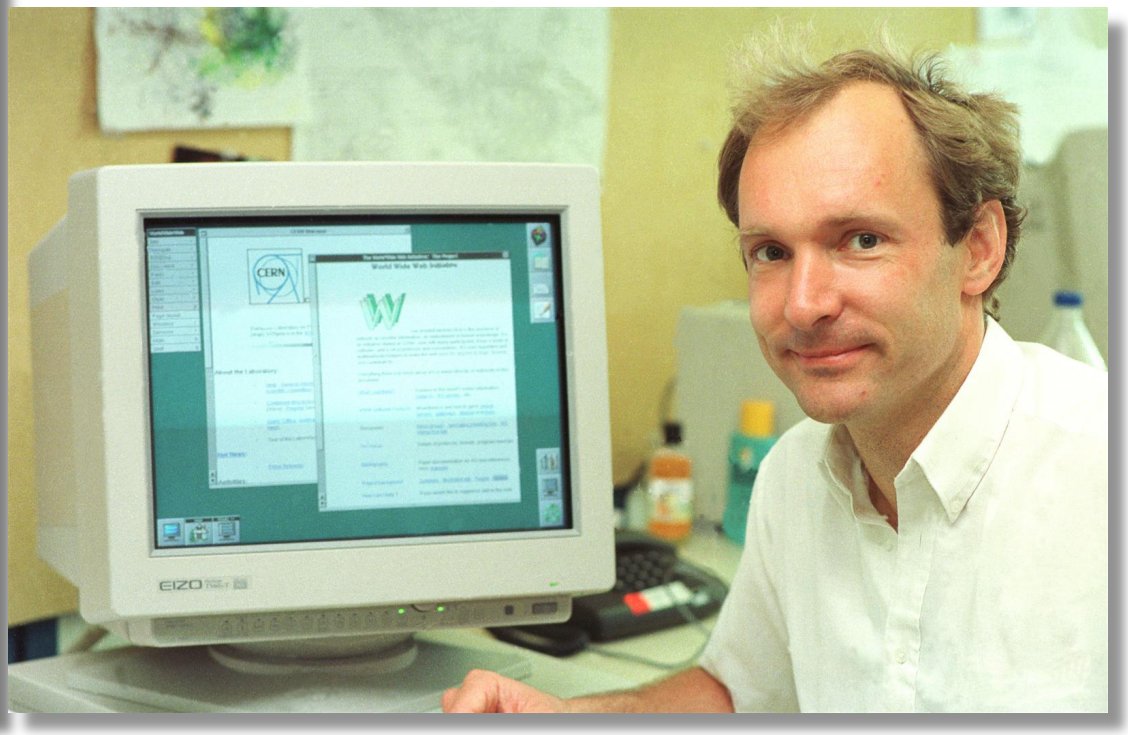
March 1989

Information Management: A Proposal

Abstract

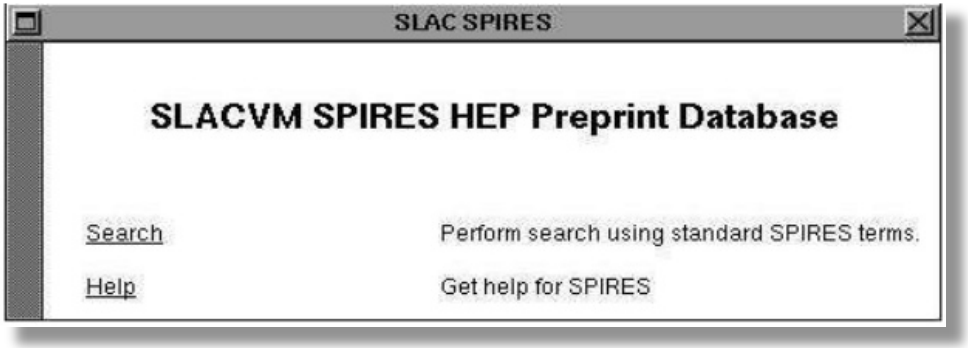
This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

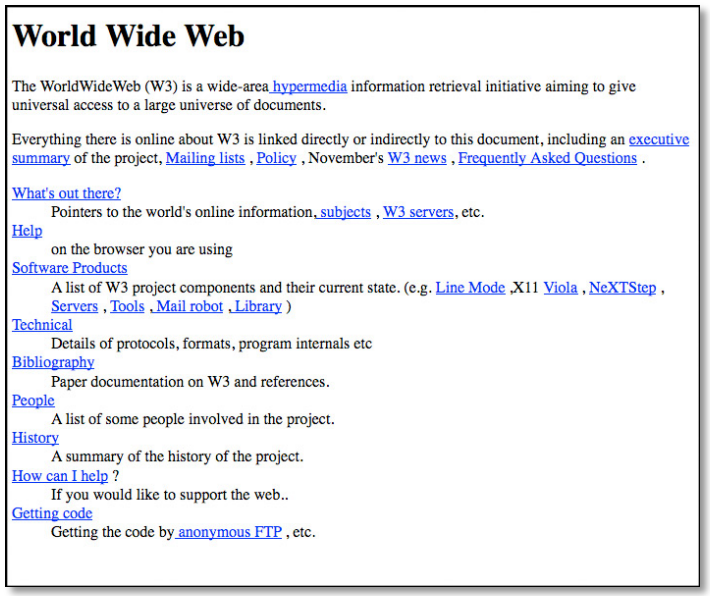
Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control



Growth of the Web

- Aug 1991 went public
  - Tim Berners-Lee posted project to alt.hypertext newsgroup and other internet groups





World Wide Web

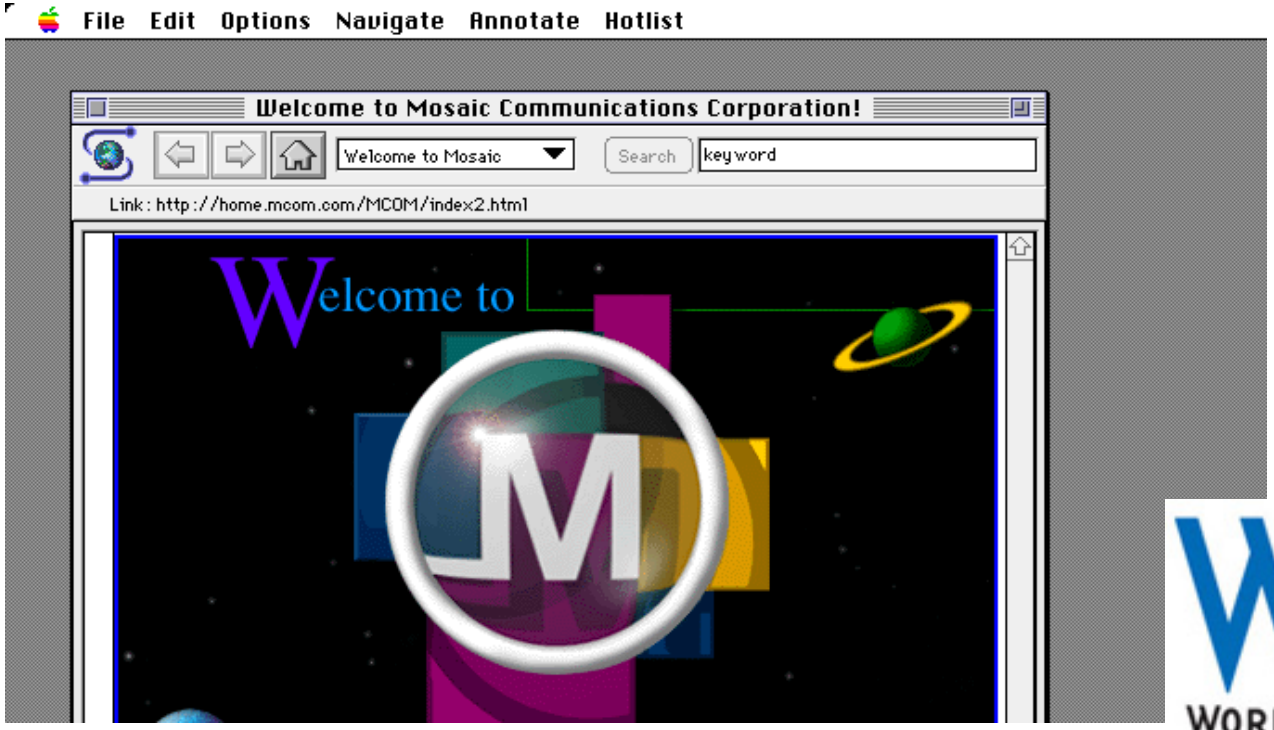
The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary of the project, Mailing lists , Policy , November's W3 news , Frequently Asked Questions .

What's out there?
Pointers to the world's online information, subjects , W3 servers, etc.
Help
on the browser you are using
Software Products
A list of W3 project components and their current state. (e.g. Line Mode ,X11 Viola , NeXTStep , Servers , Tools , Mail robot , Library)
Technical
Details of protocols, formats, program internals etc
Bibliography
Paper documentation on W3 and references.
People
A list of some people involved in the project.
History
A summary of the history of the project.
How can I help ?
If you would like to support the web..
Getting code
Getting the code by anonymous FTP , etc.

- Dec 1991 First web server in US
- 1992 rapid expansion in HEP
  - Universities and research institutes put up web servers

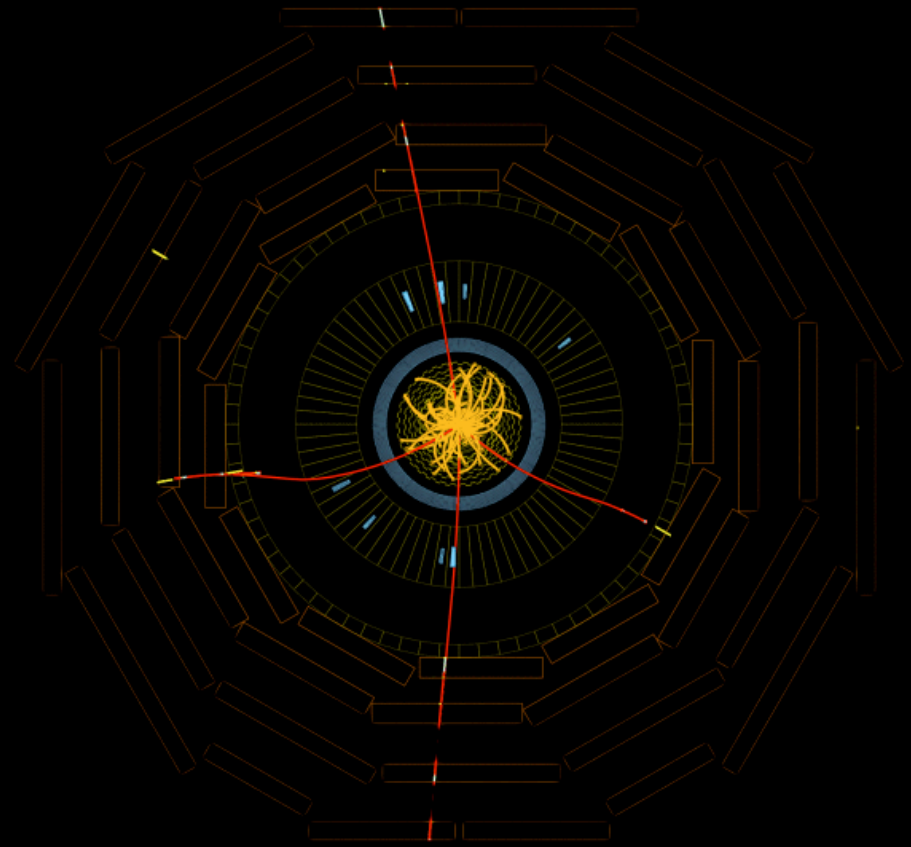
Growth of the Web



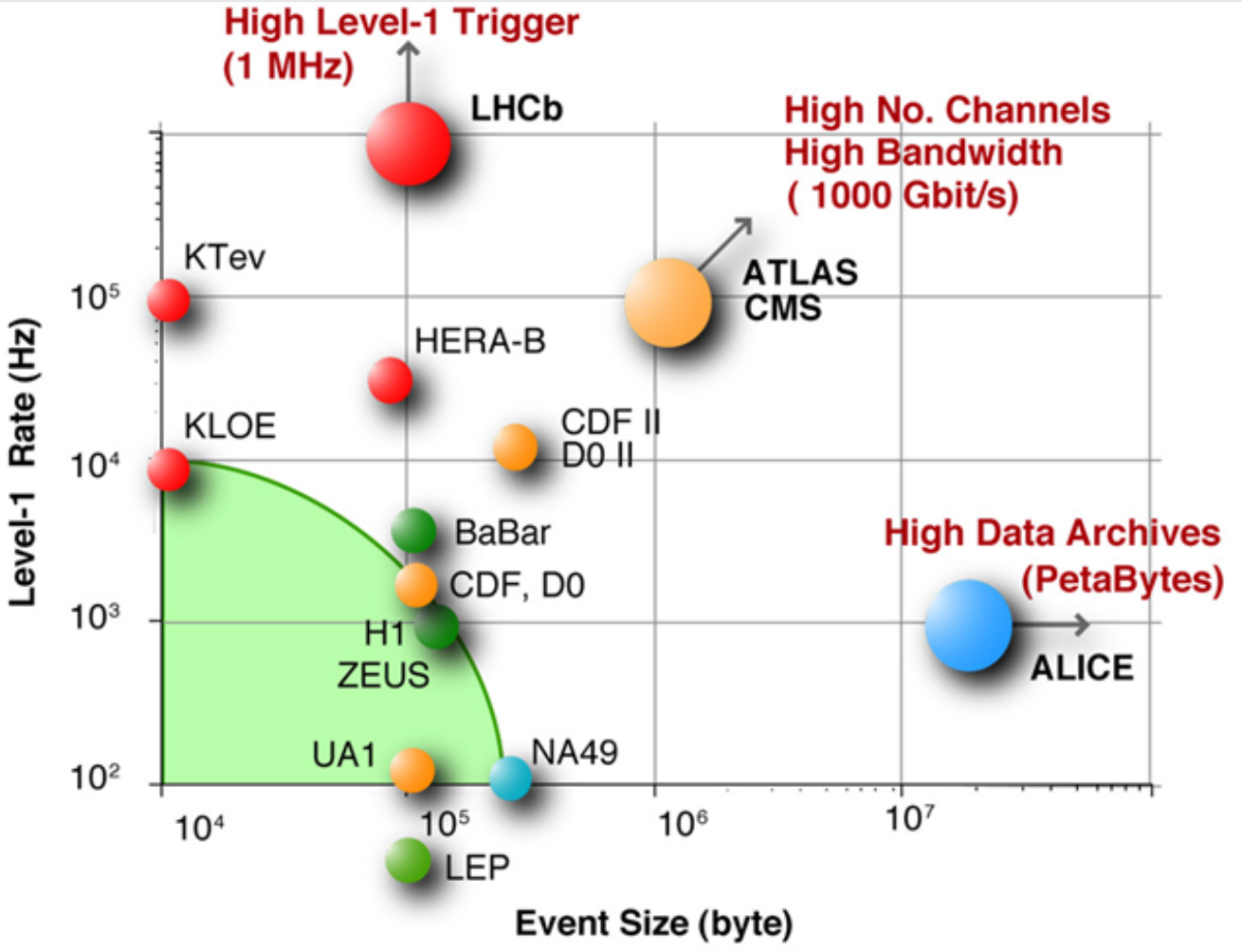
- 1993 rapid expansion across the world
  - National Center for Supercomputing Applications (NCSA) at the University of Illinois released its Mosaic browser

Computing

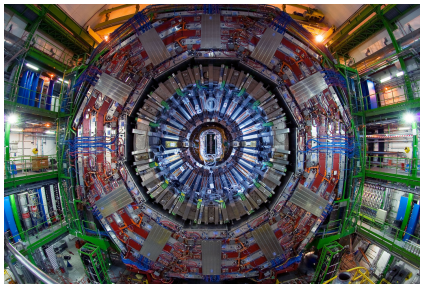
- LHC Era



The LHC Data Challenge



Big Data !



150 million sensors
Generating data 40 million times per second

➜ **Peta Bytes / sec !**

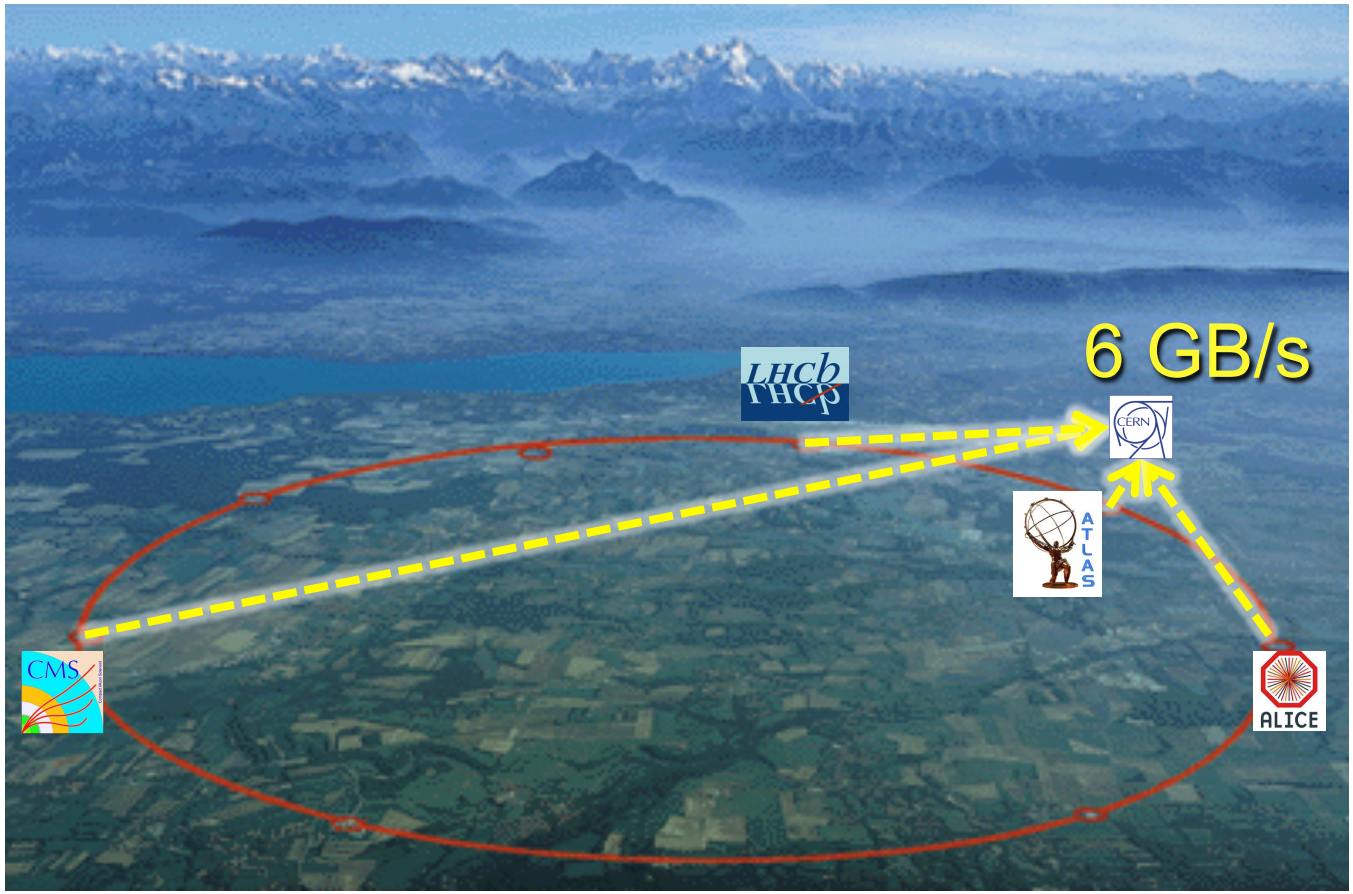
L1 Trigger
Select 100,000 per second

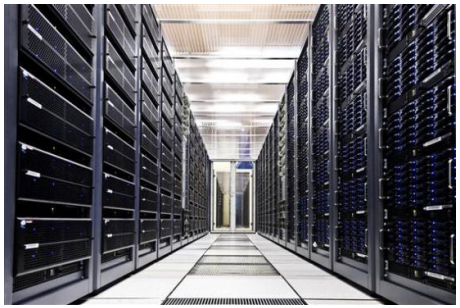
➜ **Tera Bytes / sec !**

HLT / Filter
Select 100 per second

➜ **Giga Bytes / sec !**

Primary Storage





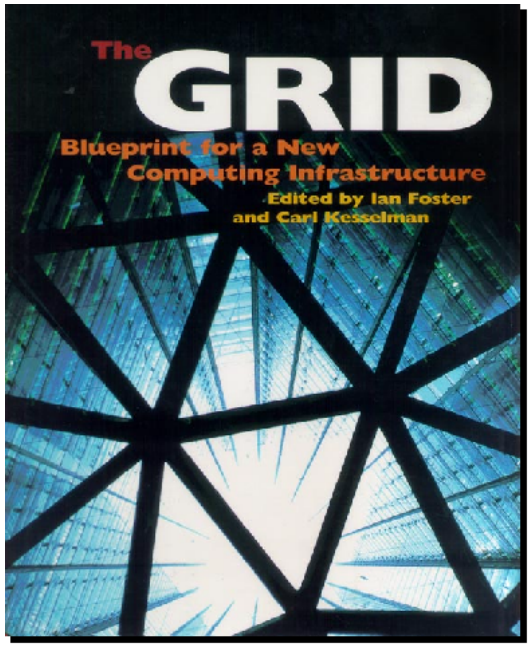
80,000 Disks
88,000 CPU Cores

18,000 1GB NICs
2,700 10GB NICs

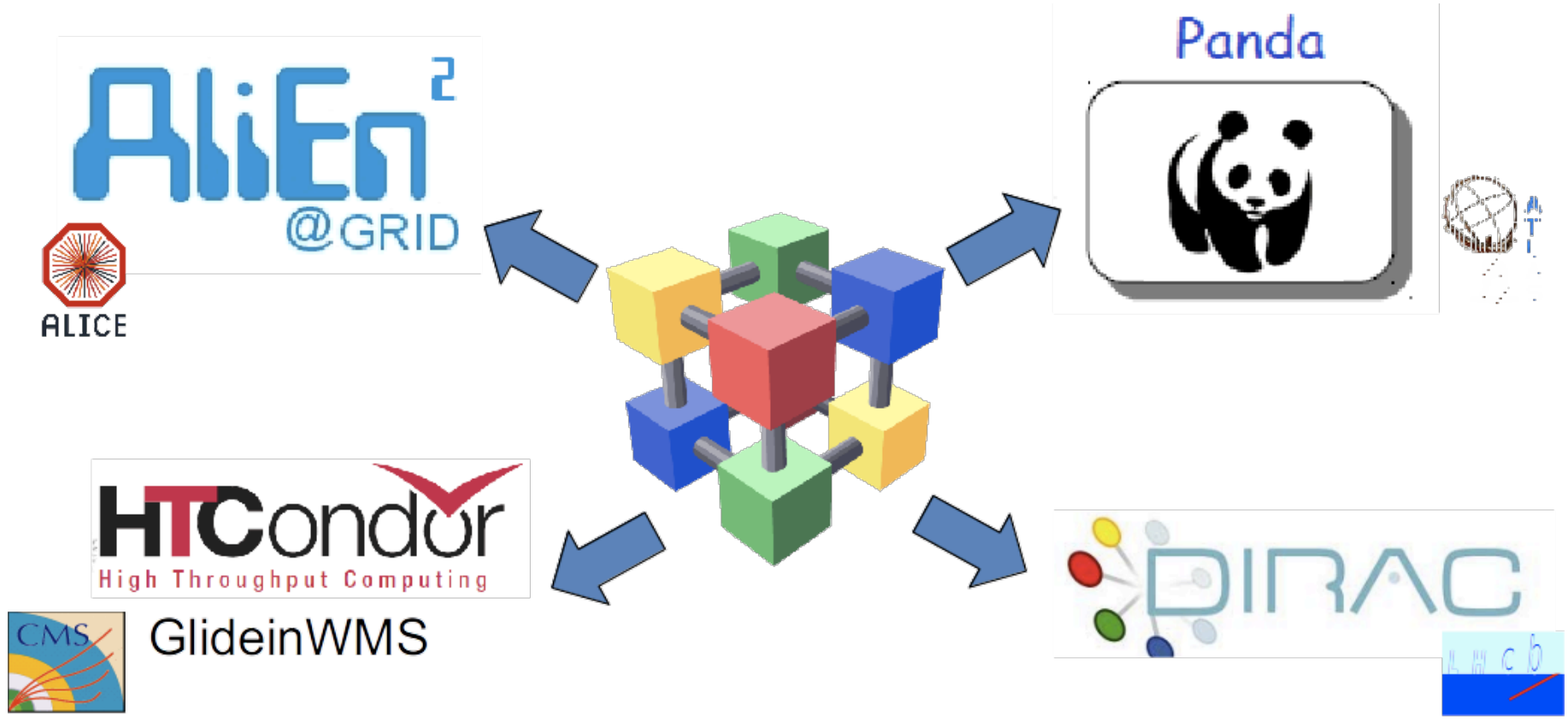
Tier-0 Extension for CERN

The Grid

- Use the Grid to unite computing resources of particle physics institutes around the world

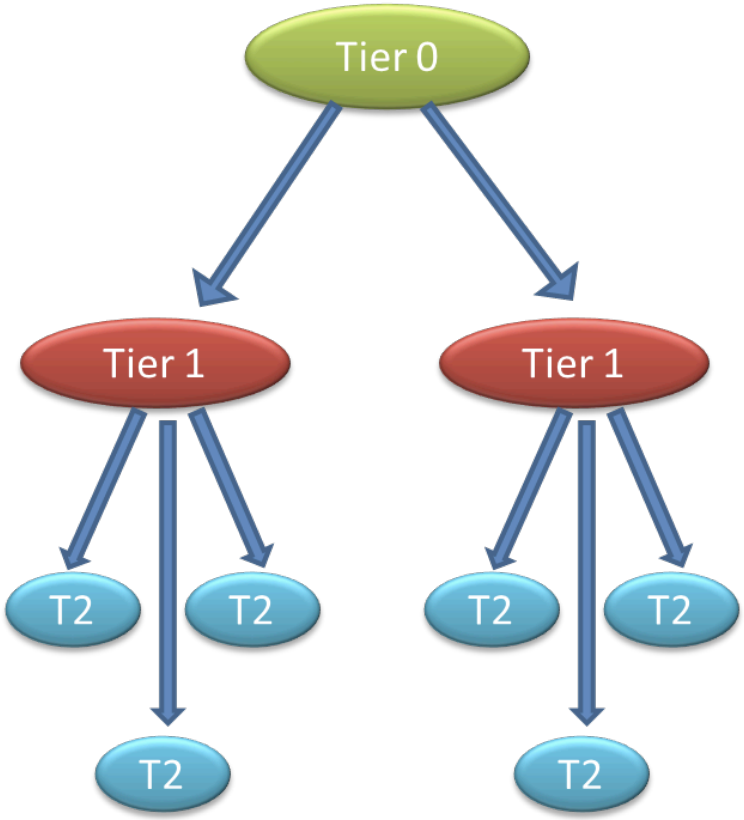


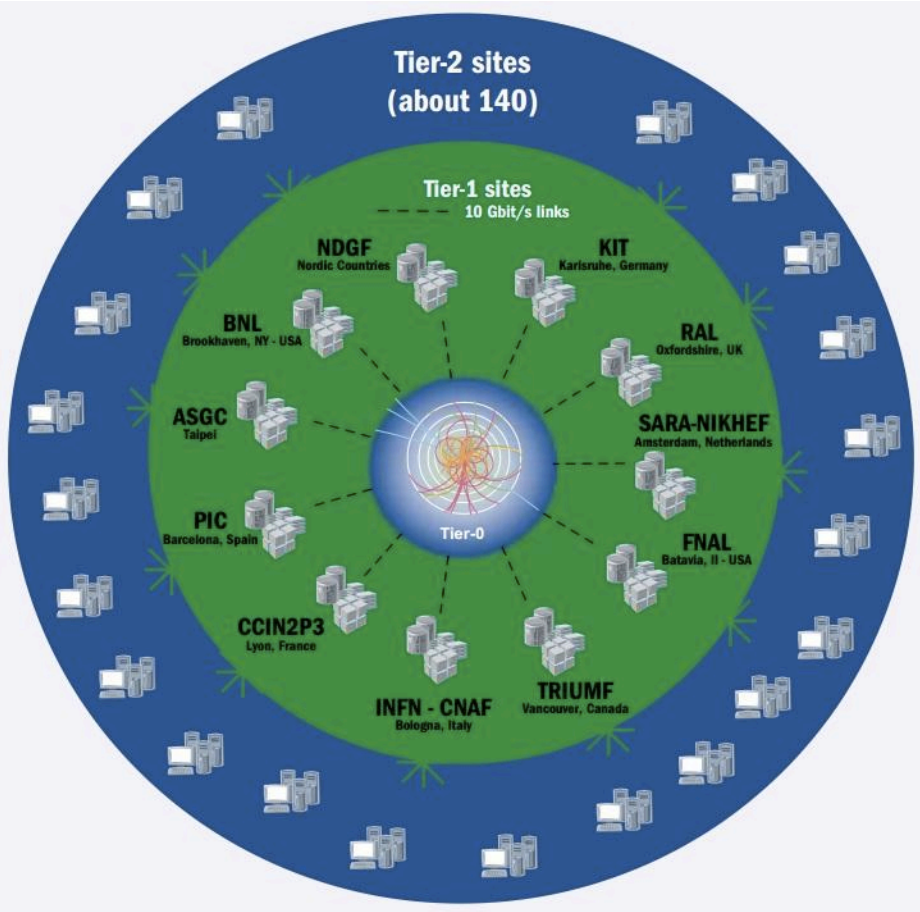
Grid at CERN



- In all cases the underlying distributed computing architecture is similar based on “pilot” job model
  - Can we scale these systems expected levels?

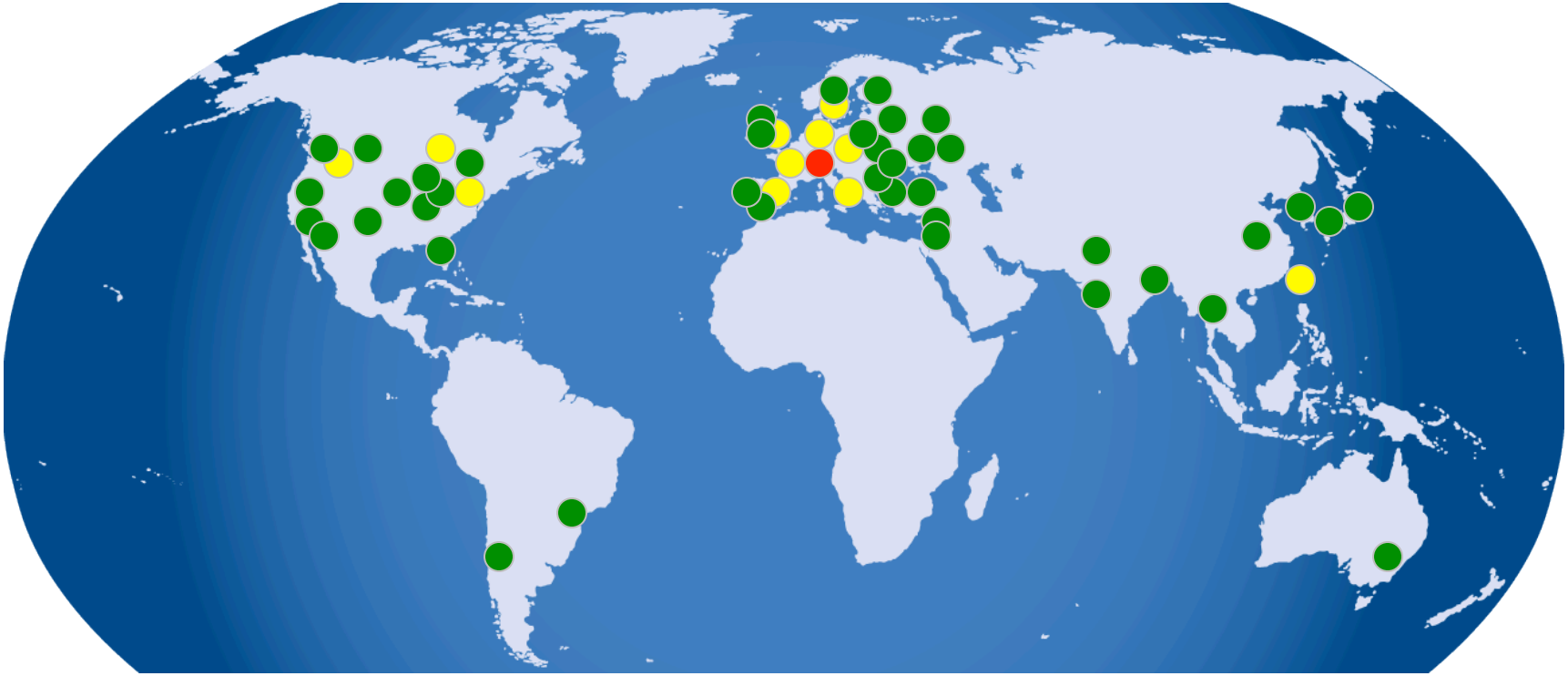
WLCG Grid model





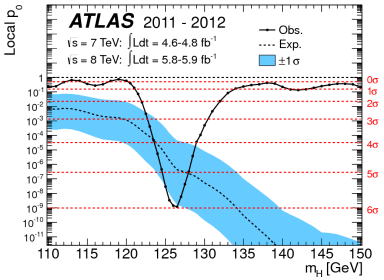
Data Inflation

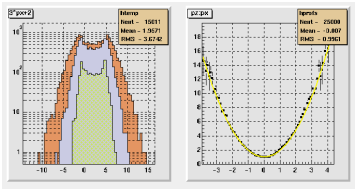
- Static storage: raw and reconstructed data
- Dynamic storage: analyzing it means transformation, reduction, transportation, replication, regeneration

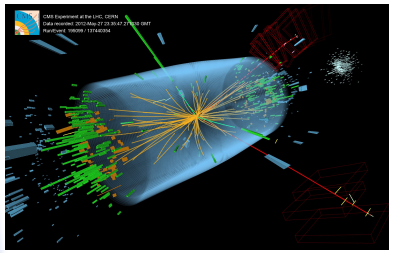


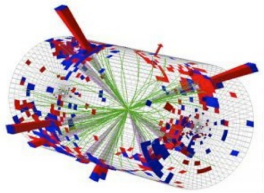
Data Reduction / Analysis

- Publication
- Reduced
- Reconstructed
- Raw









File Size

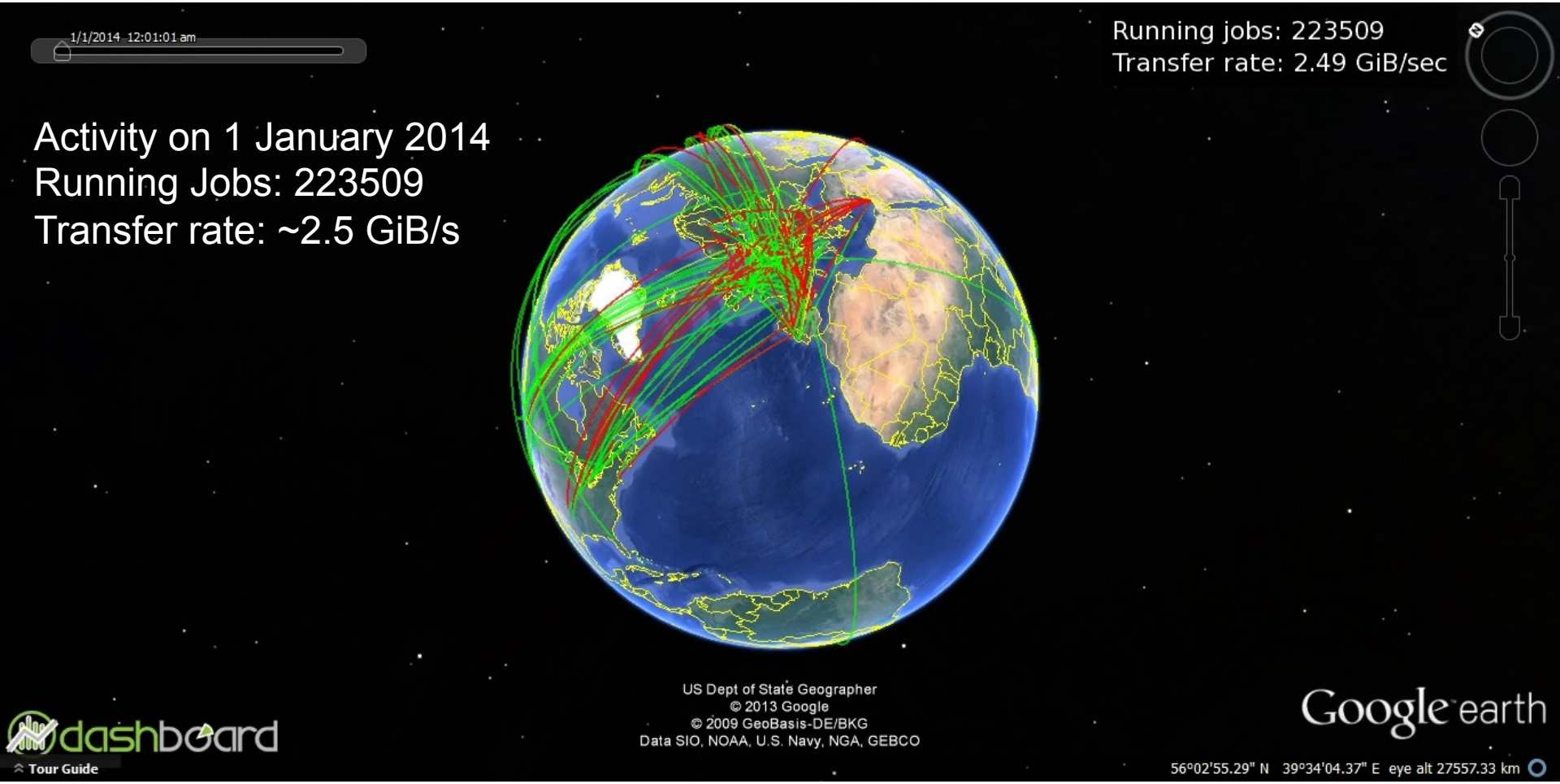
# of Files

Researchers (T1-T3)

Analysis Coordinators (T1)

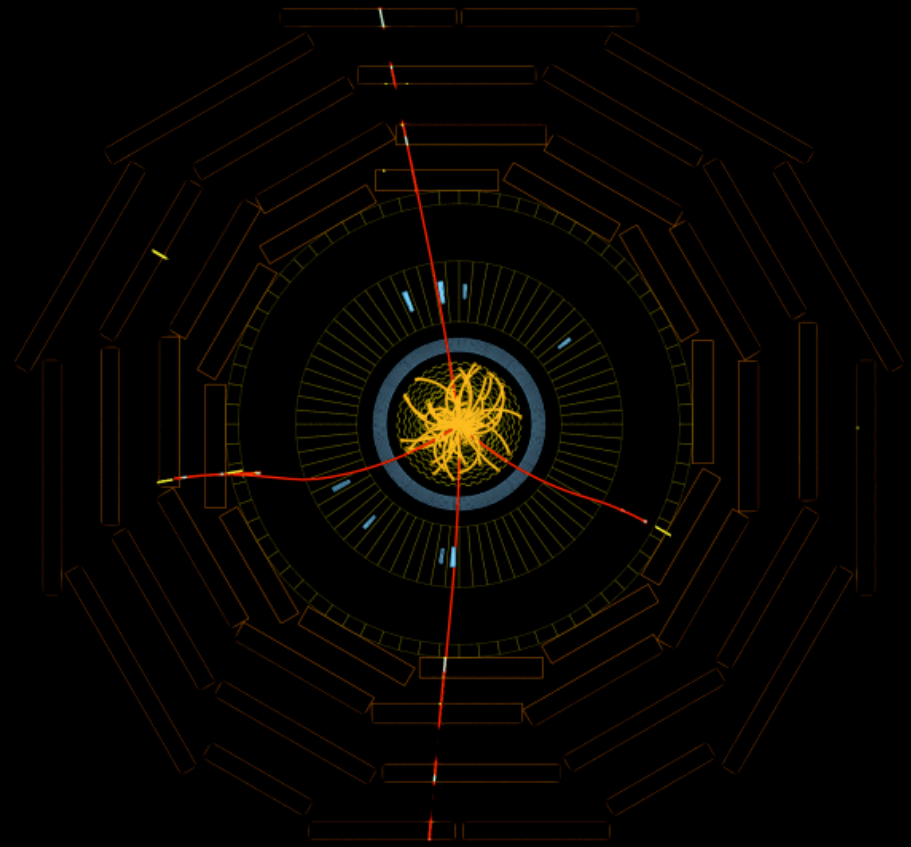
Production Managers (T0-T1)

The Grid that Never Sleeps



Computing

- Technical



Open Source Software

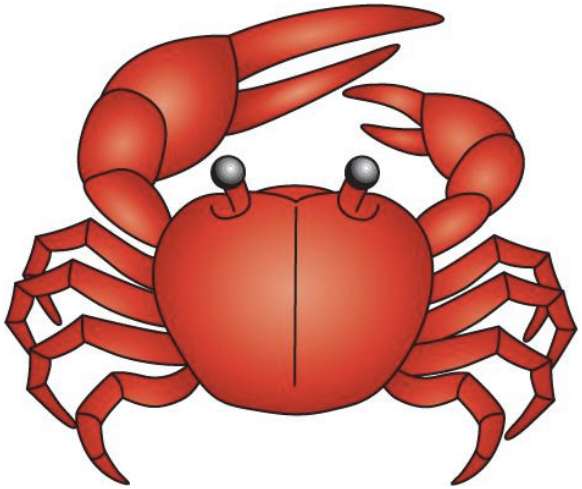
- Intellectual Property
  - Restrictions, restrictions, …
- Public Domain
  - A vast common good
  - Use, enjoy, share and build upon
  - IP expired, forfeited or inapplicable

CMS Software

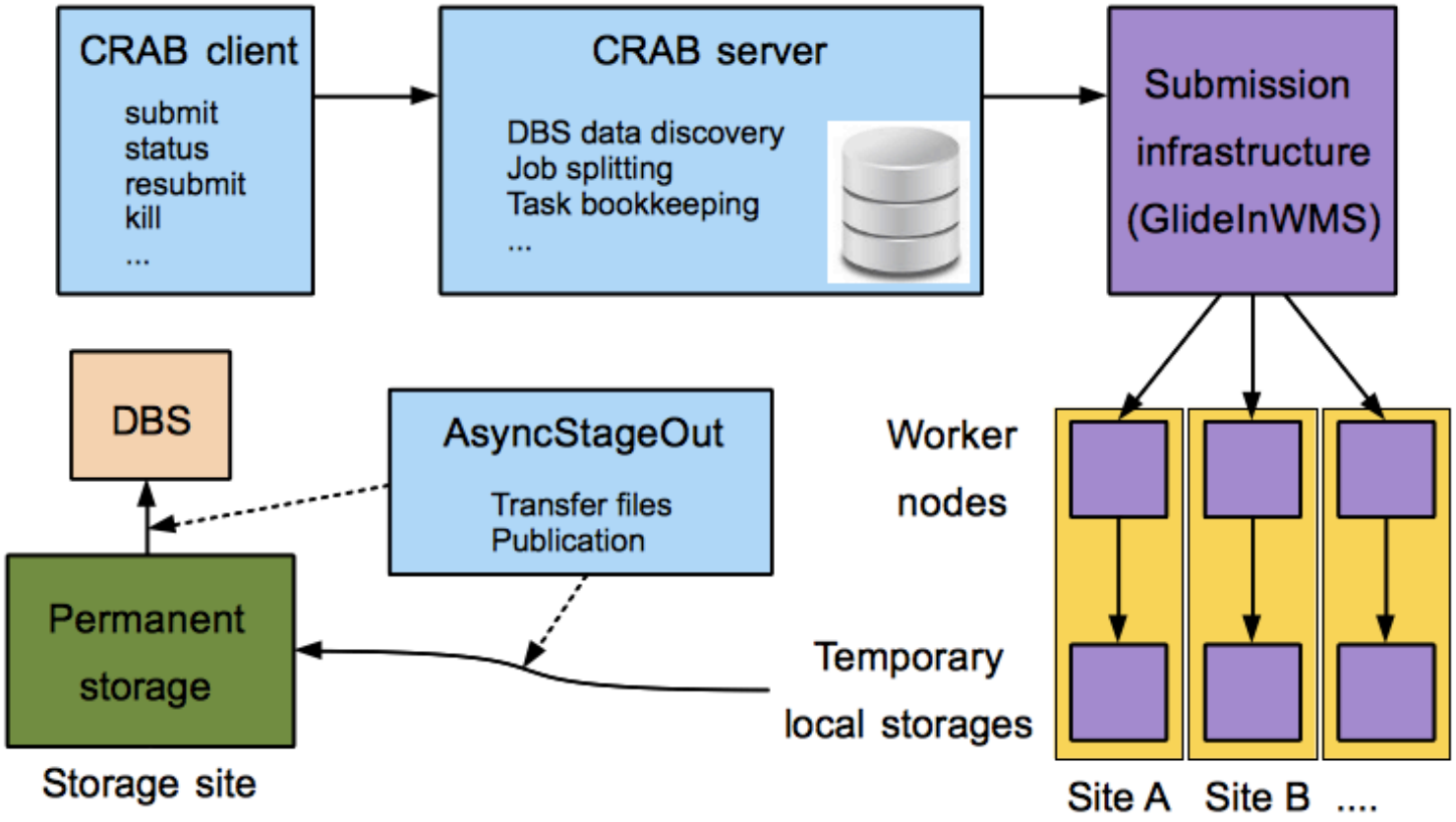
- Over 5 million lines of custom code
  - Developed over the last two decades
  - Over 1300 researches ranging from MSc students to Nobel prize candidates
  - Small core of professional software engineers mostly doing application framework, release integration and QA
  - Open source software (available on GitHub)
- Large working dataset
  - ~12 PB of RAW data acquired in 2011-13, more than double that if we count the processed via the WLCG using 100 000 cores sparse over 5 continents
  - We expect a 2-3 order of magnitude increase in data volume in the next 10 - 15 year

Users analysis (CRAB3)

- What is CRAB ?
  - CRAB: CMS Remote Analysis Builder.
  - In a few very generic words, CRAB is the software & hardware infrastructure used by CMS to submit jobs to the LHC Computing Grid.



Users analysis (CRAB3)



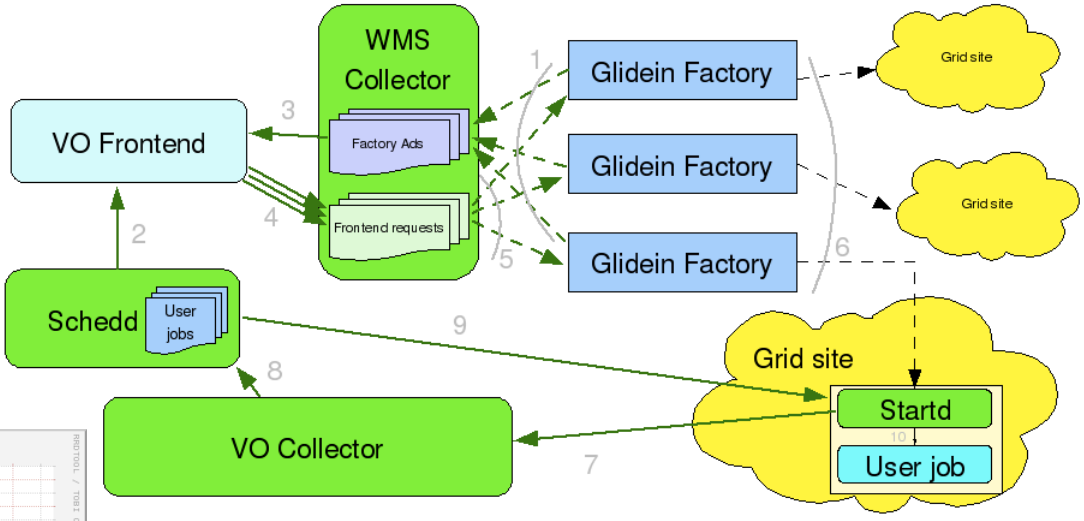
Dropbox

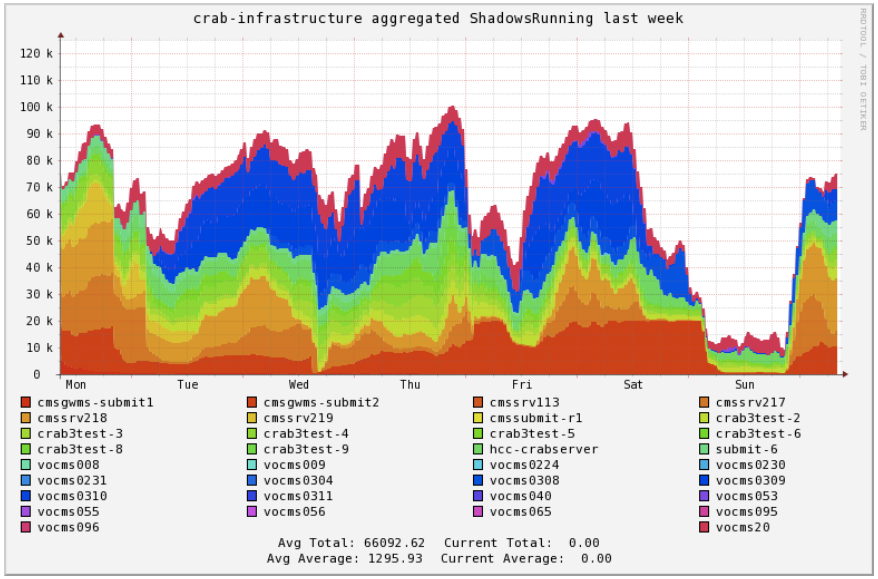
amazon web services EC2

***AAA* (*Any time*, *Any place*, Any where)**

Who runs jobs on the Grid?

- WMAgent
  - Workload Manager Agent
- CRAB3
  - CMS Remote Analysis Builder

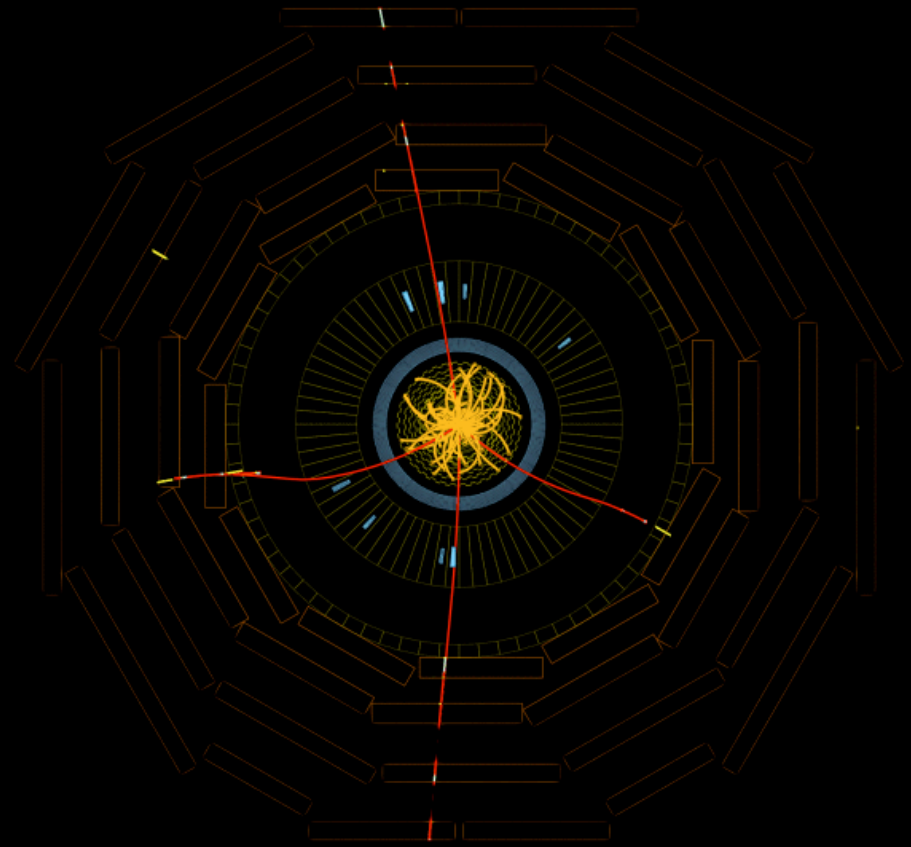




- Volunteer Computing:
  - CMS@Home
  - Amazon Cloud (EC2)

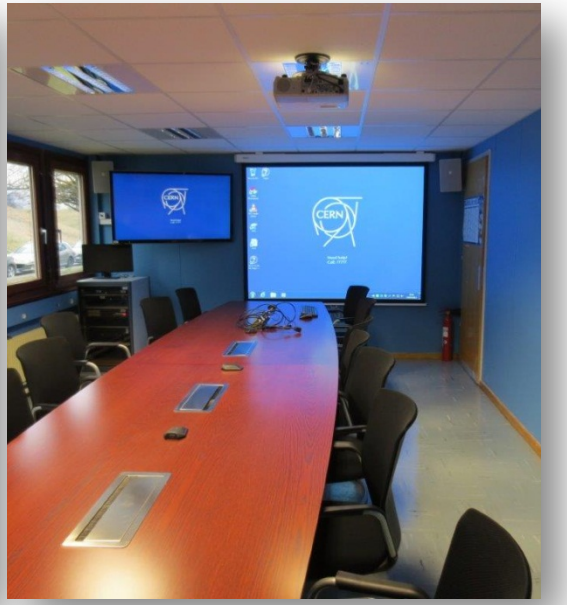
Computing

- Collaboration

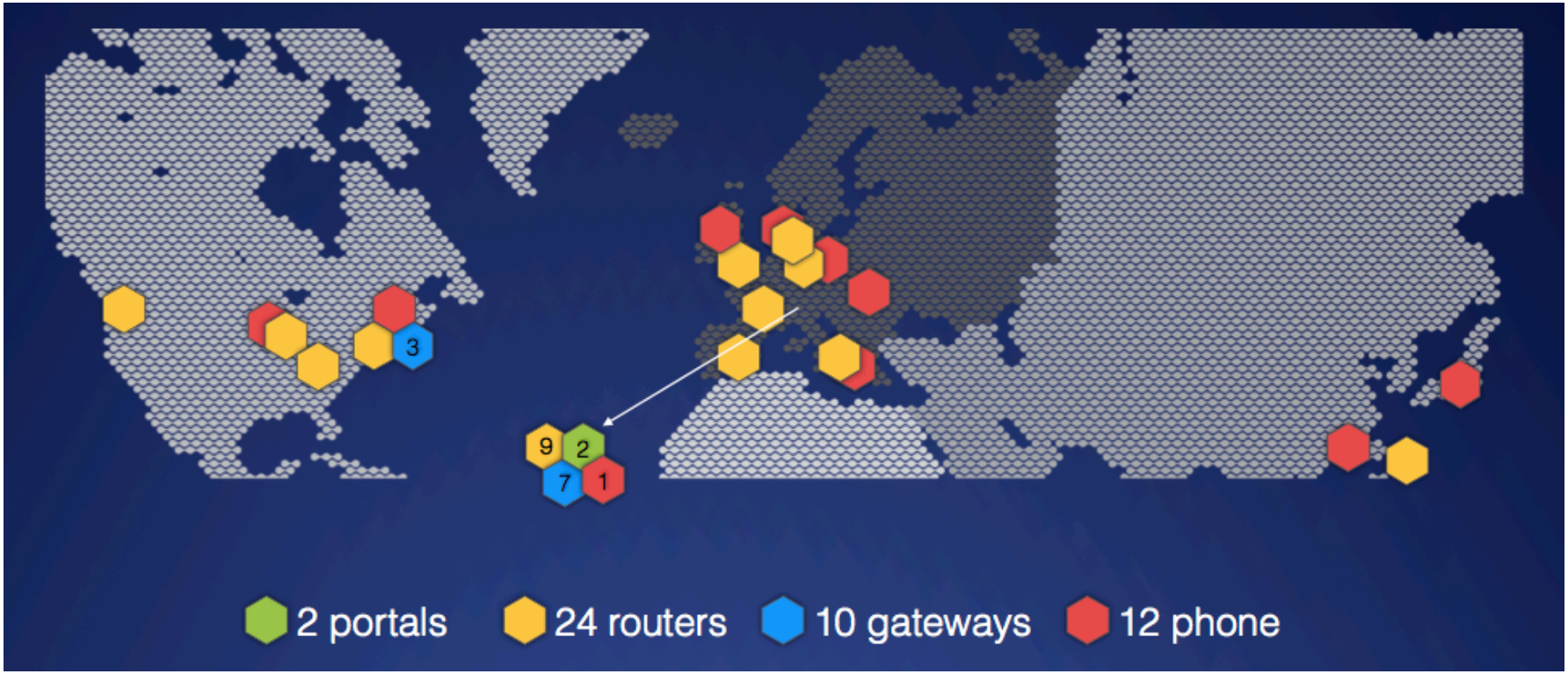


Videoconference

- 250 meeting rooms of all sizes on site
  - 90 equipped for video conference
    - Legacy + VidyoPanorama
  - 13 equipped for VC + Webcast
- 500 legacy endpoints worldwide
  - Non centrally managed

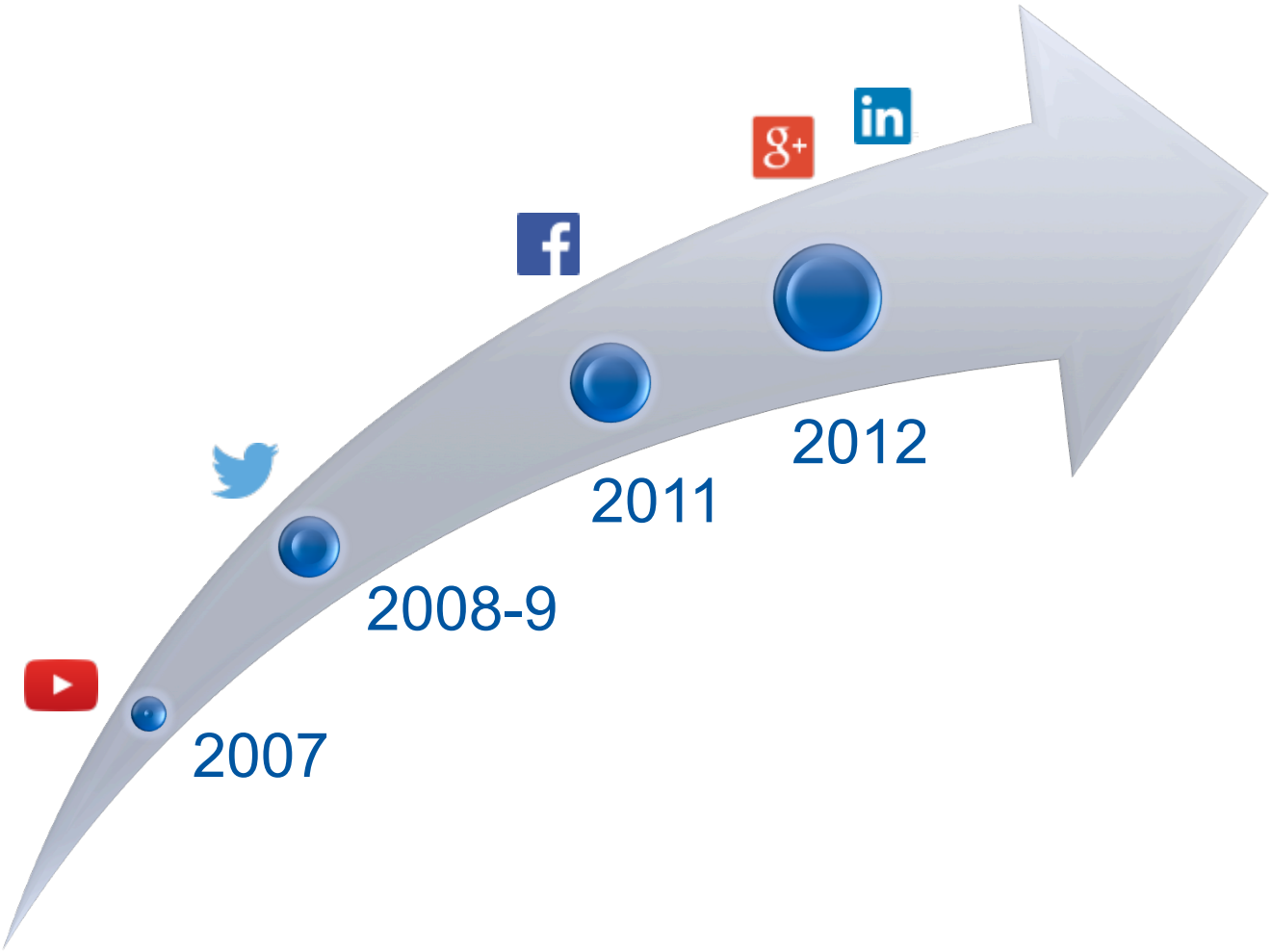


CERN Vidyo Worldwide Service



- 3200 meetings/month
- 852 simultaneous connections
- 252 in one meeting
- 50M minutes last year / 40k downloads

CERN's social media

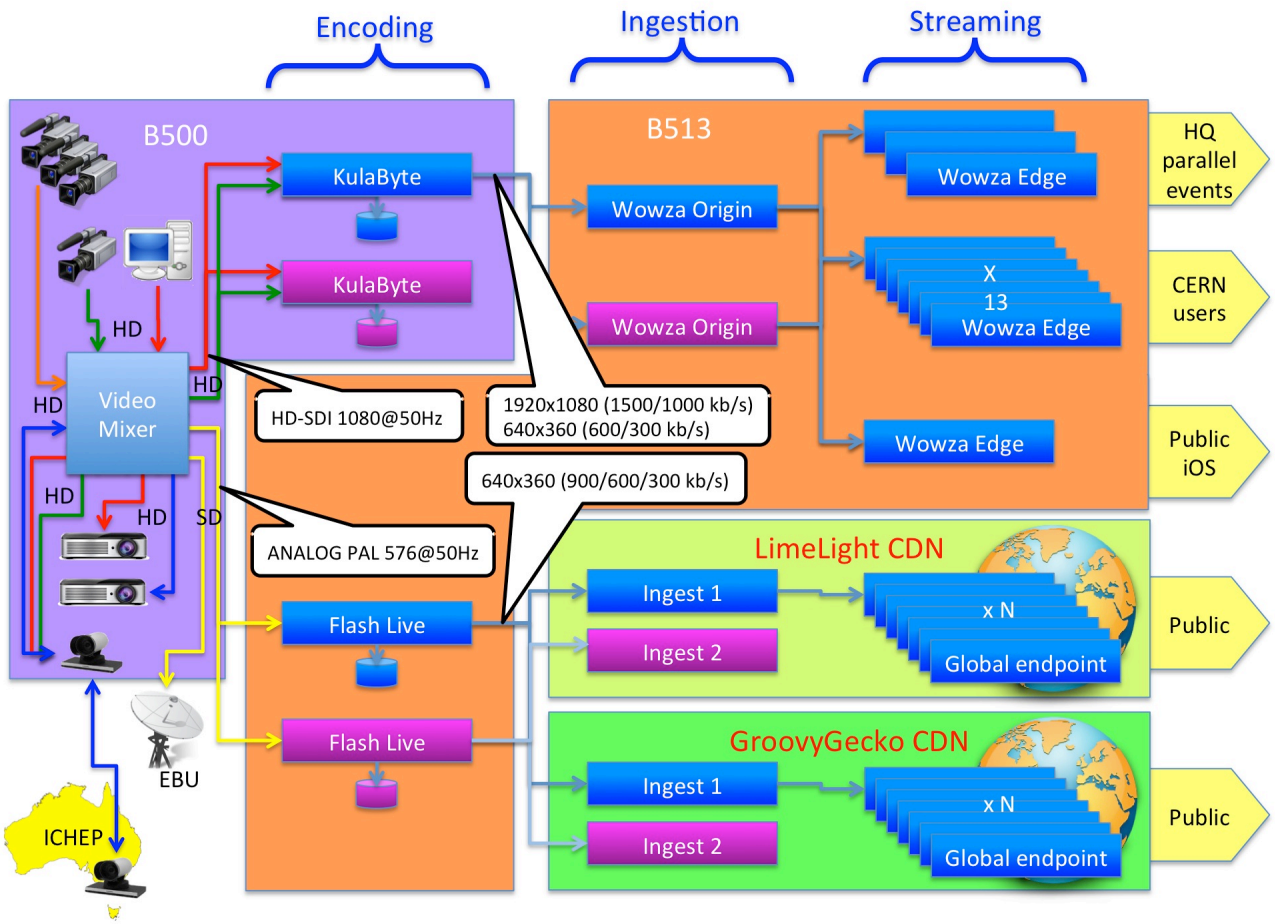


975K
309K
92K
40K
20K
12K

Impact

Most Effective International Organisations on Twitter
AVERAGE NUMBER OF RETWEETS PER TWEET
130
@CERN
European Organisation for Nuclear Research
100
@unicef
United Nations Children's Fund
82
@un
United Nations Organisation
69
@WWF
World Wide Fund for Nature
68
@greenpeace
Greenpeace

Behind Open Science



Biggest Scientific Event ever!

WebCast **~Million**

TV **~Billion**

Links (I)

- Contact:
  - David.Abdurachmanov@cern.ch
  - Justas.Balcas@cern.ch
- More information:
  - IT Department: http://information-technology.web.cern.ch
  - The LHC Grid: http://wlcg.web.cern.ch
  - Google Street view in CC:
    - https://www.google.ch/maps/@46.232624,6.045747,3a,75y,162.48h,90t/data=!3m5!1e1!3m3!1sBU7JKhoaY_H9JVPFHcH8JA!2e0!3e5?hl=en
    - http://lego-scavenger-hunt.web.cern.ch
  - IT Archives: https://it-archives.web.cern.ch

Links (II)

- Social Media at CERN
  - http://twitter.com/CERN
  - http://twitter.com/CERN_FR
  - http://facebook.com/cern
  - http://google.com/+CERN
  - http://youtube.com/CERN
  - http://linkedin.com/company/cern

Questions?

www.cern.ch