Grid History and Standards

(slides contributed by Dr. Kilian Schwarz, GSI)

Dr. Rüdiger Berlich and Christopher Jung,
Forschungszentrum Karlsruhe

DESY, Hamburg / Germany; 3.2.05

The ARPANET

- 1969 – ARPANET (Advanced Research Projects Agency Network – US ministry of defense) built with 50 kb/s
- “Resource Sharing Computer Network” using telephone cables and NCP (Network Control Program – already including telnet and FTP)
- 1973 – first international dial in into ARPANET

Dawning of the internet age

- 1984 – 1000 computers in the internet
- 1986 – IETF (Internet Engineering Task Force) founded. Open international organisation, takes care of the development of the internet architecture as well as it's functionality.
- 1986 - NSFnet (National Science Foundation) substitutes ARPANET as government network between universities and research centres.
- 1987 – 10000 computers in the internet

Supercomputer and the Internet

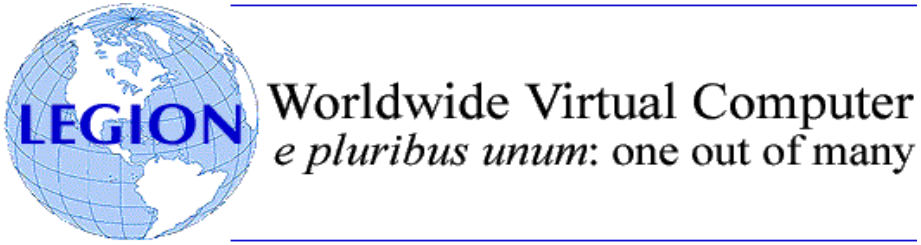
- 1988 – 60000 computers in the internet
- 1988 – NFSnet backbone 1,5 Mbs
- 1988 – IRC (Internet Relay Chat)
- 1988 - Cray Y-MP (2 GB RAM, 2,67 Gflops)
- 1988 – start of the Condor project

Condor

- A specialised workload management system. The jobs are put into a queue, condor choses where and when the jobs should run, monitors the running jobs and informs the user.
- Condor Class Ads: computers advertise their abilities via “Classified Adverts”. Condor compares them with the user requirements. Also preferences can be given.
- batch submission, local clusters
- CPU-cycle harvesting)

Legion

- 1993 – start of the projekts “Legion”
- Object orientiented approach to Grid Computing (everything is an object: files, computing resources, mass storage ...)
- But: difficult to implement
- Many potential applications have not yet been object oriented..
- Some successful applications, though



1994 - 1997

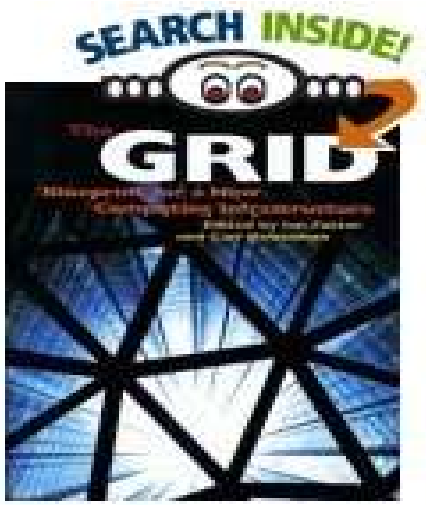
- 1995 -- Netscape IPO 3.highest in NASDAQ history
- 1995 – Java started by SUN
- 1995 – I-WAY (Information Wide Area Year) predecessor project of Globus already with I. Foster (NZ) and C. Kesselman
- 1996 -- SETI@home
- 1997 – Globus in development stage (sucessor of I-WAY)
- 1997 – start of the UNICORE project (BMBF)

Unicore

- UNICORE = Uniform Interface for Computing Resources
- A project of the German Ministry for Research and Education (BMBF)
- aim: creating a network of super computing centres and providing a uniform interface while using existing technologies.
- Written in Java (portability)
- Provides middleware functionality and a portal (GUI)
- Offers job preparation, monitoring and control, complex workflows, file management, support of Grid certificates, resource broker … soon

1997 - 1998

- 1997 -- Condor used at NCSA.
- 1997 -- “Building a computational Grid” Workshop in Argonne National Lab (ANL)
- 1997 – Storage Resource Broker (SRB) (San Diego Supercomputer Center SDSC)
- 1998 – Foster/Kesselman: “The Grid Book”
- 1998 – XML v1.0



SDSC SRB

- interface to connect various data resources via network
- Complete solution for distributed datasets
- Includes LFN/PFN schema, interfaces to various mass storage systems, high performance data movement, meta data catalogue, GUIs, user Tools for Windows, Unix and Web, as well as APIs for C, C++, Java and Python

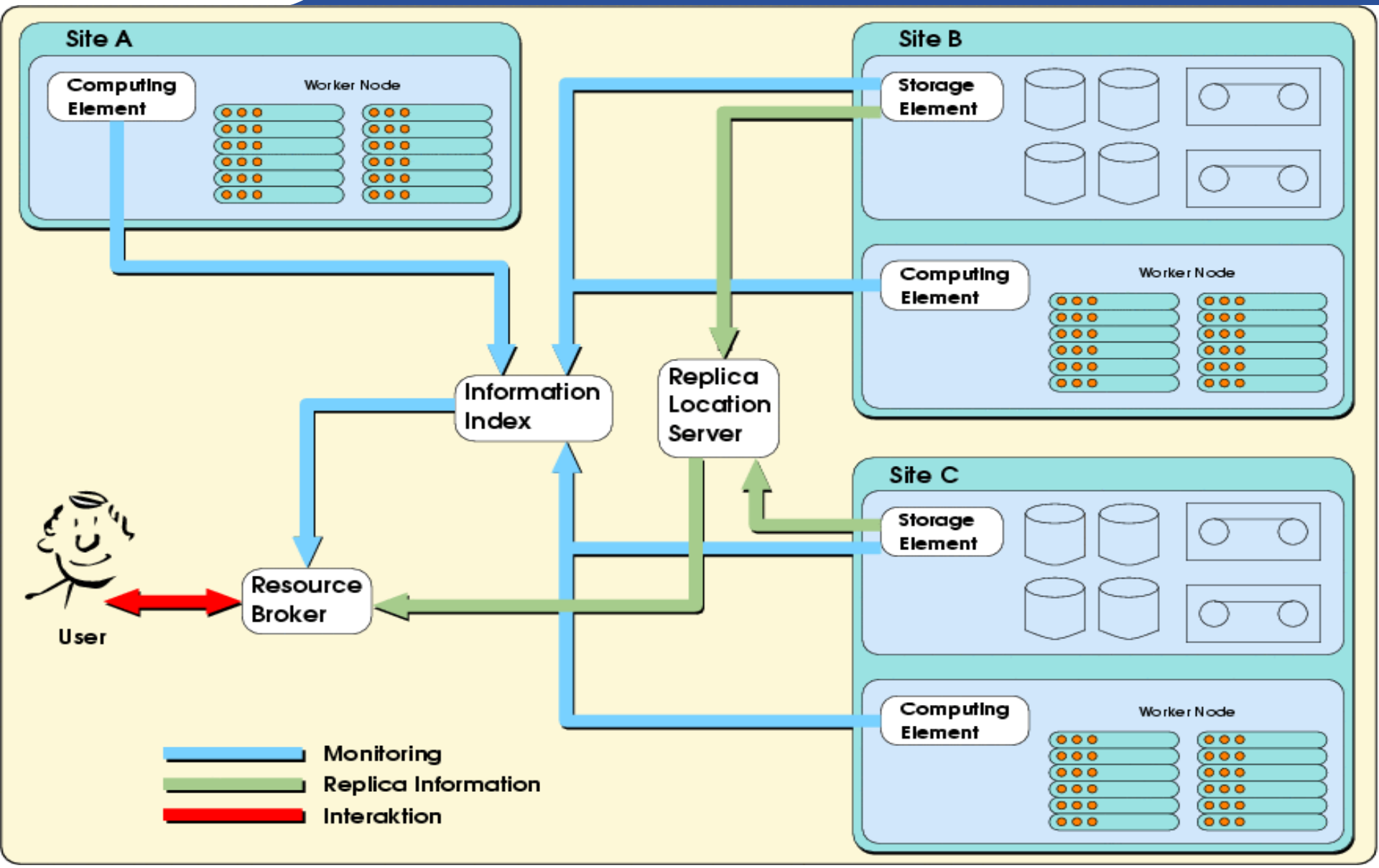
Globus 1.0

- October '98: Release of Globus 1.0
- www.globus.org
- No complete Grid, but the leading Grid tool box (most Grid projects are based on Globus)
- ANL, University of Chicago, University of Edinburgh, ... , IBM, Microsoft, ...
- Defines and implements standards for
  - Grid security (GSI)
  - Data access and transfer (GASS and GridFTP)
  - Resource management and usage (GIS and GRAM)
  - Job execution on remote clusters (globusrun)

Grid Security Infrastructure

- GSI becomes quasi standard
- Based on the free SSLeay package and uses X.509 certificates (Public Key Infrastructure) => see talk "Layout of LCG-2, VOs & CAs"
- Enables “Single sign-on” on the Grid, user identity is guaranteed with a single certificate (no frequent logins into various resources with different passwords)

Globus: Simplicity by design



1999 – 2001 big projects start up

- 1999 Grid Forum 1
- 2000 Eurogrid starts
- 2000 SUN Grid Engine (“Sun powers the Grid”)
- 2000 NASA IPG (Information Power Grid), the high performance computing and data grid of the NASA
- 2001 start of the EDG project
- 2001 start of AliEn
- 2001 Global Grid Forum 1 (American Grid Forum + Asia Pacific + European Grid Form eGRID)

Global Grid Forum

- The GGF is a forum consisting of thousands of individuals coming from industry and research. Their aim is the global standardisation of Grid computing
- Ideas: development and usage of Grid technologies
- method: user experience, guide lines, development and documentation of technical specifications
- Strong commercial participation

GGF10, Berlin, March 2004
GGF == Standardisation Gremium, similar to IETF

The Greenback: “Sun powers the Grid”

Ian Foster at
Sun booth during
SC2001, Denver

Irving Wladawsky-Berger
at GGF4 in Toronto -
$ 4 billion IBM investment

European DataGrid / LCG

- Joint three-year project of European Union
- Built on the Globus-2 Middleware
- Goal: Development of methods for the transparent distribution of data and programs
- Needed in particle physics, biology (genome project), earth observation ...
- 21 members, 15 compute centers (2-32 CPUs, up to 1 Terabyte of mass storage)
- LHC Computing Grid Middleware (LCG 2.3) partially based on EDG-2 – accesses resources with thousands of CPUs, e.g. GridKa at Forschungszentrum Karlsruhe
- Major new component: resource broker
- Project was finished in March 2004. Successor is EGEE ("Enabling Grids for eSciencE" - see below)

DataGrid in Numbers

People

>350 registered users

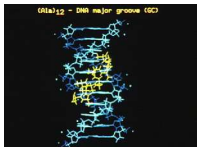
12 Virtual Organisations

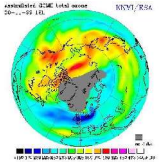
16 Certificate Authorities

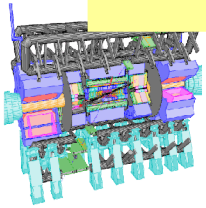
>200 people trained

278 man-years of effort

100 years funded







Testbeds

>15 regular sites

>40 sites using EDG sw

>10'000s jobs submitted

>1000 CPUs

>15 TeraBytes disk

3 Mass Storage Systems

Software

50 use cases

18 software releases

Current release 1.4

>300K lines of code

Scientific applications

5 Earth Obs institutes

9 bio-informatics apps

6 HEP experiments

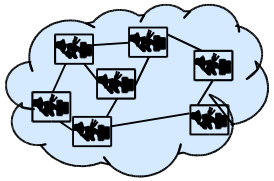
Further Development

- **Further developments in 2003**
  - Further iterative improvements to middleware driven by users needs
  - More extensive testbeds providing more computing resources
  - Prepare EDG software for future migration to Open Grid Services Architecture

- **Interaction with LHC Computing grid (LCG)**
  - LCG deploys LCG-1 service in July
  - Main components of EDG 2.0 release build the basis for LCG middleware

- **New EU project**
  - Make plans to preserve current major asset of the project:
    probably the largest Grid development team in the world
    - EoI for FP6 (www.cern.ch/egee-ei)
    - **EGEE – Enabling Grids for E-Science and industry in Europe**

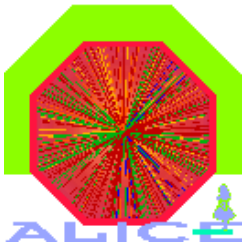
Fundamental Goal of LCG



To help the experiments' computing projects

Phase 1 – 2002-05
prepare and deploy the environment for LHC computing

Phase 2 – 2006-08
acquire, build and operate the LHC computing service



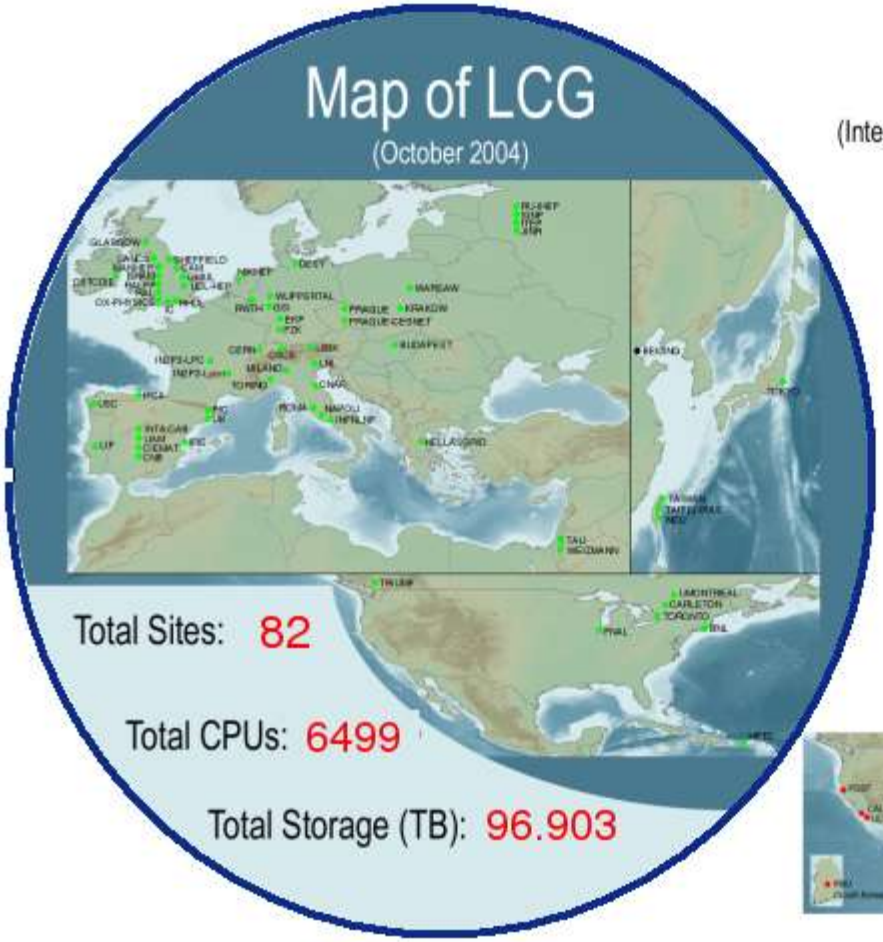
SC2 – Software & Computing Committee

- SC2 includes the four experiments, Tier 1 Regional Centres
- SC2 identifies common solutions and sets requirements for the project

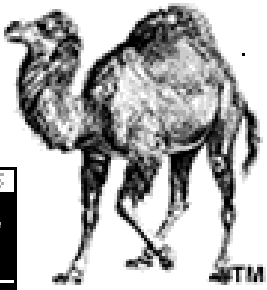
PEB – Project Execution Board

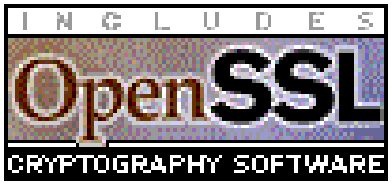
- PEB manages the implementation
  - organising projects, work packages
  - coordinating between the Regional Centres

LCG-2: geographical scope



- Pure Open Source project, started as part of ALICE collaboration (CERN)
- Small development team (very different from EDG)
- Pragmatic approach (what do we have, how can we make it work)
- 3 Million lines of code (cmp. Linux kernel: ca. 5.5 Mio LOC)
- 99 % of the code taken from publicly available packages, mostly Perl
- Only about 1 % of the code had to be developed in addition
- Similar functionality to EDG framework
- Based on WebServices (SOAP, XML)
- Used in other projects, e.g. MammoGrid (UK), a breast cancer database
- See http://alien.cern.ch





Grid Projects with HEP collaboration

European projects

the globus project™
www.globus.org

LCG

Nordic Testbed for Wide Area Computing and Data Handling

CROSSGRID

GriPhyN 2 Data Intensive

iVDGL

DataTAG

DataGRID

GRIDPP UK Particle Physics

INFN Istituto Nazionale di Fisica Nucleare

AliEn @GRID

Many national, regional Grid projects -- GridPP(UK), INFN-grid(I), NorduGrid, Dutch Grid, …

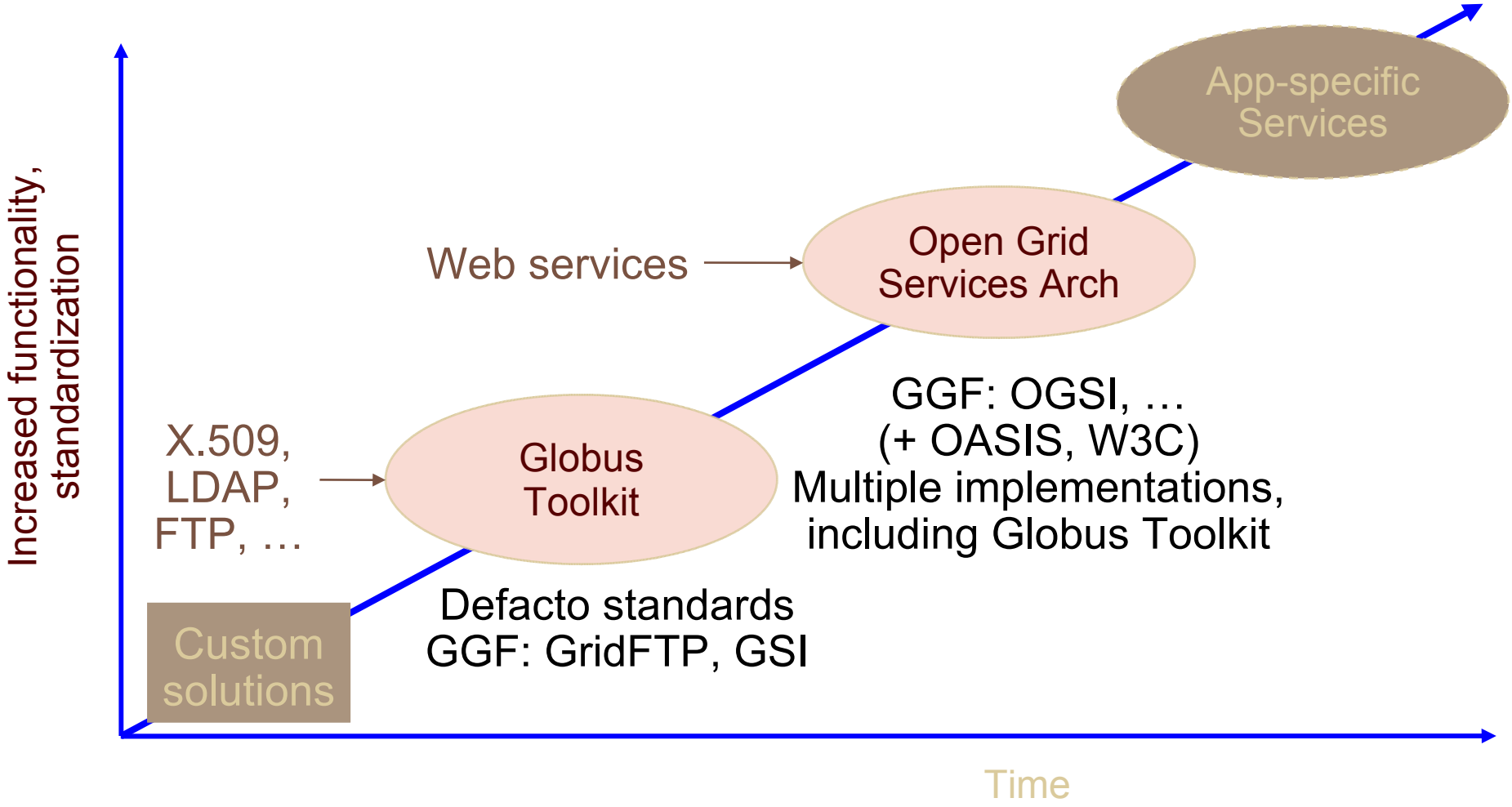
Condor High Throughput Computing

US projects

EGEE Enabling Grids for E-science in Europe

UNICORE

Grids and Open Standards

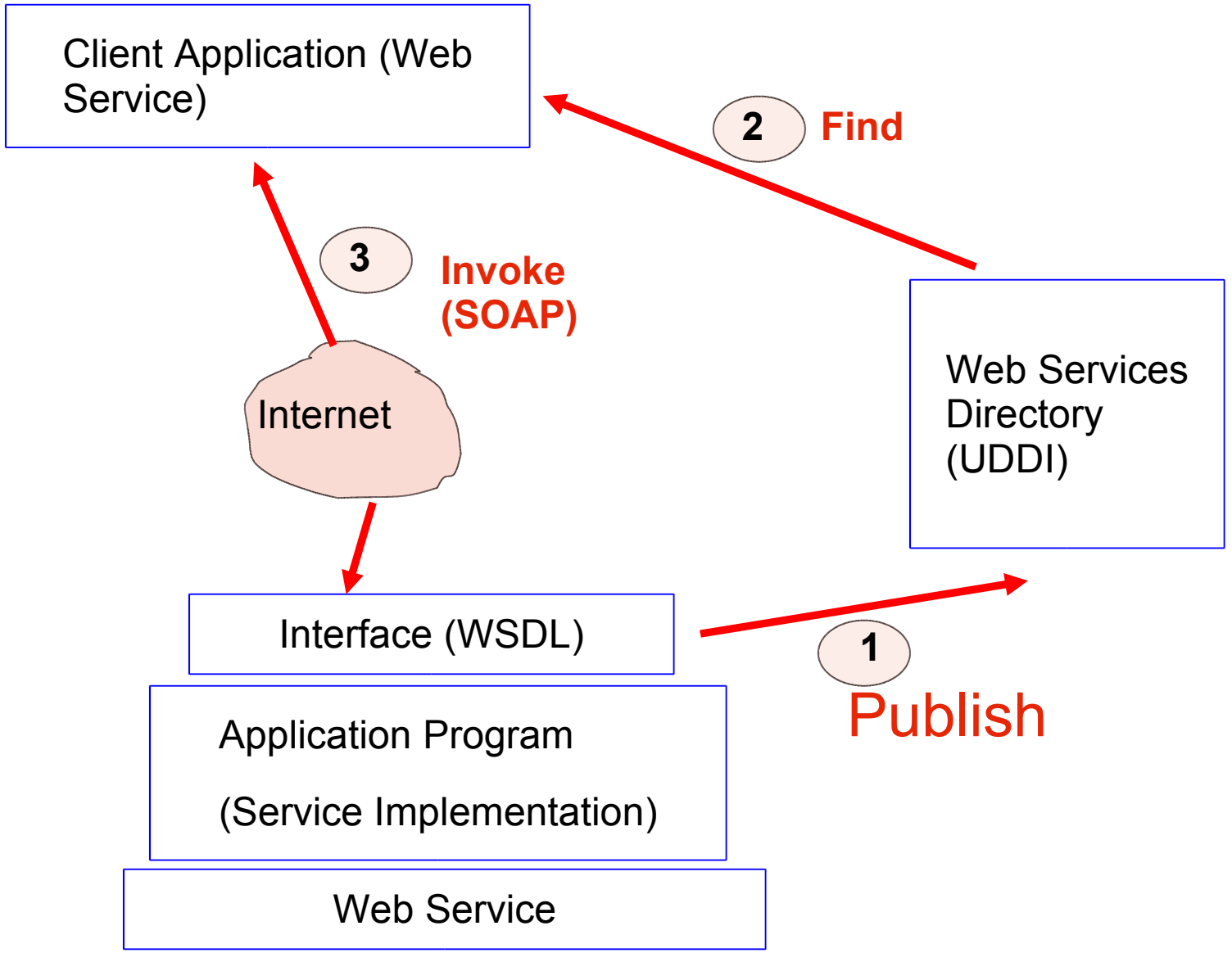


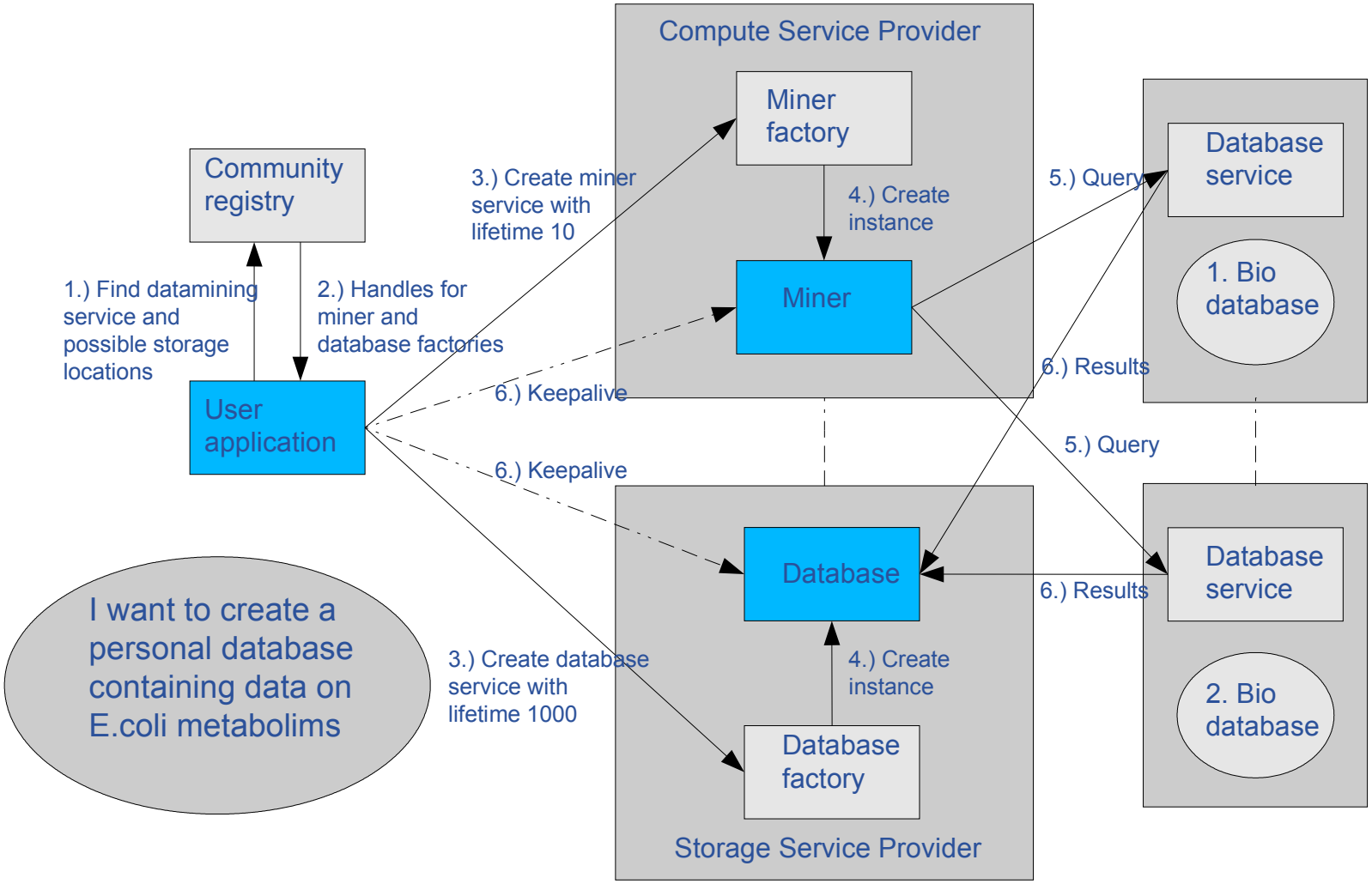
2001 - Webservices

- 2001 – WSDL suggested to W3C
- 2001 – XML Schema 1.0
- 2001 – GEANT, the EU wide Gigabit network activated

Web Services

- GGF 4, Toronto 2002: better approaches are being discussed
- problem: GT2 is too much like a distributed batch system which is not really applicable for business processes
- New idea: network of services, small programms with standardised interfaces, geographically distributed
- Decentrality secures fast growing of Grid technology
- GT3: “Grid Services”
- GT4: Web-service based framework


Client Application (Web Service)
2
Find
3
Invoke (SOAP)
Internet
Web Services Directory (UDDI)
Interface (WSDL)
1
Publish
Application Program
(Service Implementation)
Web Service

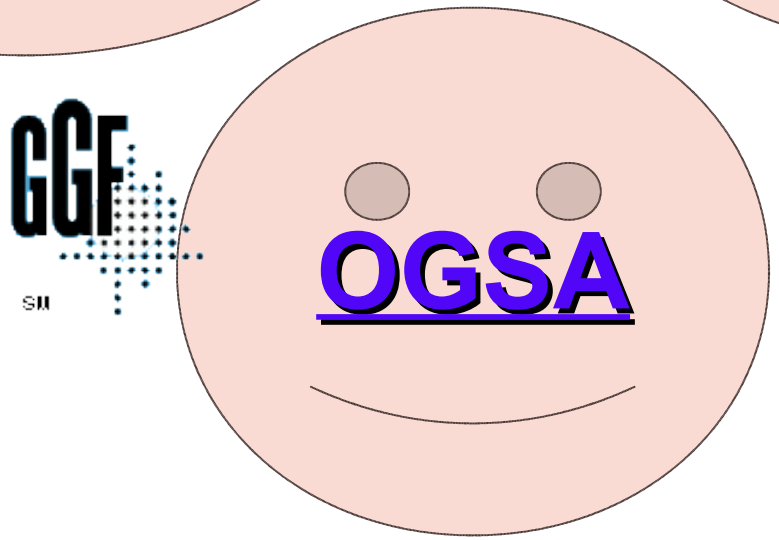

Compute Service Provider
Miner factory
Community registry
3.) Create miner service with lifetime 10
4.) Create instance
5.) Query
Database service
1. Bio database
1.) Find datamining service and possible storage locations
2.) Handles for miner and database factories
Miner
6.) Keepalive
6.) Results
User application
6.) Keepalive
5.) Query
Database
Database service
6.) Results
I want to create a personal database containing data on E.coli metabolims
3.) Create database service with lifetime 1000
4.) Create instance
2. Bio database
Database factory
Storage Service Provider

Two technologies become one

The Grid:
eg. „The Globus Toolkit“
Widely adopted as Grid technology solution

+

Web Services:
a popular standards-based framework for accessing network applications



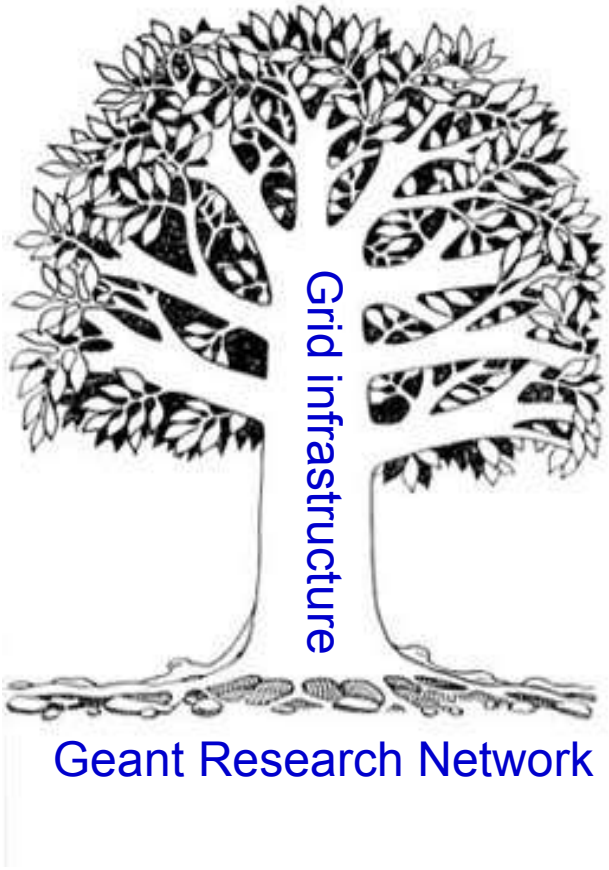
OGSI, OGSA, GT3 + GT4, WSRF

- OGSI defines the essential building blocks for distributed systems – the fundamental mechanisms on which the OGSA platform is built.
- In the core of OGSI: originally Grid Services
- Ongoing development of webservice standards at W3C
- GGF takes care of Grid Services
- Existence of two almost identical technologies leads to uncertainty
- New initiative: substitute OGSI through new standard
- OGSA shall be based directly on web services (WSRF) (no grid services)
- WSRF: 6 specifications, one grid service specific life time management

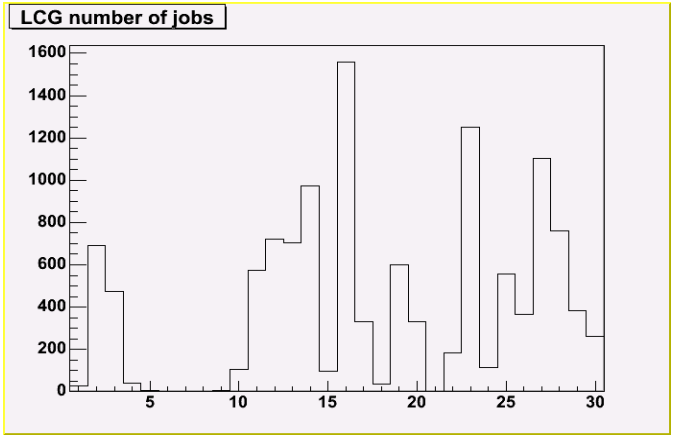
- The grid and web service communities work on the standardisation of WSRF in the framework of OASIS (organisation for the advancement of Structured Information Standards)
- Growing partnership of OASIS, W3C and IETF (Internet Engineering Taskforce)
- Only lowest layer of the specification changes, higher-level services remain untouched

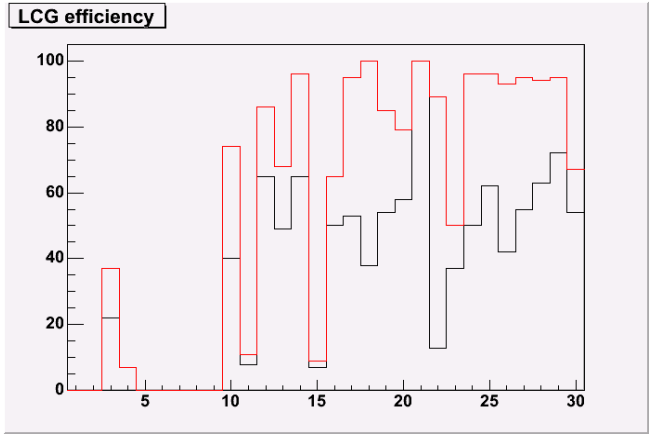
From Grid Research to Grid Deployment

- **Goal**
  - Create a wide European Grid production quality infrastructure on top of present and future EU RN infrastructure
- **Build on**
  - EU and EU member states major investments in Grid Technology
  - International connections
  - Several pioneering prototype results
  - Large Grid development teams in EU
  - Requires major EU funding effort
- **Approach**
  - Leverage current and planned national and regional Grid programmes
  - Work closely with relevant industrial Grid developers, NRENs and US-AP projects



- Standardisation is an important goal but cannot always be achieved. Thus interoperability of different solutions is equally important.
- Alongside sophisticated features, a user-friendly behaviour of Grid components is important to end-users
- Research collaborations must find the right balance between development flexibility and man-power.
- Support and Training play a crucial role in generating a critical mass of users.
- Quality of existing solutions must be ensured by extensive testing, e.g. using data challenges at regular intervals in the context of the LCG project.





Conclusion

- The Grid is not history, but there is quite a bit of history in Grid Computing
- Standardisation is an ongoing process
- Following standards might be a better approach than inventing them

Thanks for listening !

We'd like to thank the German Federal Ministry of Education and Research, BMB+F, as well as Forschungszentrum Karlsruhe / Germany for their continuous interest and support !

bmb+f - Förderschwerpunkt
Hadronen - und Kernphysik
Großgeräte der physikalischen Grundlagenforschung