
CERN

Science and society: The impact of computing at CERN

Bob Jones
Head of CERN openlab

CERN was founded 1954: 12 European States

"Science for Peace"

Today: 21 Member States

Member States: Austria, Belgium, Bulgaria, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Israel, Italy, the Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland and the United Kingdom

Candidate for Accession: Romania

Associate Member in Pre-Stage to Membership: Serbia

Applicant States for Membership or Associate Membership: Brazil, Cyprus, Pakistan, Russia, Slovenia, Turkey, Ukraine

Observers to Council: India, Japan, Russia, Turkey, United States of America; European Commission and UNESCO

The Mission of CERN

- **Push back** the frontiers of knowledge

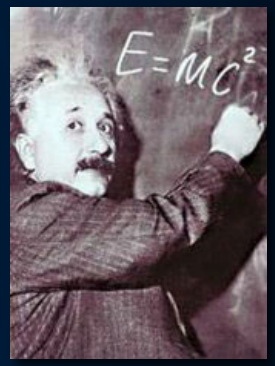
  E.g. the secrets of the Big Bang ... what was the matter like within the first moments of the Universe's existence?

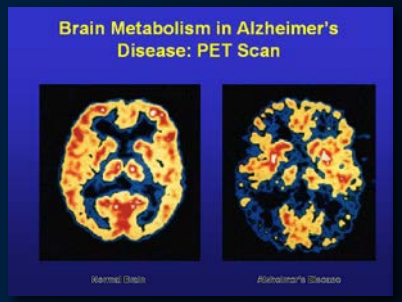
- **Develop** new technologies for accelerators and detectors

  Medicine - diagnosis and therapy

- **Train** scientists and engineers of tomorrow

- **Unite** people from different countries and cultures



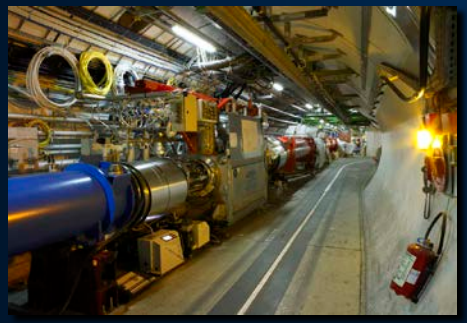


Innovation
Education
CERN
uniting people
Research

CERN: Particle Physics and Innovation

- Interfacing between fundamental science and key technological developments

- CERN Technologies and Innovation

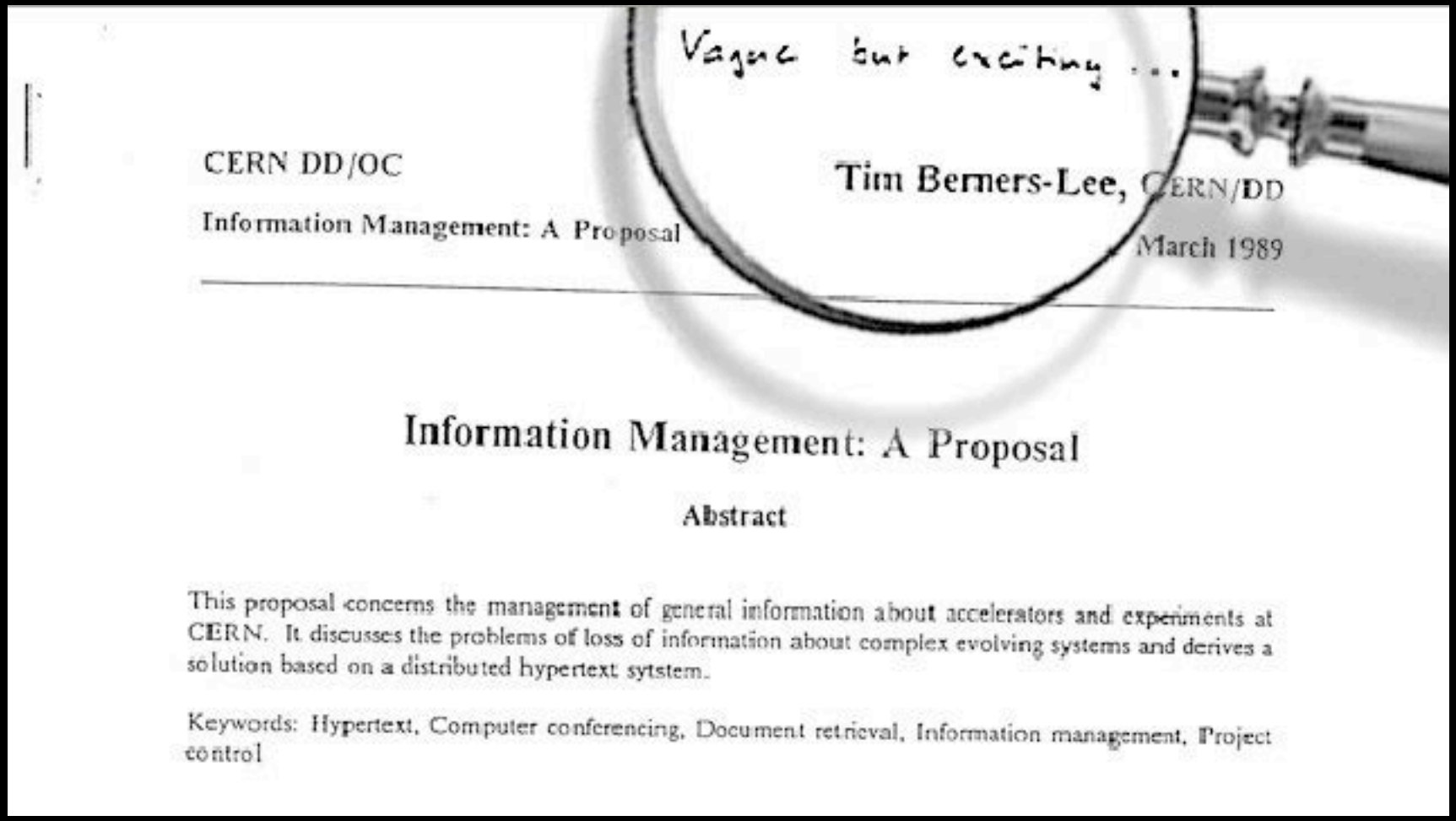


Accelerating particle beams

Detecting particles

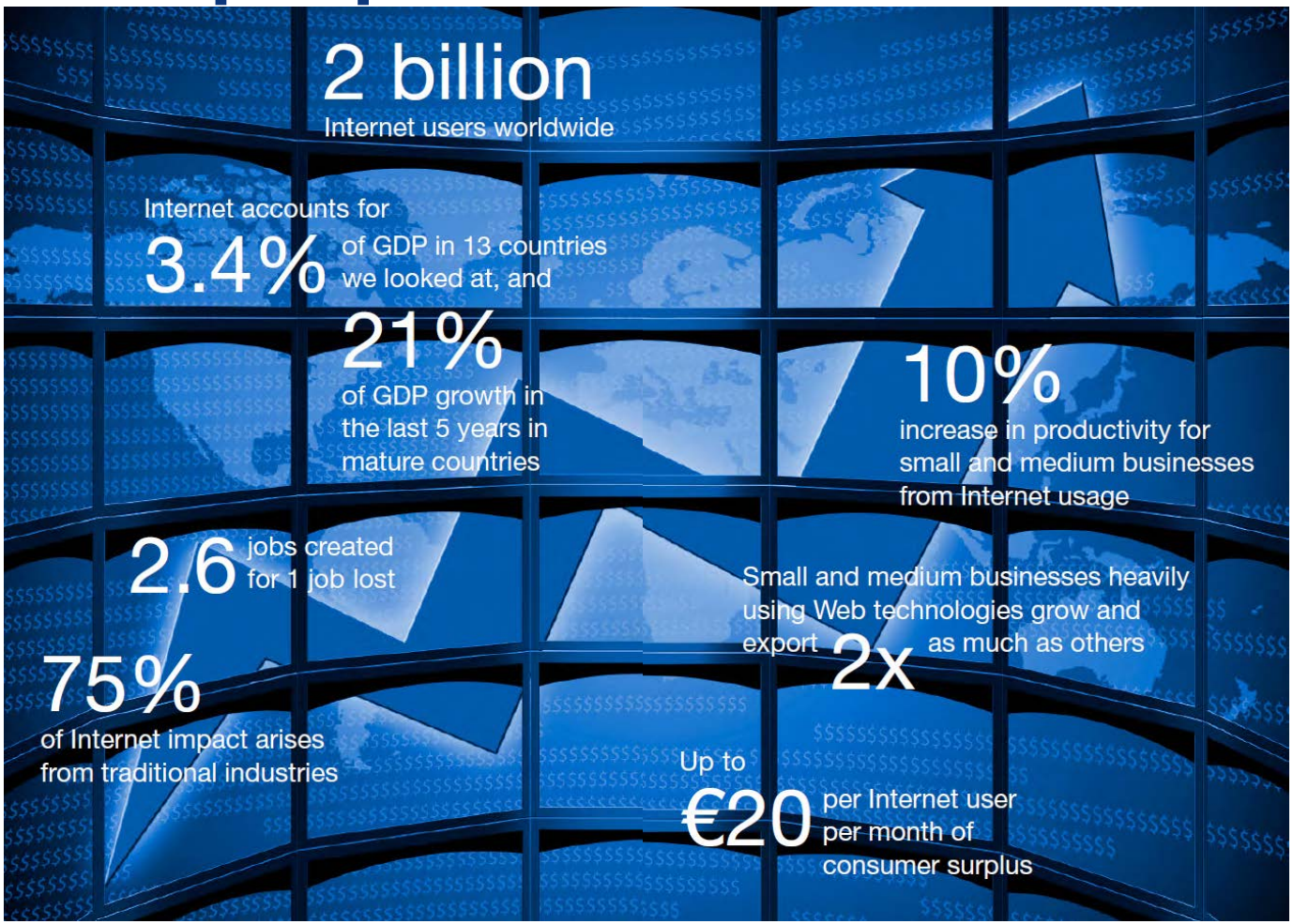
Computing

CERN


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


Web25TH
ANNIVERSARY
W3C
20

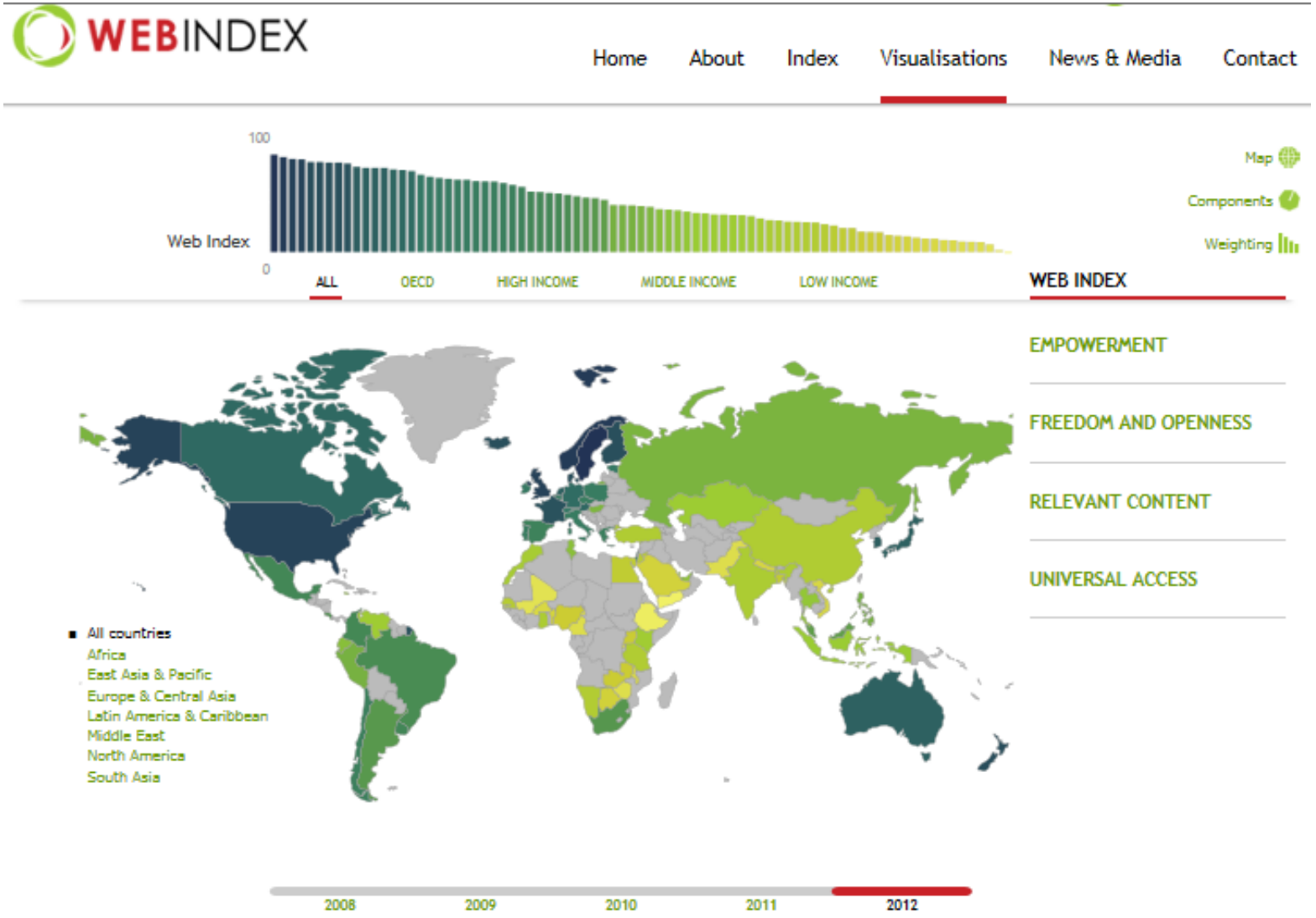
In 2 decades the web has changed from a tool for researchers to the daily reality for billions of people



Source & image: Internet matters: The Net's sweeping impact on growth, jobs and prosperity McKinsey Global Institute, 2011

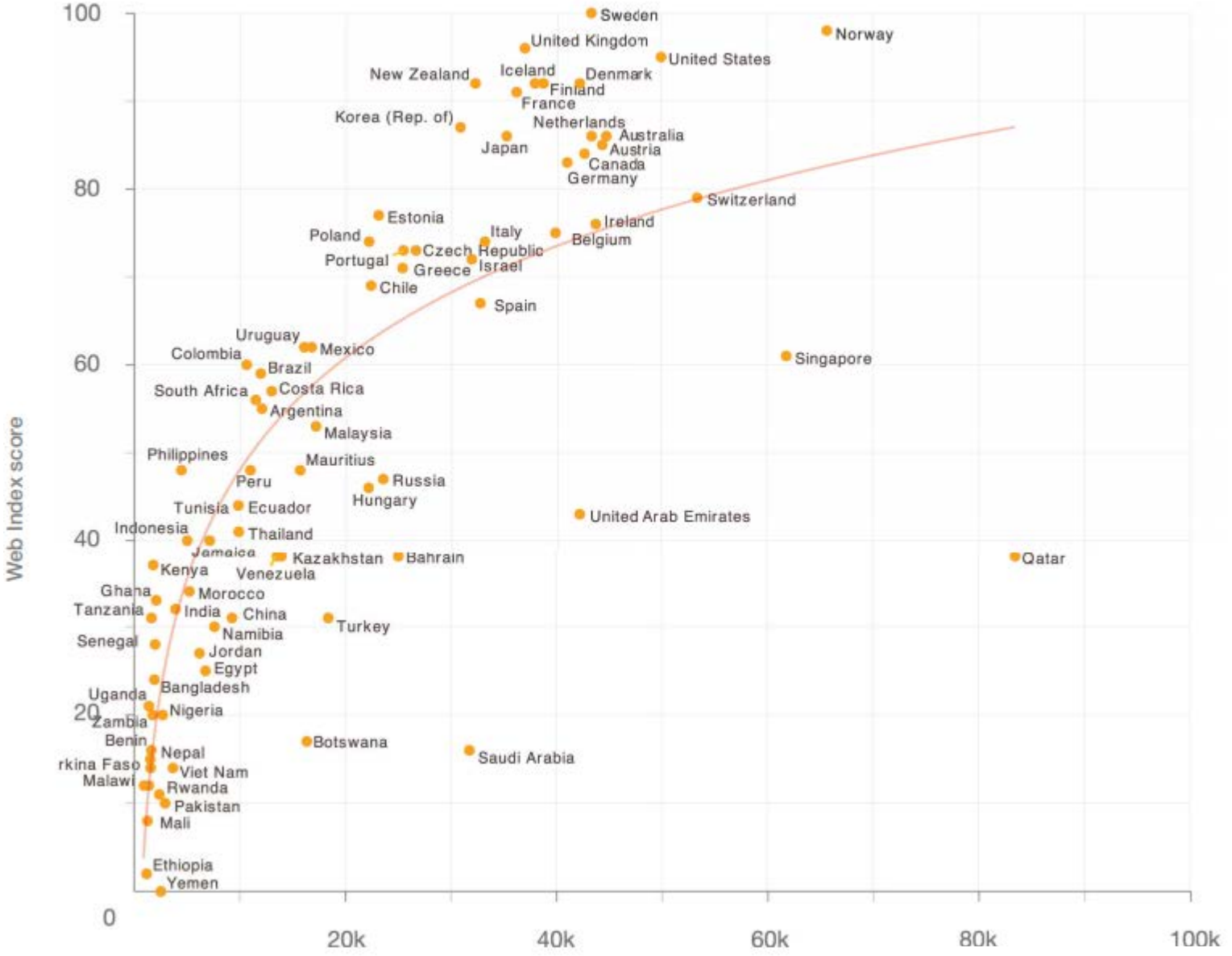
Communication is a fundamental social process, a basic human need and the foundation of all social organisation… Everyone, everywhere should have the opportunity to participate and no one should be excluded from the benefits that the Information Society offers.

World Summit on the Information Society Declaration, 2003.



https://thewebindex.org/

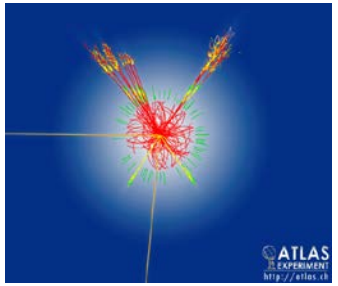
GDP PER CAPITA VS. WEB INDEX SCORE



Citizen Cyberscience Centre

Promote web-based citizen participation in science projects as an appropriate low cost technology for scientists in the developing world

- **Partners:** CERN, UN Institute for Training and Research, Univ. of Geneva
- **Sponsors:** IBM, HP Labs, Shuttleworth Foundation
- **Technology:** open src platforms for internet-based distributed collaboration
- **Projects:**
  - **Computing for Clean Water** optimizing nanotube based water filters by large scale simulation on volunteer PCs
  - **AfricaMap** volunteer thinking to generate maps of regions of Africa from satellite images, with UNOSAT
  - **LHC@home** new volunteer project for public participation in LHC collision simulations, using VM technology



Open Access

Results of publicly-funded research should be made freely available to all, not just those who can afford subscriptions to the scientific journals in which they are published

- **Access to the literature with no restriction for any reader**
- **Publishing without financial barriers for any author**
- **Maintain and stimulate a wide choice of high-quality journals**
- **An “author-friendly” copyright agreement**
- **High peer-review and editorial standards**
- **Competition among journals**
- **Get spiraling subscription costs under control**

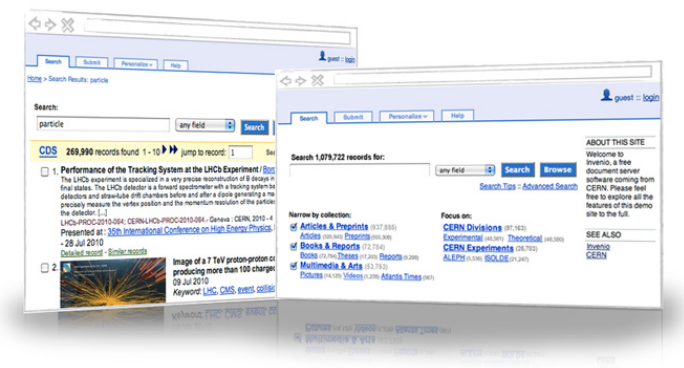
Open Access

CERN | IT Information Technology Department

http://scoap3.org/

Helping developing countries

The **Invenio** open source software suite, developed by CERN and used for LHC publications, enables you to run your own digital library on the web. The software covers all aspects of digital library management from document ingestion through classification, indexing, and curation to dissemination



UNESCO & CERN have sponsored digital library schools in Africa (Rwanda 2009, Morocco 2010, Senegal 2011)

CERN has donated computers to Africa (Morocco, Ghana, Egypt) to help capacity building.

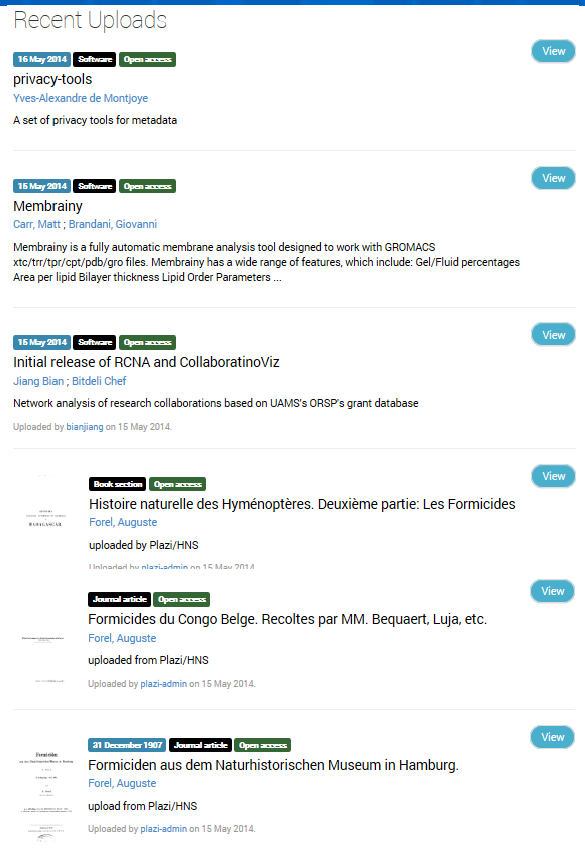
Open Repository for Research Results

zenodo

Research. Shared.

Zenodo is an open dependable home for the long-tail of science, enabling researchers to share and preserve any research outputs in any size, any format and from any science.

Store papers, data & software and make it citable



- **Research. Shared.** – all research outputs from across all fields of science are welcome!
- **Citeable. Discoverable.** – uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Community Collections** – accept or reject uploads to your own community collections (e.g workshops, EU projects or your complete own digital repository).
- **Funding** – integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- **Flexible licensing** – because not everything is under Creative Commons.
- **Safe** – your research output is stored safely for the future in same cloud infrastructure as research data from CERN's Large Hadron Collider.
- **DropBox integration** – upload files straight from your DropBox.

Powered by: INVENIO

http://zenodo.org/

OpenAIRE CERN SEVENTH FRAMEWORK PROGRAMME e-infrastructure

And what about the LHC?

Nobel Prize in Physics 2013

The Nobel Prize in Physics 2013 was awarded jointly to François Englert and Peter W. Higgs *"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider".*

To find the Higgs you need 3 things

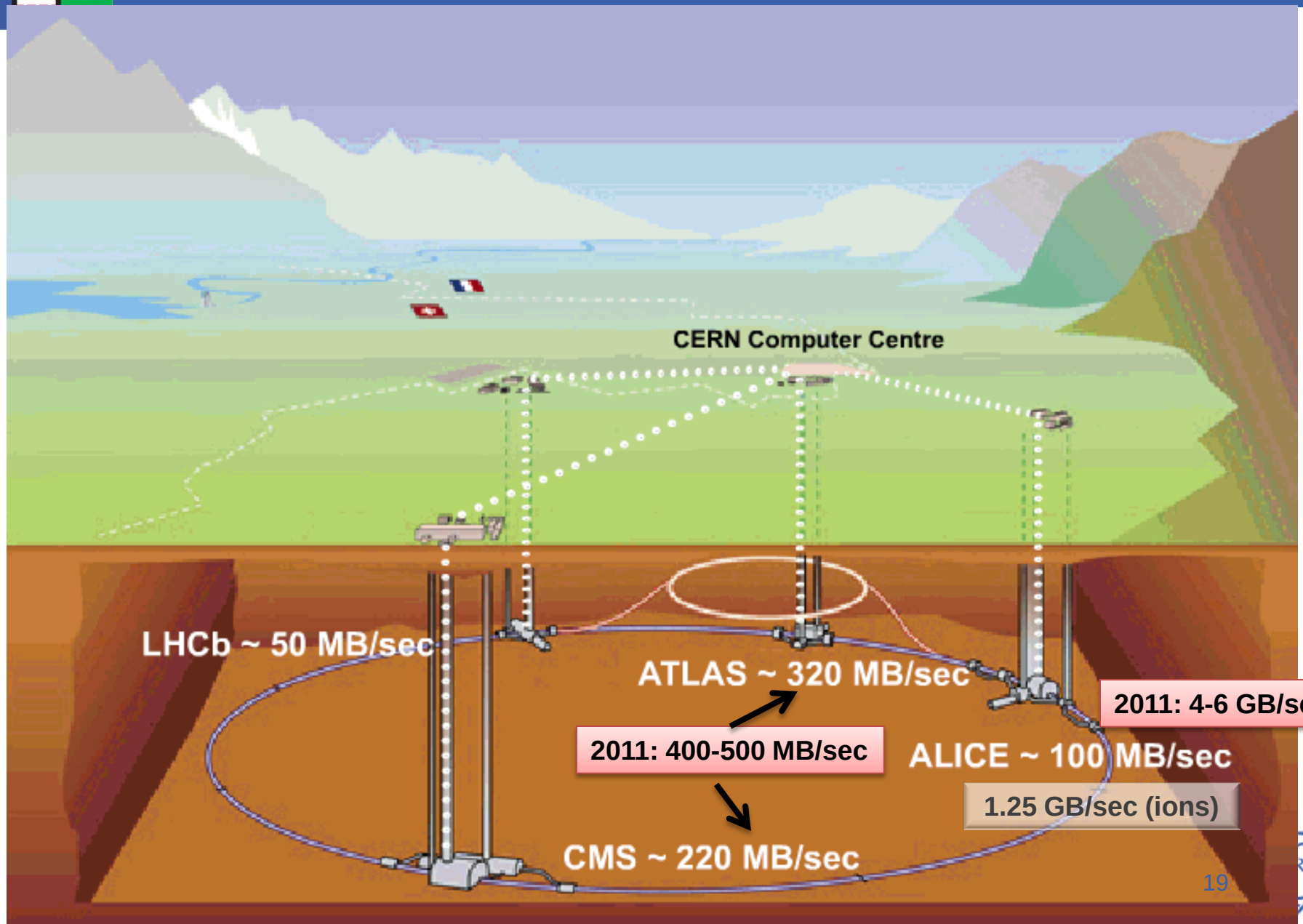
The Accelerator

The Experiments

The GRID

Rolf-Dieter Heuer, DG, CERN, July 4 2012

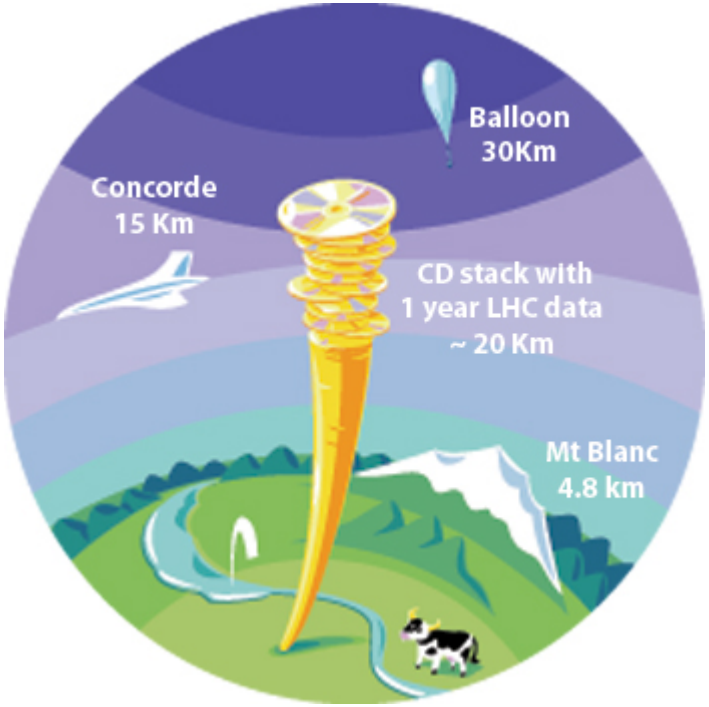
Data Acquisition, First pass reconstruction, Storage & Distribution



The LHC Data Challenge

CERN IT Department

- The accelerator will run for more than 20 years
- Experiments are producing about **25 Million Gigabytes** of data each year (about 1000 years of DVD movies!)
- LHC data analysis requires a computing power equivalent to **~100,000 of today's fastest PC processors**
- Requires many cooperating computer centres, as CERN can only provide ~15% of the capacity



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Solution: the Grid

CERN IT Department

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

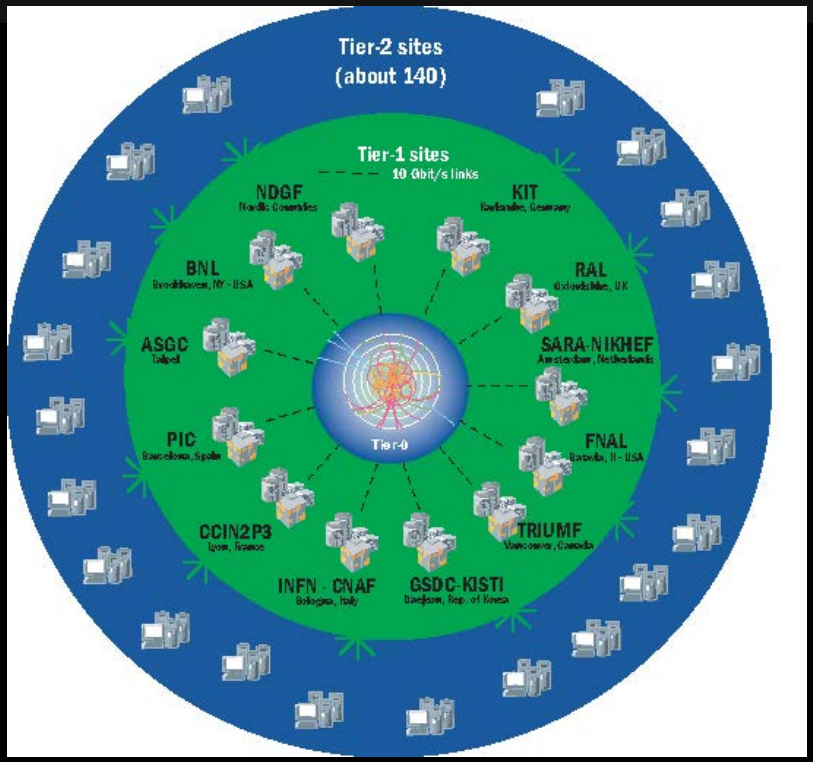
The **World Wide Web** provides seamless access to information that is stored in many millions of different geographical locations

The **Grid** is an infrastructure that provides seamless access to computing power and data storage capacity distributed over the globe

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

The Worldwide LHC Computing Grid

- A distributed computing infrastructure to provide the production and analysis environments for the LHC experiments
- Managed and operated by a worldwide collaboration between the experiments and the participating computer centres
- The resources are distributed – for funding and sociological reasons
- Our task was to make use of the resources available to us – no matter where they are located



Tier-0 (CERN):
- Data recording
- Initial data reconstruction
- Data distribution

Tier-1 (12 centres + Russia):
- Permanent storage
- Re-processing
- Analysis

Tier-2 (~140 centres):
- Simulation
- End-user analysis

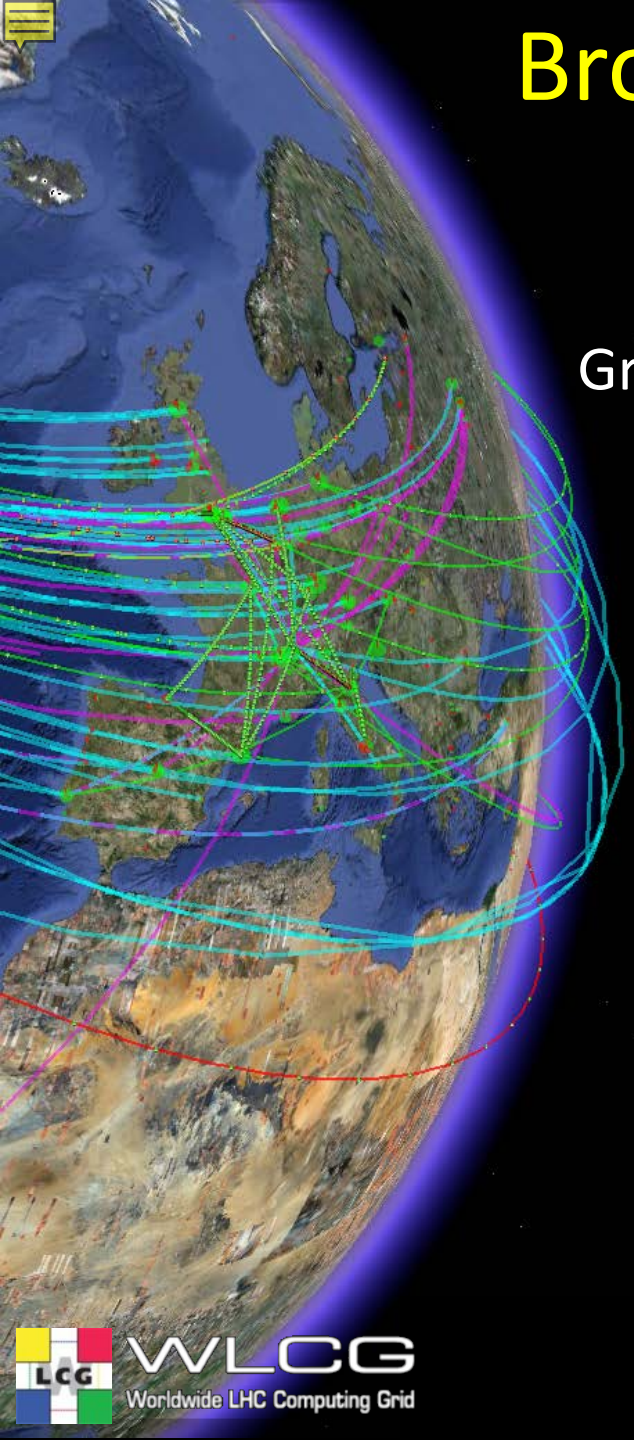
- **~ 160 sites, 35 countries**
- **300000 cores**
- **200 PB of storage**
- **2 Million jobs/day**
- **10+ Gbps links**

WLCG Worldwide LHC Computing Grid

WLCG video

http://ml-server01.cern.ch/files/DataDeluge/06%20Data%20Deluge%20Tier%20Map%2001_20140210163408609 3.mp4

Broader Impact of the LHC Computing Grid



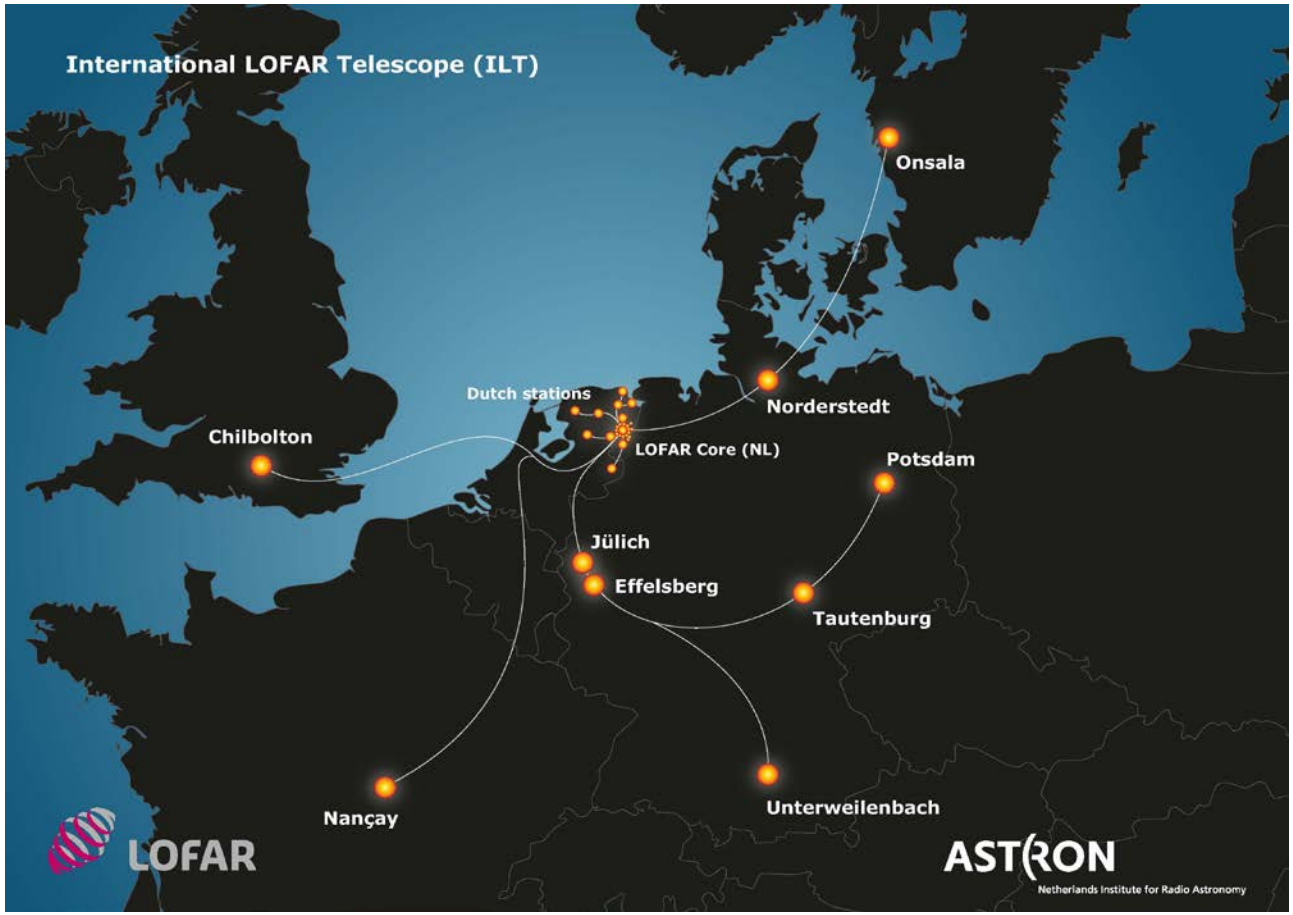
Grid has benefited the wider scientific community

- Europe (EC FP7):
  - Enabling Grids for E-sciencE (EGEE) 2004-2010
  - European Grid Infrastructure (EGI) 2010--
- USA (NSF):
  - Open Science Grid (OSG)
- Asia:
  - Japan, Korea, Taiwan etc.

Archeology
Astronomy
Astrophysics
Civil Protection
Comp. Chemistry
Earth Sciences
Finance
Fusion
Geophysics
High Energy Physics
Life Sciences
Multimedia
Material Sciences
… Zoology

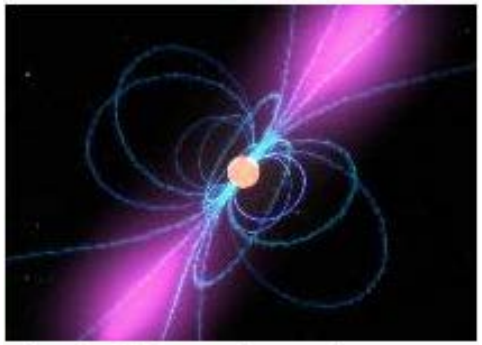
WLCG
Worldwide LHC Computing Grid

Physical Sciences



Low-Frequency Array for radio astronomy

Helping astronomers find pulsars



Clouds of charged particles move along the pulsar's magnetic field lines (blue) and create a lighthouse-like beam of gamma rays (purple).

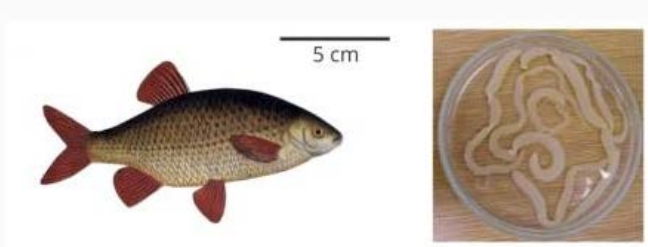
(Image: NASA via wikicommons)

Natural Sciences

Correlate data from millions of calculations to unveil the rock structure of an oil field under the North Sea

Image: wikicommons

Trace the tapeworms infecting Northern African fish back to Europe

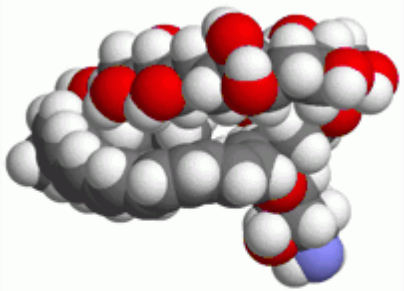


The rudd (left) introduced in North African lakes is often infected with several *Ligula* tapeworms. Image to scale.

W. Bouzid, J. Štefka et al. (2013)
Pathways of cryptic invasion in a fish parasite traced using coalescent analysis and epidemiological survey.
Biological Invasions

Life Sciences

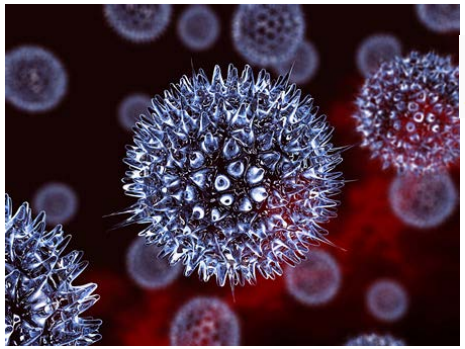
Designing better antibiotics



Three-dimensional model of the Amphotericin B molecule. (Source: wikicommons)

A. Neumann, M. Baginski and J. Czub. 2010. How Do Sterols Determine the Antifungal Activity of Amphotericin B? Free Energy of Binding between the Drug and Its Membrane Targets. *Journal of the American Chemistry Society*, 132: 18266–18272. doi:10.1021/ja1074344 (abstract)

Hunting for viruses



L van der Hoek et al. (2004) Identification of a new human coronavirus. *Nat Med 10*: 368–373.

Tracking a biomarker for Alzheimer's disease

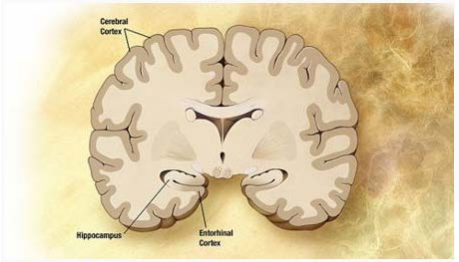


Illustration showing a brain at the preclinical stage of Alzheimer's disease, highlighting the location of the hippocampus.

Cover, K.S., et al. (2013) A standard benchmark for assessing the reproducibility of brain atrophy measures in Alzheimer's using the ADNI1 data set. Poster presented at the AAIC 2013 in Boston, MA.

Information Society Technologies

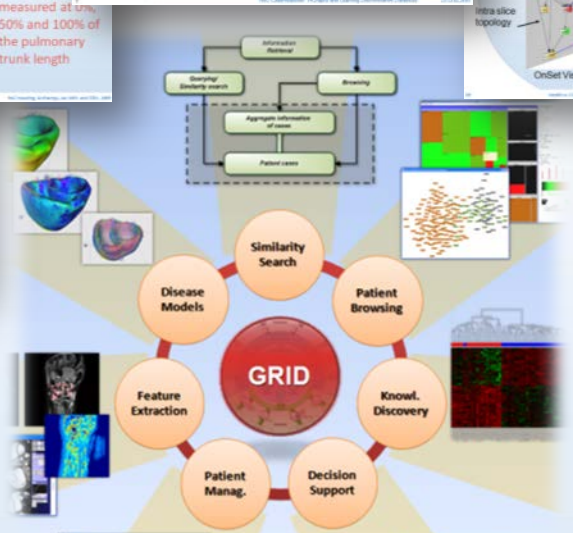
Similarity Search

Measurement of Pulmonary Trunk

Temporal Modelling

RV and LV Automatic Modelling

Visual Data Mining



Surgery Planning

Genetics Profiling

Personalised Simulation

Treatment Response

Semantic Browsing

Inferring Outcome

Biomechanical Models
Tumor Growth Modelling


Department
for Business
Innovation & Skills
Science & Technology Facilities Council
www.stfc.ac.uk
SCIENCE & ENGINEERING WEEK
is playing a leading role in the world's biggest scientific experiment
the Large Hadron Coll
CERN in Geneva - recre
the conditions tha
existed a trillionth of a second after the beginning of the Universe.

Sample of Business Applications

- **SMEs**
  - NICE (Italy) & GridWisetech (Poland): develop services on open source middleware for deployment on customer in-house IT infrastructure
  - OpenPlast project – (France) Develop and deploy Grid platform for plastics industry
  - Imense Ltd (UK) - Ported gLite application and GridPP sites
- **Energy**
  - TOTAL, UK - Ported application using GILDA testbed
  - CGGVeritas (France) – manages in-house IT infrastructures and sells services to petrochemical industry
- **Automotive**
  - DataMat (Italy) – Provides grid services to automotive industry

LATEST NEWS

CellProfiler on the Cloud
Secure Gene Sequencing Portal
Constellation Technologies
Model-based Analysis of Tiling-arrays for ChIP-Chip

POPULAR

About Constellation Technologies
Constellation Technologies
Model-based Analysis of Tiling-arrays for ChIP-Chip
Constellation Technologies and Cambridge University Explore the Cloud Together
Middleware Technology

Technology

Middleware Technology

Middleware is a "glue" software that pools together various computing resources to create the Grid. It allows users to securely access the integrated computing and storage resources of in a way similar to accessing an enourmous virtual computer. EGEE develops and deploys a middleware distribution called gLite (pronounced "gee-lite"). gLite is a result of collaborative efforts of more than 80 people in 12 different academic and industrial research centers as part of the EGEE Project. gLite provides a framework for building grid applications tapping into the power of distributed computing and storage resources across the Internet.

The gLite middleware stack provides the user both with foundation level and higher level services. Foundation level services ensure security, resource access and systems to monitor grid activity. These provide the basis for a consistent and dependable production infrastructure. Higher level services provided by gLite include job management, data catalogues and data replication, providing applications with the tools to build end-to-end solutions. Other third party projects complete a rich ecosystem built on the gLite foundation services.

Constellation Technologies will be developing cloud computing solutions for the next generation of Internet based on gLite.

Qarnot Computing video:

Trim: 00:10 - 02:34
http://vimeo.com/38095665

CERN openlab in a nutshell

- A science – industry partnership to drive R&D and innovation with over a decade of success
- Evaluate state-of-the-art technologies in a challenging environment and improve them
- Test in a research environment today what will be used in many business sectors tomorrow
- Train next generation of engineers/employees
- Disseminate results and outreach to new audiences

PARTNERS

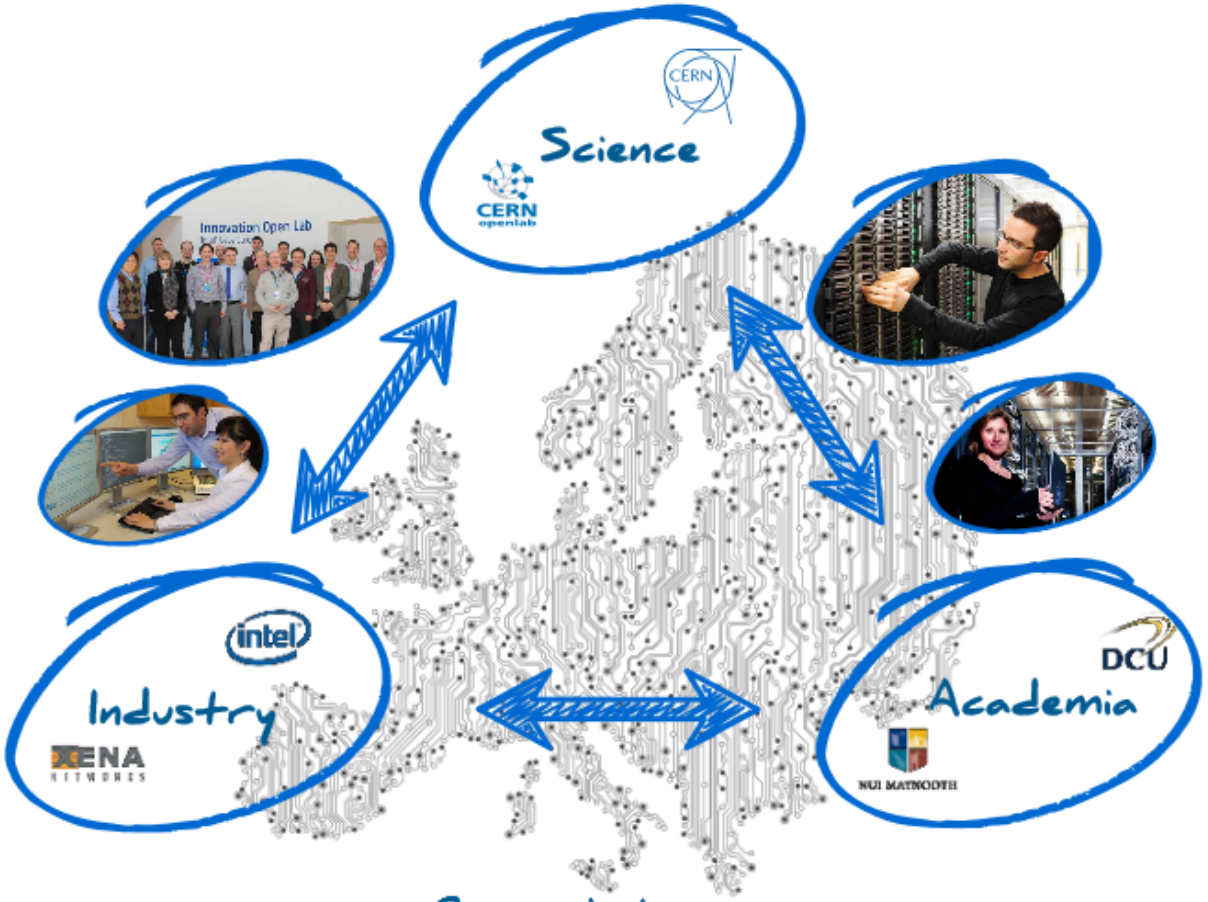
CONTRIBUTOR

ASSOCIATE

CERN openlab

ICE-DIP 2013-2017:
The Intel-CERN European Doctorate Industrial Program

A public-private partnership to research solutions for next generation data acquisition networks, offering research training to five Early Stage Researchers in ICT



Research topics:

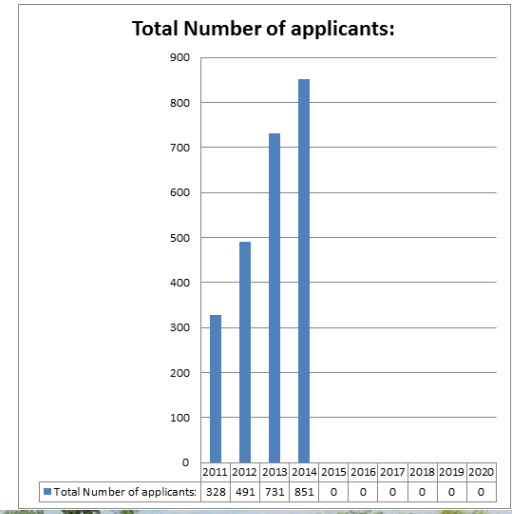
- Silicon photonics systems
- Next generation data acquisition networks
- High speed configurable logic
- Computing solutions for high performance data filtering

SEVENTH FRAMEWORK PROGRAMME
CERN
MARIE CURIE ACTIONS

Summer Student Program

9 week residential work/study programme hosted at CERN

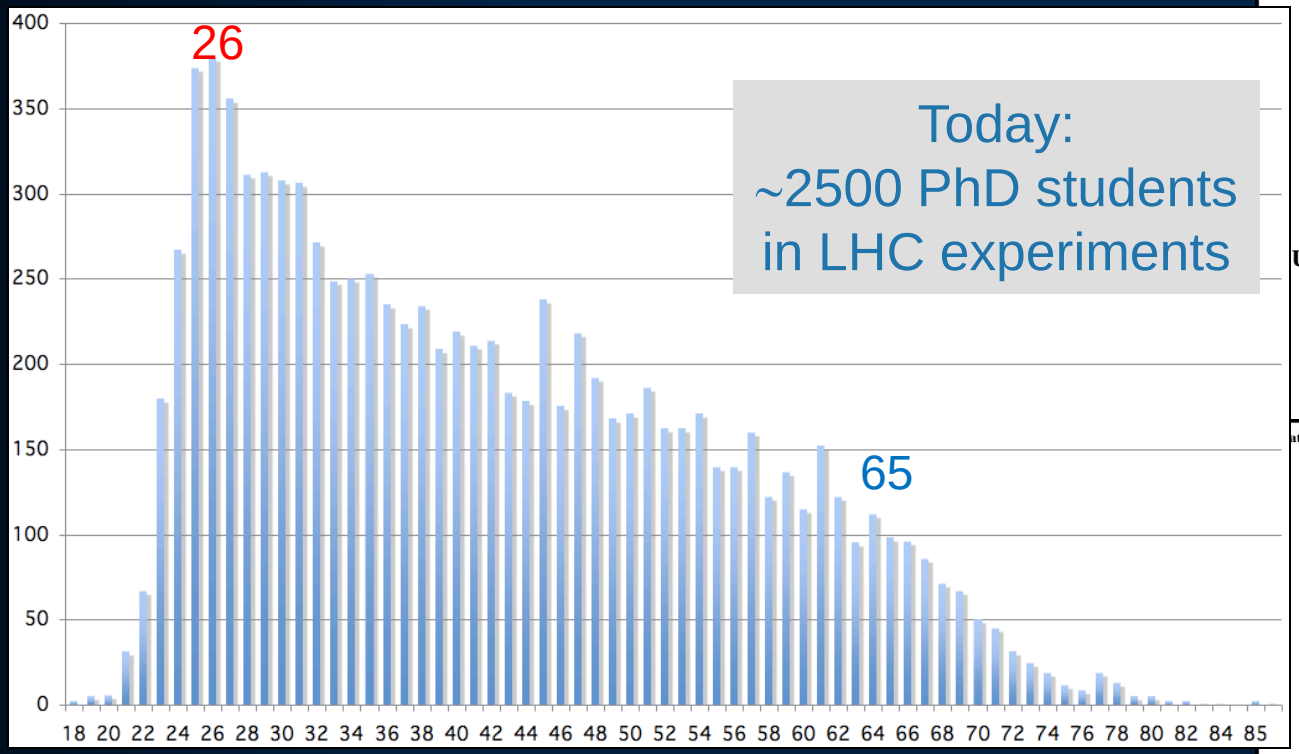
- 720+ applicants (2013)
- 22 selected candidates
- 13 lectures (including external labs)
- Student lightning talks session
- 22 technical reports



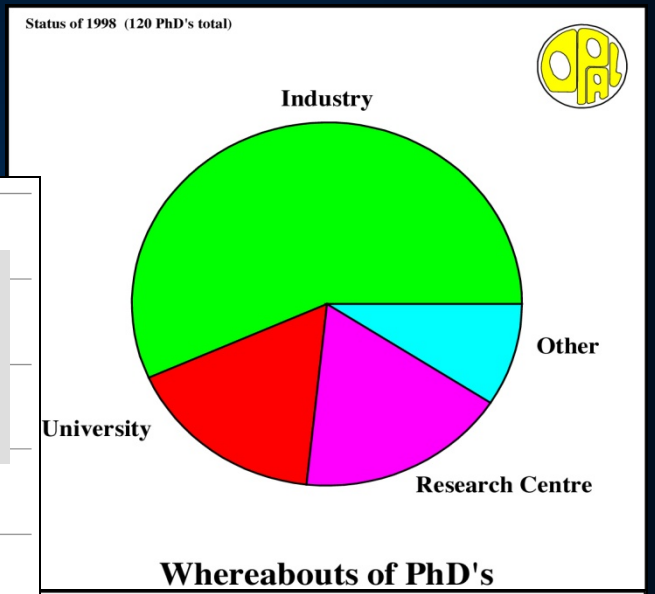
Age Distribution of Scientists

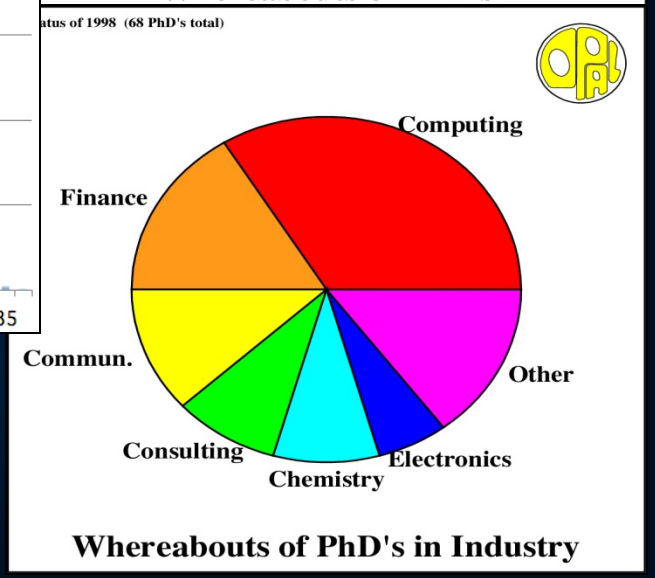
- and where they go afterwards

Survey in March 2009



Today:
~2500 PhD students
in LHC experiments

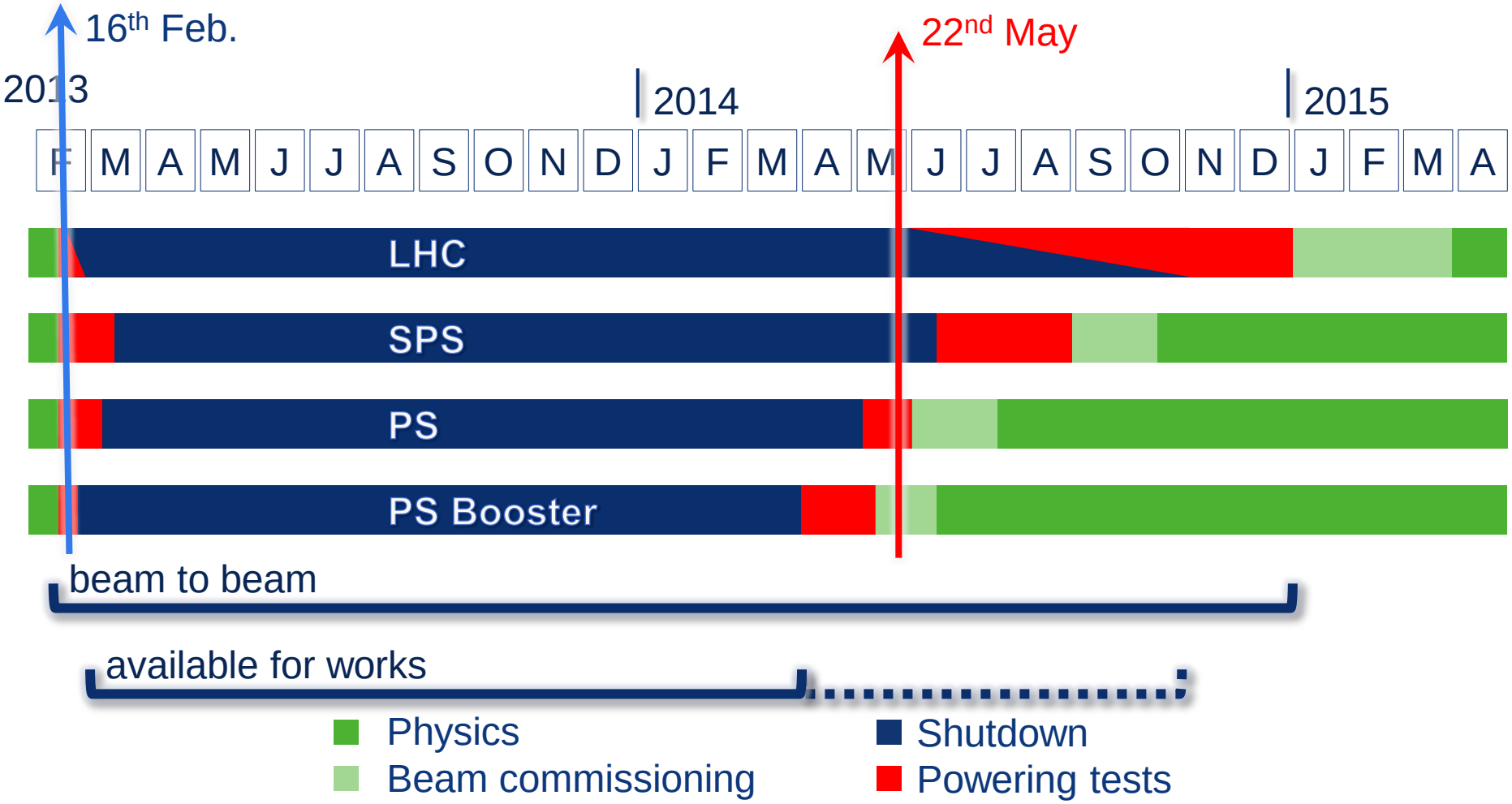




They do not all stay: where do they go?

Where do we go next?

LHC Long Shutdown 1 (Feb'13 to Dec'14)



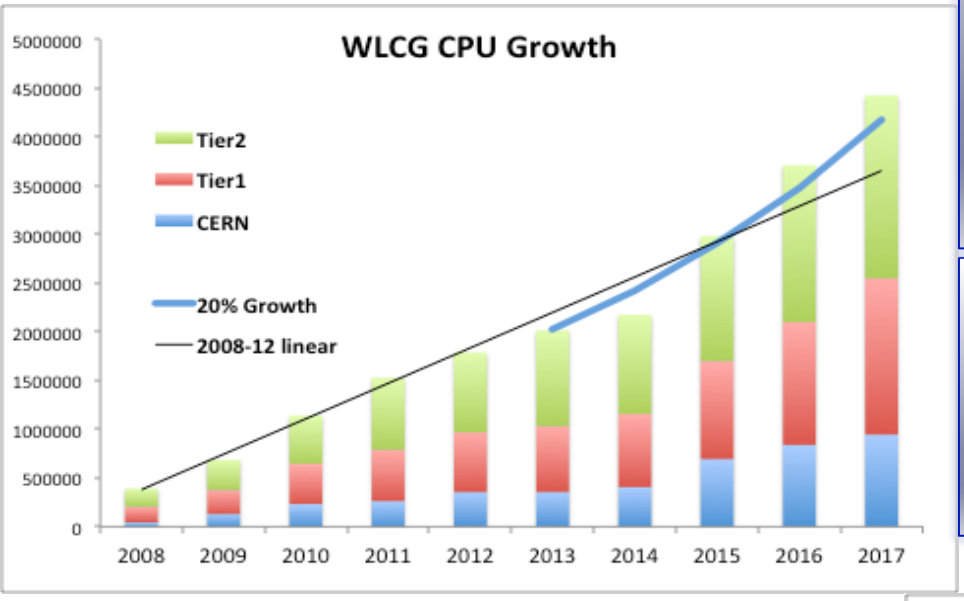
Evolution of Computing requirements

Higher trigger (data) rates driven by physics needs

Based on understanding of likely LHC parameters;

Foreseen technology evolution (CPU, disk, tape)

Experiments work hard to fit within constant budget scenario



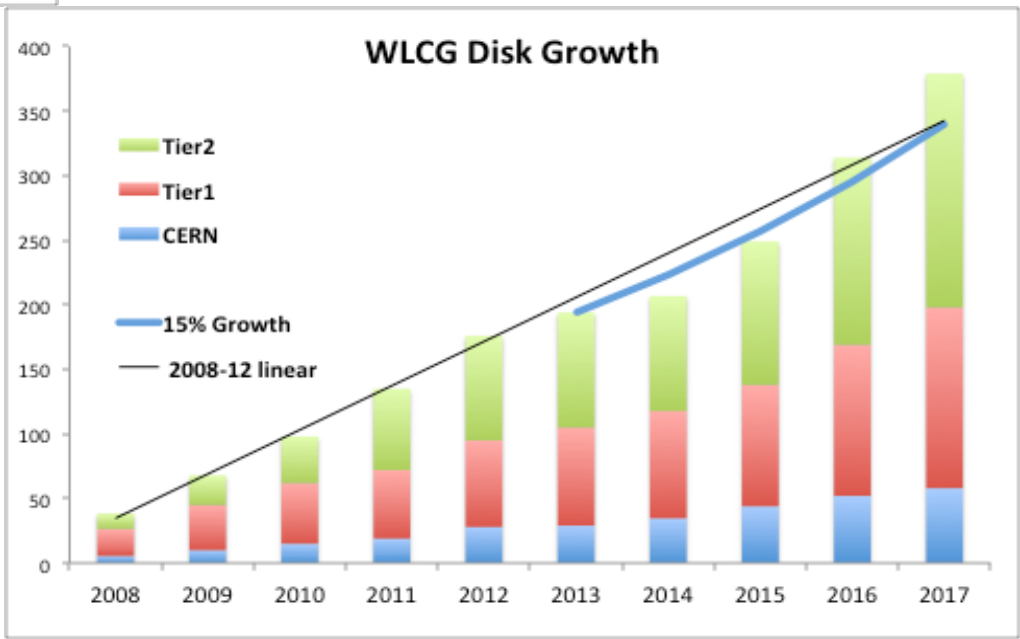
Estimated evolution of requirements 2015-2017

2008-2013: Actual deployed capacity

Line: extrapolation of 2008-2012 actual resources

Curves: expected potential growth of technology with a constant budget

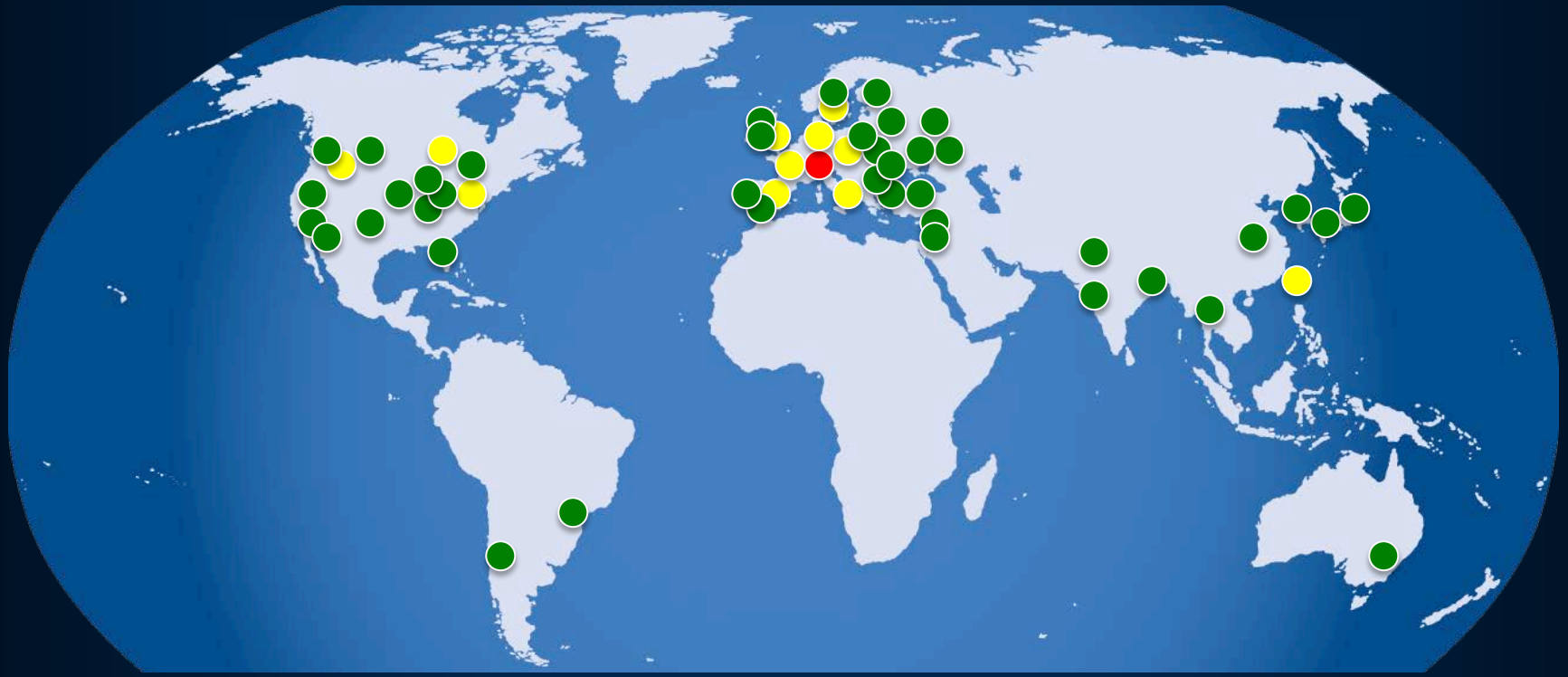
- CPU: 20% yearly growth
- Disk: 15% yearly growth



CERN

WLCG

From grids to clouds



Expand the grid so it is:

- More efficient & powerful
- More open - engage public & commercial service providers

May 2014: 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 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

PIC port d'informació científica

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

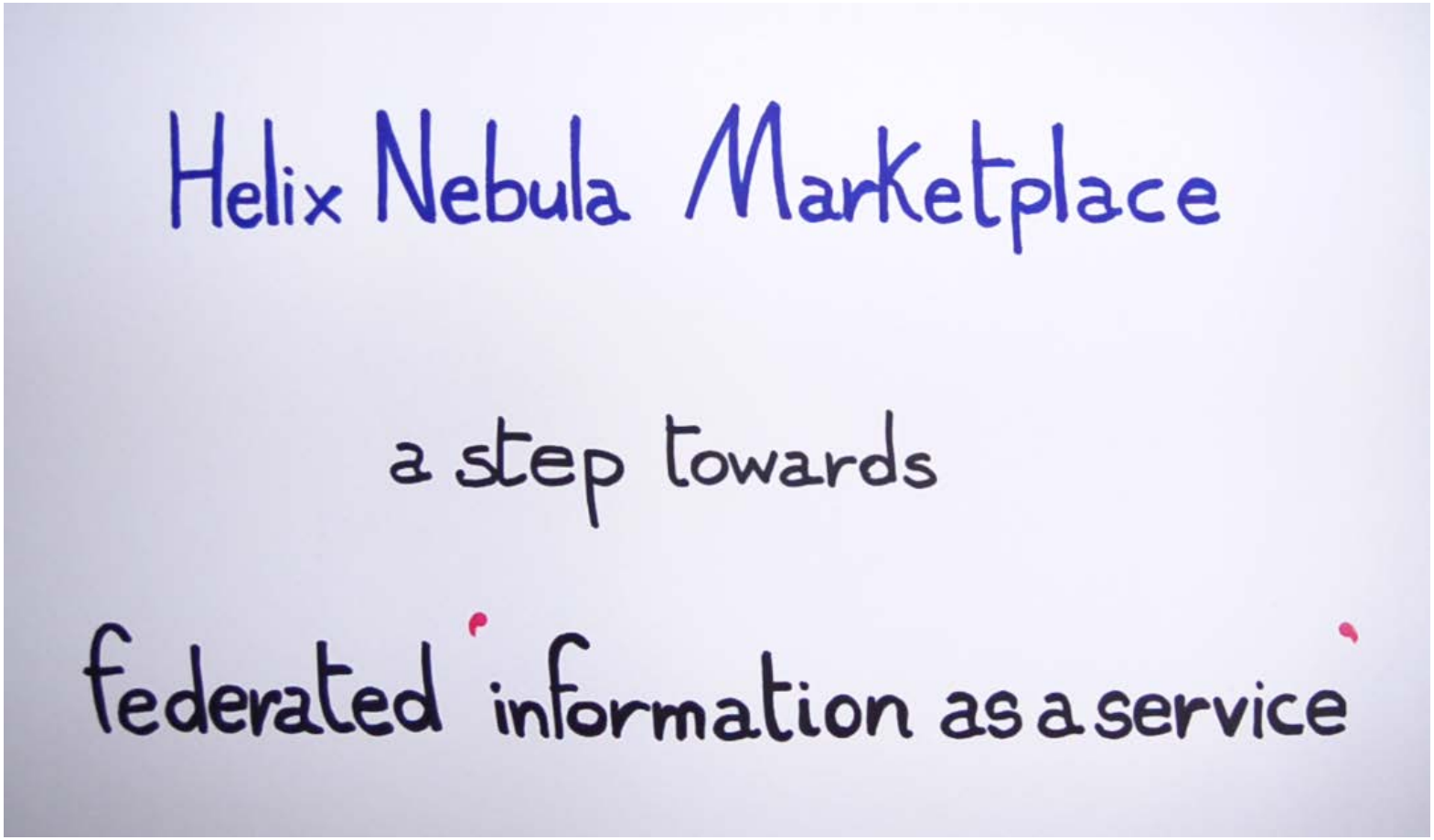
Suppliers

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

Adopters

AWST Capgemini CLOUDEO cnrs ESO Ifremer NEPHOS TECHNOLOGIES orange Business Services Telefónica terradue 20 THALES Yandex

HNX video



https://www.youtube.com/watch?v=kn6gCpgxifg
http://www.helix-nebula.eu/media/videos

Future IT Challenges

CERN openlab publishes a whitepaper on future IT challenges in scientific research

22 May 2014

Geneva, 22 May 2014. CERN openlab[1], the public-private partnership between CERN[2], leading IT companies and research institutes, released today a whitepaper on future IT challenges in scientific research to shape its upcoming three-year phase starting in 2015.

96% of our universe is still unknown and the challenges ahead for the scientific community are striking. More than ever, computing plays a critical role in helping uncover our universe's mysteries. Scientific research has seen a dramatic rise in the amount and rate of production of data collected by instruments, detectors and sensors in the recent years. The LHC detectors at CERN produce a staggering one petabyte of data per second, a figure that will increase during the next LHC run starting in 2015. New international research infrastructures are being deployed and are expected to produce comparable—or even greater—amounts of data in various scientific domains, such as neurology, radio astronomy or genetics, and with instruments as diverse as Earth observation satellites, high-performance genomic sequencers, neutron diffractometers or X-ray antennas. More than ever, collaboration will play a vital role in enabling discoveries.

In this context, CERN openlab together with a number of European laboratories, such as EMBL-EBI, ESA, ESRF, ILL, and researchers from the Human Brain Project, as well as input from leading IT companies, have published a whitepaper defining the ambitious challenges covering the most crucial needs of IT infrastructures in domains such as data acquisition, computing platforms, data storage architectures, compute provisioning and management, networks and communication, and data analytics. A number of use cases in different scientific and technological fields are described for each of the six major areas of investigation.

Continuous collaboration between the research infrastructures and IT companies is more critical than ever to make sure scientific objectives and technological roadmaps are aligned. In the current CERN openlab phase, Huawei, Intel, Oracle, Siemens are openlab partners, while Rackspace is a contributor and Yandex an associate. This whitepaper, which results from six months of reflection among IT experts and scientists, represents an exciting context for the CERN openlab public-private partnership in the years to come. It sets the goals, the technical expertise and identifies educational programs required, providing opportunities for future collaboration among CERN, other European laboratories, international scientific projects and leading IT companies to push the limits even further in support of many more years of outstanding scientific discoveries.

- Data analysis facility
  - Preserve applications
  - Secure remote researcher access
- Secure data federation
  - Federated identity
  - Role based data access
- Remote management of analysis facility
  - Secure remote access for administration
  - Isolation of roles
- Research clouds at scale
  - Elastic access to large compute resources
  - Project based authentication, provisioning and resources

CERN's ambitious research programme will keep it at the forefront of science and innovation.

The bright and enthusiastic young people that CERN attracts will continue to develop IT solutions to meet these challenges and serve society as a whole.

The material and work presented is the result of many people, organisations and projects such as:

CERN and the openlab project
EGEE & EGI
Helix Nebula initiative
W3C & The Web Foundation
WLCG collaboration

Thank you for your attention

CERN

Accelerating Science and Innovation